The brain as a generative model: information-theoretic surprise in learning and action

Dissertation

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Erziehungswissenschaft und Psychologie

der Freien Universität Berlin

vorgelegt von

Sam Gijsen, M. Sc.

Berlin, 2022

Erstgutachter: Prof. Dr. Felix Blankenburg, Freie Universität Berlin

Zweitgutachter: Prof. Dr. Stefan Kiebel, Technische Universität Dresden

Tag der Disputation: 25.04.2023

Acknowledgments

First of all, I would like to thank Prof. Felix Blankenburg for the supervision and always offering time for discussion and help. I am very grateful for all the freedom I was entrusted over these last years which allowed me to explore various topics, but also for the occasional and necessary course corrections.

I would like to thank Prof. Arno Villringer for functioning as a second supervisor and his deliberate advice. Furthermore, I am grateful to Prof. Stefan Kiebel for agreeing to review my dissertation as well as to Prof. Radoslaw Martin Cichy, Prof. Steffi Pohl, and Dr. Timo Torsten Schmidt for agreeing to join the doctoral committee.

My thanks go out to the Berlin School of Mind and Brain and the Deutsche Akademischer Austauschdienst for the continued support, financially and otherwise.

I also thank all the great members of the NNU, past and present, as well as associates: Alex, Christian, Daniela, Dirk, Evgeniya, Gian, Guido, Isil, Jona, Lisa, Pia, Robert, Till, Timo, Yuan-Hao, and especially Miro, for working on PhD challenges together. I really appreciate everyone's help and thank you for making the time at the NNU so enjoyable.

I am sincerely grateful to the family and friends around me. To Fiona, for always being there and for the invaluable inspiration. To my parents, Chrit and Lilian, and my sisters, Karolien and Suzan, for their support as well as enabling and encouraging me to choose my own way.

TABLE OF CONTENTS

Abstract

Our environment is rich with statistical regularities, such as a sudden cold gust of wind indicating a potential change in weather. A combination of theoretical work and empirical evidence suggests that humans embed this information in an internal representation of the world. This generative model is used to perform probabilistic inference, which may be approximated through surprise minimization. This process rests on current beliefs enabling predictions, with expectation violation amounting to surprise. Through repeated interaction with the world, beliefs become more accurate and grow more certain over time. Perception and learning may be accounted for by minimizing surprise of current observations, while action is proposed to minimize expected surprise of future events. This framework thus shows promise as a common formulation for different brain functions.

The work presented here adopts information-theoretic quantities of surprise to investigate both perceptual learning and action. We recorded electroencephalography (EEG) of participants in a somatosensory roving-stimulus paradigm and performed trial-by-trial modeling of cortical dynamics. Bayesian model selection suggests early processing in somatosensory cortices to encode confidence-corrected surprise and subsequently Bayesian surprise. This suggests the somatosensory system to signal surprise of observations and update a probabilistic model learning transition probabilities. We also extended this framework to include audition and vision in a multi-modal roving-stimulus study. Next, we studied action by investigating a sensitivity to expected Bayesian surprise. Interestingly, this quantity is also known as information gain and arises as an incentive to reduce uncertainty in the active inference framework, which can correspond to surprise minimization. In comparing active inference to a classical reinforcement learning model on the two-step decision-making task, we provided initial evidence for active inference to better account for human model-based behaviour. This appeared to relate to participants' sensitivity to expected Bayesian surprise and contributed to explaining exploration behaviour not accounted for by the reinforcement learning model. Overall, our findings provide evidence for information-theoretic surprise as a model for perceptual learning signals while also guiding human action.

Zusammenfassung

Unsere Umwelt ist reich an statistischen Regelmäßigkeiten, wie z. B. ein plötzlicher kalter Windstoß, der einen möglichen Wetterumschwung ankündigt. Eine Kombination aus theoretischen Arbeiten und empirischen Erkenntnissen legt nahe, dass der Mensch diese Informationen in eine interne Darstellung der Welt einbettet. Dieses generative Modell wird verwendet, um probabilistische Inferenz durchzuführen, die durch Minimierung von Überraschungen angenähert werden kann. Der Prozess beruht auf aktuellen Annahmen, die Vorhersagen ermöglichen, wobei eine Verletzung der Erwartungen einer Überraschung gleichkommt. Durch wiederholte Interaktion mit der Welt nehmen die Annahmen mit der Zeit an Genauigkeit und Gewissheit zu. Es wird angenommen, dass Wahrnehmung und Lernen durch die Minimierung von Überraschungen bei aktuellen Beobachtungen erklärt werden können, während Handlung erwartete Überraschungen für zukünftige Beobachtungen minimiert. Dieser Rahmen ist daher als gemeinsame Bezeichnung für verschiedene Gehirnfunktionen vielversprechend.

In der hier vorgestellten Arbeit werden informationstheoretische Größen der Überraschung verwendet, um sowohl Wahrnehmungslernen als auch Handeln zu untersuchen. Wir haben die Elektroenzephalographie (EEG) von Teilnehmern in einem somatosensorischen Paradigma aufgezeichnet und eine *trial-by-trial* Modellierung der kortikalen Dynamik durchgeführt. Die Bayes'sche Modellauswahl deutet darauf hin, dass frühe Verarbeitung in den somatosensorischen Kortizes *confidence corrected surprise* und *Bayesian surprise* kodiert. Dies legt nahe, dass das somatosensorische System die Überraschung über Beobachtungen signalisiert und ein probabilistisches Modell aktualisiert, welches wiederum Wahrscheinlichkeiten in Bezug auf Übergänge zwischen Reizen lernt. In einer weiteren multimodalen *Roving-Stimulus*-Studie haben wir diesen Rahmen auch auf die auditorische und visuelle Modalität ausgeweitet. Als Nächstes untersuchten wir Handlungen, indem wir die Empfindlichkeit gegenüber der erwarteten *Bayesian surprise* betrachteten. Interessanterweise ist diese informationstheoretische Größe auch als Informationsgewinn bekannt und stellt, im Rahmen von *active inference*, einen Anreiz dar, Unsicherheit zu reduzieren. Dies wiederum kann einer Minimierung der Überraschung entsprechen. Durch den Vergleich von *active inference* mit einem klassischen Modell des Verstärkungslernens (*reinforcement learning*) bei der zweistufigen Entscheidungsaufgabe konnten wir erste Belege dafür liefern, dass *ac-*

tive inference menschliches modellbasiertes Verhalten besser abbildet. Dies scheint mit der Sensibilität der Teilnehmer gegenüber der erwarteten *Bayesian surprise* zusammenzuhängen und trägt zur Erklärung des Explorationsverhaltens bei, das jedoch nicht vom *reinforcement learning*-Modell erklärt werden kann. Insgesamt liefern unsere Ergebnisse Hinweise für Formulierungen der informationstheoretischen Überraschung als Modell für Signale wahrnehmungsbasierten Lernens, die auch menschliches Handeln steuern.

List of abbreviations

AP Alternation probability

BOLD Blood-oxygen-level dependent

BS Bayesian surprise

CS Confidence-corrected surprise

DC model Dirichlet-Categorical model

EEG Electroencephalography

fMRI Functional magnetic resonance imaging

HMM Hidden Markov model

MB Model-based

MF Model-free

ms Millisecond

PS Predictive surprise

RL Reinforcement learning

SARSA State-action-reward-state-action (algorithm)

S1 Primary somatosensory cortex

S2 Secondary somatosensory cortex

SP Stimulus probability

TP Transition probability

List of original research articles

Study 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.

Study 2

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

Study 3

Gijsen, S., Grundei, M. & Blankenburg, F. Active inference and the two-step task. Scientific Reports 12, 17682 (2022).

* Shared authorship

Chapter 1

Introduction

Humans face significant uncertainty when operating in the world. The brain only has indirect access to the environment as it relies on signals arising in the early sensory system, which has considerable neural noise. Furthermore, it has to deal with fundamental inconclusivity such as a 2D image on the retina being an ambiguous projection of a 3D object. A prominent viewpoint of how our brains overcome such challenges states that we represent uncertainty and perform probabilistic inference (Knill and Pouget, 2004). Using a generative model, which includes information about how possible sensory input is caused by hidden states in the world, the brain is posited to use previous experience and knowledge of statistical dependencies in the environment to resolve uncertainty. This ultimately furnishes a best guess of the environmental cause of sensory input, with the reliance on prior knowledge increasing when sensory measurements are less reliable. More specifically, probability distributions encoding *prior* beliefs about hidden states, $p(s)$, are combined with a *likelihood* describing the probabilities that sensory data is generated by certain states, $p(y|s)$. These complementary pieces of information may be combined to yield a *posterior* belief $p(s|y)$. This inferential process is described by Bayes theorem:

$$p(s|y) = \frac{p(s)p(y|s)}{p(y)}$$

where $p(y)$ is the overall probability of observing y across all states s included in the generative model. Due to the central role Bayes theorem plays in probabilistic inference, the hypothesis that computation in the brain rests on analogous principles is commonly referred to as the *Bayesian brain hypothesis*. Despite a great surge of interest in the previous two decades, principles underlying the Bayesian brain hypothesis are already found in the work of Helmholtz, who proposed viewing perception as *analysis-by-synthesis* (Helmholtz, 1856). Fundamentally, perception is not considered as the recording of raw input, but by appreciating the faced uncertainties one is naturally led to consider a reliance on prior knowledge. The act of perception then corresponds to making inferences about the causes

of input, while learning corresponds to estimating the parameters of the generative model (Friston, 2005). For a task like learning, it has been thought since at least the work of Hume to be impossible without prior information, with this knowledge nowadays embedded in No Free Lunch theorems (Wolpert, 2012). What aspects of human functioning are Bayesian and to what extent, are open problems relevant to much of contemporary neuroscience. Nevertheless, a formal account of the encapsulation of prior information in an internal, generative model and its inversion forms the basis for current widespread views on brain function.

Once it is assumed that prior knowledge and its integration with new evidence is crucial to human functioning, it becomes necessary for this knowledge to be optimized so as to stay synchronized with the environment. This process of aligning the generative model with novel information has been described as surprise minimization (Friston, 2005) and it may follow intuitively that selecting the model and its parameters which render new observations least surprising is an appropriate strategy. As exact Bayesian inference can be intractable, the brain is often assumed to engage in approximate inference instead (Daw et al., 2008; Findling et al., 2021). In case belief updates are approximated using variational inference can the process be shown to correspond to minimizing surprise (Friston et al., 2006). This suggests a plausible neural framework for probabilistic inference for perception and learning. By borrowing concepts regarding surprise originating in information theory (Shannon, 1948), which has close links with probability theory, surprise can be used as a measure of how informative outcomes are given currently held beliefs. This approach has been applied in psychology and neuroscience to describe attentional processes (Itti and Baldi, 2009) and in machine learning (Schmidhuber, 2010). For the study of (perceptual) learning, various surprise functions (Modirshanechi et al., 2022) have been used to describe learning behaviour itself (Nevo and Erev, 2012; Schwartenbeck et al., 2015) and model neural signals (Mars et al., 2008; Kolossa et al., 2015; Amado et al., 2016).

By applying surprise to future potential outcomes, an expectation of surprise may be computed. As expected surprise corresponds to the estimated mismatch of current beliefs with the (future) environment, it can be interpreted as expected information gain (Schwartenbeck et al., 2013; Friston et al., 2015). In other words, it addresses the question of 'How much is there to learn?'. Expected surprise has therefore shown up in optimal exploration strategies by promoting agents to seek and resolve uncertainty (Sun et al., 2011;

Little and Sommer, 2013). Surprise minimization is thus a perspective that may be used to describe probabilistic inference in the service of perception and learning on the one hand, and action on the other. Here we apply information-theoretic surprise quantities to the empirical study of both perceptual learning and action, which have traditionally been described predominantly in a separated fashion and have only recently began to be merged. By leveraging model comparison methods, we investigate the role of surprise functions 1) to explain neural learning signals as captured by EEG and 2) to explain human behaviour on a decision making task.

First, we will briefly review key ideas related to predictive processing in perceptual learning and a traditional view of action in terms of reinforcement learning. Afterwards, we describe how information-theoretic views on surprise minimization may be used to bridge concepts in learning and action using Bayesian and active inference. Finally, the introduction will present the used experimental paradigms and the modeling approach.

1.1 The brain as a generative model

Under the Bayesian brain hypothesis, the role of prior knowledge is to enable a 'best guess' of an environmental cause of ambiguous and noisy sensory input. The more accurately prior knowledge captures the environment, the better resulting inference may be. The specifics of the context are likely to determine which information may be leveraged. For example, when considering the temporal domain, environmental structures that change only slowly over time are better learned across longer time-horizons. This allows for stable knowledge that is unaffected by short-term noise. One would also expect that this knowledge is innate or learned early in life, such that it may be continuously exploited thereafter. A well-known example concerns the biased human visual perception of line orientation. Specifically, humans are thought to be biased toward cardinal (horizontal and vertical) orientations, which matches the distribution of line orientations in natural scenes (Girshick et al., 2011). Furthermore, infant development of inferring visual line orientation starts soon after birth and continues into early childhood (Siu and Murphy, 2018). Even more generally, the hierarchical structure of the visual world, a constant across evolutionary time, is thought to have contributed to the human hierarchical visual system (Lee and Mumford, 2003). Ultimately, environmental structure may be recoverable from an organism that maintains a generative model of its environment (Kiebel et al., 2008). Beside statistics which are stable

over long time frames, prior beliefs should also capture current stimulus statistics. This has been tested by exposing people to stimulus distributions and observing that reproduced samples consider estimates of the mean and variance of the empirical distributions (Jazayeri and Shadlen, 2010). Less explicitly, a representation of uncertainty has been recovered from neural data (Strange et al., 2005; Bestmann et al., 2008; McGuire et al., 2014; van Bergen et al., 2015) as well as behaviour on learning (Meyniel et al., 2015; Heilbron and Meyniel, 2019) and decision making tasks (Bland and Schaefer, 2012; van Bergen and Jehee, 2019). As mentioned, Bayes provides the degree to which prior knowledge should be traded-off against incoming sensory evidence. This trade-off should depend on how confident we are about current beliefs and how reliable our observations are. To test the weighting of prior and incoming information, noise may be manipulated to experimentally trigger changes in information integration (Mareschal et al., 2013), with basic Bayesian accounts providing good descriptions of human behaviour (Stocker and Simoncelli, 2006; Vilares et al., 2012; Wei and Stocker, 2015). In general, the literature suggests a) that humans rely on prior knowledge as well as b) this information being dynamically weighted against novel evidence.

1.1.1 Bayes and perceptual learning

Given a generative model of the environment, which encodes knowledge about how sensory inputs result from causes, prediction becomes possible. A theoretical paradigm of predictive processing has become dominant in many branches of neuroscience, of which the free energy principle and predictive coding are prominent examples. Generally traced back to similar starting points as the Bayesian brain hypothesis (Hohwy, 2018), this predictive processing paradigm frames prediction as a central function of the brain. It allows for a formulation that leads to approximate Bayesian inference without requiring any knowledge of Bayes theorem. Indeed, as exact Bayesian inference is often intractable, Bayesian brains may deviate from strictly Bayes-optimal inference by using approximation and are in theory compatible with certain probabilistic reasoning errors (Smith et al., 2022). Often, predictions are assumed to be generated constantly, with resulting errors fulfilling the important role of feedback, guiding belief updates to lead to better prediction in the future. However, how much weight should be given to a prediction error in changing prior beliefs? Again, the optimal influence is provided by Bayes and depends on uncertainty of the prior and the likelihood. As such, prediction error schemes include precision terms to set the learning rate. This amounts to

the claim that not only point-estimates are modeled, but also their associated uncertainty. As predictions are corrected and beliefs become more accurate through iterative interaction, the world model of an observer should come to more accurately represent the environment. As we will expand on later, schemes may propose a form of approximate Bayesian inference through minimizing prediction errors.

The predictive coding scheme constitutes a prominent example and assumes the brain to contain hierarchies which generate sensory input from higher-level causes (Friston, 2005). As a consequence, specific roles are assigned to feedforward (or bottom-up) and backward (or top-down) connections. Top-down connectivity represents the environmental causal structure, carrying predictions, while bottom-up connections signal resulting prediction errors. Although a theory such as predictive coding may be grouped under the umbrella of the Bayesian brain, it is defined through its assumptions that instantiate specific, testable claims about the brain. As summarized by Gershman (2019), these include a hierarchical, probabilistic model with a specific approximation family, free energy approximation, and optimization scheme. The result is that the computational description may be attempted to be concretely mapped to neuronal populations and their connectivity, allowing for more specific, testable hypotheses (Bastos et al., 2012).

This framing is attractive as it potentially offers a cohesive description of perception, learning, and action by applying prediction error minimization to different timescales (Millidge et al., 2021). Perception then corresponds to inference about latent environmental states on short timescales and aims to surmise causes of sensory information in the moment (Bogacz, 2017). Learning may be mapped to dynamics unfolding across experiences and interactions with the environment relating to adjustments of the world-model (Friston, 2010). Finally, prediction errors may be minimized by sampling data that best aligns with predictions generated by the subjective model (Friston et al., 2016). There are consequently two hypothesized pathways of optimizing predictions: changing the input itself by acting on the external world or adjusting the model internal to the agent.

Especially predictive coding has received considerable attention investigating its relation to redundancy reduction and efficient coding (Huang and Rao, 2011). Earlier work already noted that rather than fully transmitting sensory input via forward connections, only relaying the part of the data that was not predicted leads to minimum redudancy in neural signaling (Mumford, 1992). This was initially hypothesized for early sensory processing

using concepts related to information theory (Attneave, 1954; Barlow et al., 1961), with Rao and Ballard (1999) providing important initial empirical support concerning visual receptive field effects. The contemporary and neuroscientific theoretical forms of predictive coding and free energy minimization still stay close to the concept of minimal redundancy (Huang and Rao, 2011; Millidge et al., 2021). This is an especially important characteristic for neuroscience as resource constraints are expected to have selected for highly efficient algorithms across evolutionary time.

The predictive processing framework has been successfully applied to explain a variety of phenomena. These include end-stopping (Rao and Ballard, 1999), bistable perception (where perception oscillates between two interpretations of a constant stimulus; Weilnhammer et al. (2017)), as well as perceptual illusions (Watanabe et al., 2018) and repetition suppression (Auksztulewicz and Friston, 2016). Further, classic learning signals such as the mismatch negativity and P300 as recorded by electroencephalography (EEG) have been found to relate to error-signaling and updating dynamics in the perceptual learning domain (Lieder et al., 2013; Kolossa et al., 2015; Nassar et al., 2019). Similarly, functional MRI has been used to map brain networks to prediction error signaling (Fouragnan et al., 2018). The framework also shows promise in aiding the understanding of psychopathology, with example applications including autism (Van Boxtel and Lu, 2013) and psychosis (Sterzer et al., 2018). Further, representations of predictions and prediction errors appear to be neurally represented in a segregated manner (Muckli et al., 2015; Kok et al., 2016; Lawrence et al., 2019). However, studies commonly do not investigate whether alternate models may also account for the data (Walsh et al., 2020). In sum, despite originating in a perceptual learning context, the predictive processing paradigm and in particular the predictive coding implementation has seen widespread application in cognitive computational neuroscience.

1.1.2 Action via reinforcement learning

Implementations of the predictive processing paradigm have some overlap with the field of reinforcement learning (RL), which has been extensively applied in neuroscience to study human action and, to a lesser degree, learning. Nevertheless, RL had a different starting point, which is arguably captured to a significant degree by the reward hypothesis. It states that goals may be fully specified as the maximization of the expected value of the cumulative sum of reward (Sutton and Barto, 2018), with reward being a scalar value at

each time step. Essentially, the reward signal is used to specify *what* the agent should achieve without communicating *how* it should be achieved. The field of RL is nowadays incredibly broad and populated by a diverse set of ideas, yet the concept of relying on reward per se is still present in some of the prominent literature (Silver et al., 2021).

In traditional RL accounts, action-selection uses a mapping from states (e.g. locations in space or game board positions) to actions based on the relative action values, which correspond to estimated cumulative reward. This information is described in a value function, which in neuroscience is often modeled to be learned via reward prediction errors. This process corresponds to updating the estimated value by computing the delta between this estimate and the actual observed sum of rewards (Rescorla, 1972). Versions of such a delta-rule have been shown to approximate Bayesian inference for certain problems (Nassar et al., 2010; Wilson et al., 2013). Distributional reinforcement learning, which estimates a distribution over reward rather than relying on a single scalar, has been suggested as an extension to this framework (Bellemare et al., 2017). Despite the novelty of the application, promising evidence has already been provided for its neural implementation in humans (Dabney et al., 2020; Lowet et al., 2020).

Classical accounts of reinforcement learning have been highly influential by providing accurate models for dopamine neurons as encoding value and prediction error (Schultz, 1998; Schultz et al., 2017). A large body of literature has also provided great insights into the underlying neurobiology of choice behaviour (Lee et al., 2012) and the associated reward prediction error computation (Garrison et al., 2013). Additionally, the manner of temporal discounting has been investigated in many task-settings (Dayan and Niv, 2008), which describes the manner by which humans downweigh distant rewards compared to near-term rewards. A further important contribution of RL has investigated the distinction between model-free and model-based inference (Daw et al., 2005; Gläscher et al., 2010). A model-based strategy uses knowledge of the environmental structure, for example by considering possible transitions between states. This information can be used to more accurately assign value updates to estimates following an action or sequence of actions as well as enable planning. A model-free algorithm is instead more habitual and does not rely on a task model. Rather, updating happens by associating selected actions and visited states directly with obtained rewards. Humans have been suggested to use a combination of model-free and model-based estimates (Daw et al., 2011), although this is likely to depend on the context

(Kool et al., 2016; Castro-Rodrigues et al., 2022). This distinction between strategies has also been leveraged to investigate psychopathology such as gambling disorder (Wyckmans et al., 2019) and obsessive-compulsive disorder (Voon et al., 2015).

1.2 Information theoretic concepts to investigate learning and action

Despite conceptual overlap such as prediction errors and approximations to Bayes, RL has predominantly been used to study choice behaviour, while predictive processing has been mainly applied to the domains of learning and perception. However, ideas which spawned from the Bayesian brain hypothesis have the potential to merge these sub-fields and describe them in a more unified manner. We will introduce concepts originating in information theory with a focus on surprise, and show how by using these ideas surprise minimization can be applied across domains. This will in turn enable a description of how we investigated learning and action using information-theoretic surprise.

Information theory was originally formulated for a communication system and is based on probability theory, with information closely tied to predictability (Shannon, 1948). If a stimulus can be predicted, it is said to carry little information. This is also connected to the concept of minimal redundancy discussed above, which can be equated to minimal predictability and maximal informativeness (Barlow et al., 1961; Spratling, 2017). The information content, also called surprisal or simply surprise, can be defined for an event or observation $y \in Y$. Given a (prior) belief about environmental statistics $\pi_t(\theta)$ at time t, the informational surprise is:

$$PS(y) = -\log \int_{\theta} p(y|\theta)\pi_t(\theta)d\theta = -\log p(y) \tag{1.1}$$

where the negative logarithm is taken of the subjective (i.e., estimated) probability of observing y having marginalized over model parameters. This original definition of Shannon's surprise will be referred to as predictive surprise $PS(y)$ so as to dissociate it from alternative definitions introduced below and to free up the term surprise to refer to the general concept of observations deviating from probabilistic prediction. Equation 1.1 states that if the probability of observing y is low, the event would be considered surprising and consequently informative. In essence, this is a probabilistic formulation of a prediction error, although it is unsigned. A generative model enabling accurate predictions will generate low surprise,

suggesting an intuitively link between surprise minimization, predictive ability, and a fit generative model. As discussed, mismatches between predicted and actual sensory input play an important role in probabilistic inference. Faraji et al. (2018) argue that such errors of prediction result in *puzzlement surprise*, concerning the initial realization of an existing discrepancy between beliefs and observation.

In the context of biology, a potential issue with predictive surprise as puzzlement surprise is that it is insensitive to confidence. Faraji et al. (2018) make the point that many everyday events are highly unlikely to occur, such as parking next to a specific car in a large parking lot, but at least on a experiential level do not seem to instantiate surprise as they are irrelevant to us. In response, it is hypothesized that surprise also requires a commitment to a belief. For example, the realization that the car in question belongs to a family member living abroad may suddenly elicit significant surprise. The authors proposed confidence-corrected surprise, a quantity which scales with the entropy of the prior. Entropy is another information-theoretic concept and denotes the expected information or predictive surprise of observing the outcome of a random event:

$$H(Y) = \mathbb{E}\left[PS(Y)\right]$$

For the current purpose, we apply it to the belief distribution π_t to read out the agent's confidence or commitment to a belief, said to correspond to the negative entropy:

$$-H(\pi_t) = \int_\theta \pi_t(\theta) \log \pi_t(\theta) d\theta$$

Given a probability distribution, entropy is low if a large portion of the probability mass is assigned to a small area of the hypothesis space, as prior to further observations the outcome is highly predictable. The confidence-corrected surprise is then defined as the Kullback-Leibler divergence (see below) between the agent's prior and the posterior of a naive observer, which observed y with an uninformed prior $\hat{\pi}(\theta)$:

$$CS(y) = KL\left[\pi_t(\theta)||\hat{\pi}(\theta|y)\right] \tag{1.2}$$

Confidence-corrected surprise reads out not only the subjective probability of an event, but also depends on the confidence associated with this estimated probability. When the

underlying learning model is fixed, modeling neural signals as predictive or confidence-corrected surprise allows for the investigation whether surprise computation is sensitive to distributional uncertainty.

We may also define a relative entropy term by taking the Kullback-Leibler divergence between two probability distributions (Kullback and Leibler, 1951), e.g. the Bayesian prior $\pi_t(\theta)$ and posterior $\pi_{t+1}(\theta)$ (after having observed y):

$$KL\left[\pi_t(\theta) \| \pi_{t+1}(\theta)\right] = \int \pi_t(\theta) \log \frac{\pi_t(\theta)}{\pi_{t+1}(\theta)} d\theta \tag{1.3}$$

This quantity, when applied to a prior and posterior belief distribution, is also known as Bayesian surprise and can be used to capture the change of the internal generative model in response to new observations (Itti and Baldi, 2009).

$$BS(y) = KL\left[\pi_t(\theta) || \pi_{t+1}(\theta)\right] \tag{1.4}$$

It is not categorized as an initial form of puzzlement surprise as its computation requires the agent's posterior, requiring that beliefs have been updated (Faraji et al., 2018). Consequently, by quantifying the change to the belief distribution in response to new data, Bayesian surprise is a readout function that enables inference of belief updates of a probabilistic learner. A comparison with puzzlement surprise may be interpreted as a contrast between signals of model inadequacy and model updating (Figure 1A). Finally, confidence-corrected surprise can be shown to increase with predictive surprise, Bayesian surprise, and the negative entropy of $\pi_t(\theta)$ (Faraji et al., 2018).

The concept of information as defined by Shannon has received considerable attention in biology and neuroscience and led to the idea of information as a crucial biological resource. This is supported by the observation that animals expend substantial amounts of energy to both obtain and process information (Laughlin et al., 1998; Laughlin, 2001). The evolution of these ideas was bootstrapped by the field of cybernetics, which broadly dealt with systems displaying circular causality or feedback (Wiener, 2019), such as an organism acting on the world and thereby generating and perceiving new input. Research has explored analogies between Shannon's initial communication model and the perception-action cycle (Klyubin et al., 2004; Lungarella and Sporns, 2006; Klyubin et al., 2008; Tishby and Polani, 2011). This approach has generated extensions to the RL framework by modeling

not only expected reward but additionally a sensitivity to informational quantities such as information-gain (Sun et al., 2011), information processing cost (Tishby and Polani, 2011), or entropy (Schwartenbeck et al., 2015). In this way, it offers additional incentives beyond pure reward-maximization which hold the potential promise of giving rise to self-organized intelligent processing.

The current thesis is concerned with incentives that result from pursuing surprise minimization, which can correspond to Bayesian inference when applying the framework of variational inference. First, recall that the Bayesian brain hypothesis does not require the brain to perform exact Bayesian inference, which becomes intractable rather quickly as complexity increases. Instead, approximate inference may be performed, with schemes falling into one of two broad categories. On the one hand, Monte Carlo methods are sampling-based and allow for numerical approximations to probability distributions and stay tractable with a limited amount of samples (Binder et al., 1993). These algorithms correspond to randomly picking one hypothesis at a time and evaluating their relative probabilities. The longer this process is repeated, the closer one's knowledge approaches the full posterior distribution. Importance sampling (including particle filters) and Markov chain Monte Carlo are classes of commonly used algorithms (Hastings, 1970). However, finite samples lead to inferential biases, some of which are systematic and have been shown to match human errors (Sanborn and Chater, 2016).

More relevant here and as an alternative to non-parametric sampling methods, inference may be turned into an optimization problem. In such *variational* approaches, an approximation to the posterior $p(s|y)$ is proposed ($q(s)$) which belongs to a family of distributions. In order to select $q(s)$, consider the following optimization problem:

$$\hat{q}(s) = \text{argmin}_{q(s)} KL\left[q(s)||p(s|y)\right] \tag{1.5}$$

with $KL\left[q(s)||p(s|y)\right] = 0$ if $q(s) = p(s|y)$, that is, if our approximation exactly matches the true posterior. Note that if the true posterior does not belong to the chosen parametric family, then this equality cannot be achieved and the approximation will always deviate. At this point, not much progress has been made because the computation of the KL divergence requires $p(s|y)$. However, a useful reformulation provides the following relation (Ostwald

et al., 2014; Blei et al., 2017):

$$\log p(y) = KL\left[q(s)||p(s|y)\right] - F\left[q(s)\right] \tag{1.6}$$

$$F\left[q(s)\right] = \sum_{s} q(s) \log \frac{q(s)}{p(y,s)} \tag{1.7}$$

The second line describes the variational free energy, $F\left[q(s)\right]$, which is known as the negative of the evidence lower bound, a quantity common in machine learning (Bishop and Nasrabadi, 2006). This is due to the log marginal likelihood, $\log p(y)$, also being known as model evidence as it can be used in Bayesian model comparison to score models based on their probability of having generated the observed data. We cannot optimize the KL divergence directly, but due to the relation in equation 1.6, minimizing $F\left[q(s)\right]$ is equivalent. This results from $\log p(y)$ being independent of $q(s)$ and thus changes in the KL divergence and $F\left[q(s)\right]$ cancel each other out. As free energy is iteratively minimized it thus also maximizes model evidence while providing an improved approximation $q(s)$ to the true posterior, of which the error bound depends on the chosen family distribution. Equivalently stated, minimizing the KL divergence between the variational and posterior distribution would minimize the free energy and render it a closer approximation to the log model evidence.

Note that we already saw the negative of $\log p(y)$ above as predictive surprise and thus engaging in this optimization problem furnishes free energy as a bound on surprise. This means that the average surprise is reduced when observing an event for a second time (Friston, 2010). Computing surprise per se is often intractable, as in realistic scenarios the state space can get very large (equation 1.1). Approximate inference provides a manner by which surprise may be estimated at the cost of giving up guarantees of finding exact posterior beliefs. In experimental work, we may also operationalize surprise by assuming simple generative models for participants and thereby enabling exact inference. As information is defined using probability theory, if it plays a role in human functioning it can only be computed using subjective (i.e. estimated) probability distributions. It is this probabilistic information which is suggested to be encoded in the internal generative model. The application of predictive coding as a neuroscientific hypothesis has provided a formal description of perception as inference by framing predictive coding as approximate Bayesian inference via variational inference and thereby bringing it under the Bayesian brain hypothesis (Friston, 2005; Millidge et al., 2021).

For completeness, we note that if no restrictions are placed on the family distribution in variational inference, then we are still left with an infinite summation or integral. To make variational inference practical, restrictions are necessary, which will no longer guarantee that the true posterior can be obtained. A common restriction is called the *mean field approximation*, which assumes that the (approximate) posterior factorizes into the product of the dimensions of state space:

$$q(s) = \prod_i q_i(s_i)$$

Although it generally works well, this approach is a considerable simplification which can bias inference similar to sampling approaches. As the occurrence of these inferential errors may be tested, these restrictions are one way through which hypotheses assuming variational inference in the brain become falsifiable. For the mean-field approximation, trial-order effects are part of the resulting biases, which have been observed in human behaviour (Daw et al., 2008; Sanborn and Silva, 2013; Gershman, 2019).

1.2.1 Surprise about the future

Next, we comment on the relevance of information-theoretic quantities related to surprise when applied to the future and action, moving beyond perception and learning. To restate, by integrating prior and novel information, the parameters of the generative model may be estimated through observation of the world. Action relates to this process in two fundamental ways. First, action is ultimately served by learning. That is to say, action is the process by which we interact with the world to attain our goals, which an accurate generative model can support, simply by providing more reliable information about the environment and through planning. By applying probabilistic inference to action, the concept of *planning-as-inference* emerges (Botvinick and Toussaint, 2012). By including action and their outcome dynamics in the generative model, an agent may infer on possible consequences and plan how to reach specific future outcomes. This has especially innovated the field of goal-directed action through the application of partially-observed Markov decision processes (POMDPs), in which agents may only have access to ambiguous observations generated by an underlying state, rather than directly observing these states (Kaelbling et al., 1998). Here, too, the usefulness of the generative model and its inversion is bounded by how well the model represents the target.

Second, integrating action and basing it on probabilistic inference provides an appealing tool, as action may take on a function beyond reward maximization per se. If maintaining an accurate generative model is crucial, actions which are predicted to bring about highly informative states or observations may be assigned greater value (Schwartenbeck et al., 2019; Sajid et al., 2021). This natural exploration strategy can lead to forms of active learning by seeking out actions that reduce uncertainty, thereby linking back action to serve learning. Such a process is enabled by explicitly representing uncertainty and by being able to infer how actions may lead to changes in beliefs. Whereas learning is concerned with adjusting model parameters to reduce surprise with regard to current observations, action may aim to reduce surprise of future observations and thereby deals with expected surprise. As discussed more detailed in the next sections, both learning and action can thus be used to infer a generative model of the environment.

More concretely, one extension of the Bayesian brain formulation to action selection features expected Bayesian surprise as a loss function. This gain can be formulated as follows:

$$\mathcal{I}(a) = \sum_{y} p(y|a) KL\left[p(s|y,a)||p(s|a)\right] \tag{1.8}$$

where the agent uses its beliefs about which future observations y result from action a and weighs their probability by the extent these observations would change the belief distributions as per the KL-divergence, corresponding to Bayesian surprise as described above (Gershman, 2019). This expectation has been viewed as information gain and its pursuit by humans has been investigated in psychology (Klayman and Ha, 1987; Nelson, 2005) and neuroscience (Yang et al., 2016; Mirza et al., 2018) and has been framed as active learning in machine learning literature, where its used to allow models to query data labels and explore efficiently (Sun et al., 2011; Houthooft et al., 2016). As uncertainty about environmental statistics is reduced by biasing action toward high Bayesian surprise, more accurate beliefs are expected to yield lower (predictive) surprise going forward. That is, maximizing short-term Bayesian surprise can help minimize predictive surprise over a longer period.

The information gain quantity from equation 1.8 also arises in active inference. This framework is derived from the free energy principle and describes the exploration-exploitation trade-off through a probabilistic inference approach (Friston et al., 2006, 2016). It presumes

an agent that maintains a generative model of its environment. Action selection is based on the idea of surprise minimization, which yields a single expression combining incentives for exploration and the realization of an agent's preferences. The resulting quantity, the expected free energy, evaluates actions based on (predictive) surprise of potential future observations. For a comprehensive account of how the expected free energy is derived from the pursuit of surprise minimization please see Friston et al. (2015). The expected free energy of action a at time t can be expressed as follows:

$$G_t(a) = -\underbrace{E_{p(o_t;\pi_t(\theta)|a_t=a)}\left[\ln p(o_t|C)\right]}_{\text{Extrinsic Value}} - \underbrace{E_{p(o_t;\pi_t(\theta)|a_t=a)}\left[KL\left(\pi_t(\theta)|o_t, a_t=a\|\pi_t(\theta)\right)\right]}_{\text{Intrinsic Value}} \quad (1.9)$$

$$= \underbrace{KL\left(p(o_t;\pi_t(\theta)|a_t=a)\|p(o_t|C)\right)}_{\text{Cost}} + \underbrace{E_{p(\pi_t(\theta))}\left[H(p(o_t|\hat{\theta}, a_t=a))\right]}_{\text{Expected Ambiguity}} \quad (1.10)$$

where $\pi_t(\theta)$ is the belief at time t about statistic θ, for example an outcome probability, with the first line highlighting the dual imperative. Specifically, the left-handed $p(o_t|C)$ denotes the prior preferences over outcomes and the right-hand side is the expected information gain term. The prior preference distribution assigns greater probability to more desirable outcomes and agents are therefore considered to 'expect' to realize their preferences. Acting so as to minimize predictive surprise (in expectation) with respect to prior preferences will then promote behaviour that leads to preference realization. On the other hand, minimizing predictive surprise of outcomes based on belief distributions of environmental statistics $\pi_t(\theta)$ requires having accurate beliefs. This entails effective learning, which is promoted by seeking information via maximizing expected Bayesian surprise. As such, the extrinsic value of a given action measures how likely prior preferences are realized, while intrinsic value scores the information gain. The second line shows a different rewriting of the expected free energy. The cost term indicates how close the belief distribution of observations conditional on action a is to the distribution over prior preferences. Meanwhile, the ambiguity term captures the entropy H of the observation likelihood under the most likely value of the probability $\hat{\theta}$, which when minimized promotes actions that are most informative. Indeed, given that entropy is expected surprise, behaviour that aims to minimize surprise of future events can be framed as entropy minimization. Practically speaking, uncertainty is often high in a novel environment and therefore information gain incentives will tend to drive behaviour. Over time, there will remain less to learn and thus behaviour will shift towards

realizing prior preferences instead. Following this theoretical introduction, next we will summarize the empirical literature on surprise relevant to the experimental work of the thesis.

1.2.2 Surprise in perceptual learning

Attempting to study learning in the absence of explicit behaviour by which internal beliefs are read out, begs the question of how the learning process can be probed. Here we focus on the concept of readout functions, which provide quantities hypothesized to be relevant to the neural learning process. To the extent that humans engage in mismatch computation, the aforementioned surprise functions can be considered as readout functions to infer on the underlying learning process as their progression over time is a function of the observations and the beliefs about task-relevant statistics. Of these, the observations are directly known to the experimenter, while the evolution of model parameters (i.e., learning) is the process of interest and depends on the (hypothesized) computational model, which is not only known but even specified. Thus, the dynamics of surprise across time can be regressed against neural signals to infer on their relative plausibility as descriptions of neural computation, as well as allowing for an analysis of the learning model.

EEG provides data with a high temporal resolution which allows for the dissociation of neural signalling of different computations across time. By comparing the evolution of EEG signal and readout quantities across time, we may infer on processes related to surprise and belief updating. Often, Bayesian learning studies focus on specific EEG components, of which the earlier mismatch negativity (MMN) and later P300 are regarded as most relevant to surprise computation and Bayesian learning. Of these, the P300 has received the most attention, with studies showing that subcomponent variability is well described by predictive surprise (Kolossa et al., 2013, 2015; Kopp et al., 2016) and Bayesian surprise (Kolossa et al., 2015; Mars et al., 2008; Seer et al., 2016). Some studies have investigated the entire peri-stimulus timewindow (Ostwald et al., 2012; Maheu et al., 2019; Modirshanechi et al., 2019). Of these, only Modirshanechi et al. (2019) compared multiple surprise functions, including confidence-corrected surprise, finding that predictive surprise was best decoded across the trial. Their work also constitutes one of two publications that include analyses of somatosensory data, with Ostwald et al. (2012) finding instead evidence for the

somatosensory MMN to reflect Bayesian surprise, although no direct surprise comparisons were included.

Besides surprise readout functions, here we additionally aim to investigate important characteristics of the underlying model. As further detailed below, these include whether the model is hierarchical, across which time horizon information is integrated, and which sequence statistics are learned. The combined investigation of surprise functions and model characteristics are rare. However, multiple studies inferred on the time horizon of learning by modeling a forgetting kinetic in combination with a non-hierarchical model. Results indicate local integration of only recent observations (Mars et al., 2008; Ostwald et al., 2012) or suggest a co-existence of multiple time horizons including global timescales (Kolossa et al., 2013; Maheu et al., 2019). In terms of sequence statistics, Maheu et al. (2019) provide evidence for transition probability learning over stimulus or alternation probability alternatives, which is in accordance with results of an earlier re-analysis of multiple behavioural studies on human learning (Meyniel et al., 2016). To the best of our knowledge, no EEG studies on surprise have attempted to also investigate whether learning happens in a hierarchical fashion. Yet, EEG signals have been shown to predict subsequent belief updating in a manner sensitive to the task context using signed prediction errors as a readout function, indicating a hierarchical learning model (Nassar et al., 2019). More commonly, the learning model is inferred by studying decision making. This literature has diverging findings, with indications for hierarchical (Behrens et al., 2007; Heilbron and Meyniel, 2019) as well as non-hierarchical learning (Summerfield et al., 2011; Farashahi et al., 2017) being common. Overall, comprehensive model comparison studies using EEG to investigate surprise are scarce, which complicates literature synthesis. In particular, different surprise functions tend to correlate strongly due to the influence of belief inadequacy on belief updating (Liakoni et al., 2021). This could lead to a scenario where certain neural signals could be described to encode either process when a model of each is only compared to a weak baseline model in absence of the other, which highlights the importance of direct comparisons within a study. Further, inferring on surprise computation in subjects is likely sensitive to the learning model and vice versa, seeing as the surprise is a function of the learning model and the learning model is inferred upon via surprise quantities.

1.2.3 Active inference for action

Active inference has received considerable theoretical attention (Friston et al., 2006; Friston, 2012) and has been investigated in simulation-based work showing its ability to solve learning and decision problems, also in dynamic environments (FitzGerald et al., 2015; Friston et al., 2015; Marković et al., 2021). Investigations using empirical data attempting to validate the framework as matching human behaviour have only more recently appeared. As shown above, expected free energy can be rewritten to include an entropy term, of which the inclusion was found to guide human decision making over and above pure reward maximization (Schwartenbeck et al., 2015a,b), while implicating a role for midbrain dopaminergic activity. When model parameters were fitted to human behaviour, these were found to be predictive of future symptomatology (Smith et al., 2022) and to significantly correlate across time (Smith et al., 2021). Active inference has also recently been used to study pathology and the resulting atypical task behaviour in comparisons to healthy controls. These analyses have described alterations in the decision making process in terms of decision uncertainty (Smith et al., 2021) and action selection precision (Smith et al., 2020) as well as belief updating proceses, indicated by differences in learning rates (Smith et al., 2022) and weighting of interoceptive signals (Smith et al., 2020). Finally, by formulating action selection as combining utility and information objectives, it continues to prescribe sensible behaviour in the absence of any rewards (Schwartenbeck et al., 2015; Sajid et al., 2021).

However, whether the resolution to the exploration-exploitation dilemma proposed by active inference captures human behaviour better than existing models based on RL remains unclear. As discussed, a traditional RL treatment of choice behaviour results in learned point estimates of (expected) reward without a representation of uncertainty. Nevertheless, the commonly used softmax operation still provides these algorithms with decision stochasticity, providing a simple yet effective capacity for exploration with a single parameter that scales the overall tendency for exploration (Sutton and Barto, 2018). It naturally leads to more exploration when action values are similar, while a more exploitative strategy emerges when considerable value differences arise. However, choice stochasticity only yields *random exploration*, while the contrasting concept of *directed exploration* assigns incentives specifically to uncertain actions. The latter is thereby an intentional process that steers ac-

tion selection so as to maximize information gain. Humans have been shown to be sensitive to the information gain associated with observations, for example while reading (Eugster et al., 2016; Kangassalo et al., 2020). Further, expected information gain of observations resulting from action provides an accurate description of visual sensing (Itti and Baldi, 2009; Yang et al., 2016; Mirza et al., 2018). However, the extent to which humans engage in directed exploration by trading off against exploitation in sequential decision making is less well-understood, with some work not observing this behaviour (Daw et al., 2006). On the other hand, evidence for such strategies has been found when the value of information and, by extension, directed exploration has been experimentally manipulated (Wilson et al., 2014; Dezza et al., 2017; Horvath et al., 2021). While evidence has also been documented for more naturalistic settings (Schulz et al., 2019; Frank et al., 2009), it is not found consistently (Riefer et al., 2017). The literature contains a diverse set of formulations for directed exploration incentives, which potentially contributes to the variability of findings.

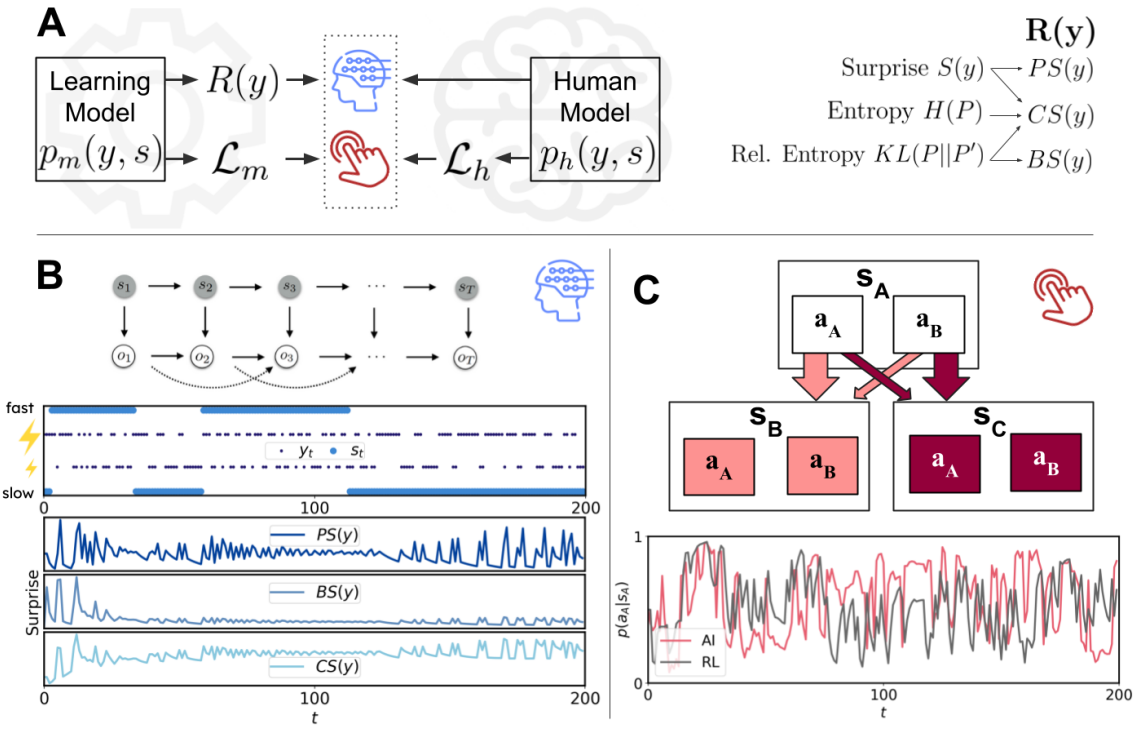


Figure 1: **An investigation of perceptual learning and action using information theoretic tools.** A) The left-hand side includes a schematic overview of the experimental approach using EEG signals and action selection. The human participant is thought to maintain a task model $p_h(y, s)$ which is updated across trials. Neural signals captured by EEG are hypothesized to reflect this process of statistical inference. Further, the task model is used to perform action selection using the participant's loss function $\mathcal{L}_h$. Meanwhile, a computational model $p_m(y, s)$ is posited as a candidate representation of the (hidden) human model $p_h(y, s)$ and the associated learning dynamics. Action selection may be simulated through a combination with a loss function $\mathcal{L}_m$. Beliefs encoded by the candidate model $p_m(y, s)$ may also be passed through a surprise readout function $R(y)$ to yield single-trial regressors, which may be fitted to the EEG data. Through model fitting we aim to align $p_m(y, s)$ with $p_h(y, s)$ and $\mathcal{L}_m$ with $\mathcal{L}_h$, while model comparison methods enable inference of the most appropriate candidate representations of human perceptual learning and action. The right-hand side provides a mapping of concepts originating in information theory to the used surprise readout functions of predictive surprise (PS), confidence-corrected surprise (CS), and Bayesian surprise (BS). B) These surprise functions are used to investigate probabilistic perceptual learning on a 'roving-stimulus' paradigm by generating single-trial regressors (Baldeweg et al., 2004). These may be compared against one another in their ability to explain EEG data. In study 1, the sequence occasionally ($p = 0.01$) switched between two recurring regimes, with the *fast* regime switching between stimulus types more frequently compared to the *slow* regime. C) Next, we investigated whether these information theoretic (surprise) quantities play a role in human action selection on the two-step task (Daw et al., 2011). Active inference (AI) posits that human behaviour is sensitive to expected Bayesian surprise and belief-entropy associated with potential actions to accomplish surprise minimization, which is compared to a traditional reinforcement learning (RL) approach as models for human action selection.

1.3 Paradigms and modeling

1.3.1 The roving-stimulus paradigm and the two-step task

Surprise and related concepts based on information theory are useful tools to investigate the Bayesian brain hypothesis. Not only do they play an important role in providing a formal description of perceptual learning as probabilistic inference, these quantities have allowed for novel frameworks of action selection. Here we introduce our approach of studying learning and action in humans using information theoretic measures.

We studied the implicit learning process of environmental statistics using the well-established roving-stimulus paradigm (Baldeweg et al. (2004); Figure 1B). It features sequences of stimuli which alternate between trains of repeated identical stimuli. Upon observation, the statistics of the generative process that governs alternations and repetitions of stimulus types is hypothesized to be estimated by participants and summarized in their internal model. From an experimental perspective, the roving paradigm is attractive as the generative statistics may be formulated in terms of transitions between stimulus types so that frequencies of stimulus identity by themselves are uninformative. This allows for the study of learning independently of the physical properties of stimuli per se. Further, the transition statistics governing the sequence were altered unbeknownst to the participant during each run. Specifically, the sequence switched infrequently between two recurring regimes, which differed in their tendency to repeat or alternate between stimulus types (Figure 1B). A dynamic rather than static environment calls for sustained learning for the duration of the sequence rather than only at the start and until an initial estimate has been formed. It also allows for an analysis of how humans adapt to dynamic contexts, which is relevant due to the non-stationary human environment and the large variety of theoretical proposals to deal with partially-observable environmental changes.

The roving paradigm has previously been applied to the study of mismatch signals (Cowan et al., 1993; Baldeweg et al., 2004), which have traditionally been elicited by presenting a rare 'oddball' stimulus that differed from a more frequently presented stimulus (Squires et al., 1975; Näätänen et al., 1978). Various explanations for the generation of these responses have been proposed. A prominent example is an adaptation-based hypothesis, stating that mismatch signals result from sensory neurons being adapted to different extents to every stimulus type (May et al., 1999; Jääskeläinen et al., 2004). Meanwhile, a

change detection perspective posits a change in physical sensory inputs to be the underlying cause (Schröger and Winkler, 1995). Successive accounts already surmised that a repetition of a stimulus might lead to a build-up of expectation, which is conceptually closely related to prediction, and that a rare regularity-breaking stimulus may result in prediction violation (Winkler and Czigler, 2012). The current approach of the Bayesian brain may be considered an extension of these initial proposals by allowing for a formal description of the signals resulting from probabilistic inference based on a generative model. This framework further provides the ability to generate single-trial predictions of neural signals, rather than perform comparisons on averaged responses. The discussed surprise functions may be considered formal descriptions of the earlier concepts of expectation violation and model adjustment (Lieder et al., 2013). Despite the relatively long history and extensive research work, much remains unknown about mismatch signaling. Consequently, this paradigm was chosen to further the development and investigation of a surprise-based approach to study the Bayesian brain in perceptual learning.

As the Bayesian brain hypothesis is commonly posited as a general description of brain function, its principles are expected to apply across modalities. Moreover, Bayesian inference can be used to prescribe the optimal integration of information from multiple modalities and therefore underlies promising candidate models of multi-modal processing (Knill and Pouget, 2004; Ernst, 2006; Cao et al., 2019). An understanding of the extent to which mismatch-related computation on roving paradigms is multi-modal and probabilistic in nature is still incomplete (Besle et al., 2005; Butler et al., 2012; Andric et al., 2017). In a further study we extend upon the aforementioned EEG work by presenting auditory, somatosensory, and visual stimuli simultaneously in a roving paradigm. By incorporating cross-modal dependencies in the stimulus-sequence generation process, we intended to apply computational modeling of Bayesian learners and surprise computation to investigate whether mismatch processing is probabilistic and multi-modal in human participants.

Action in a sequential decision making setting has historically been predominantly approached with the primary goal of reward maximization. As suggested, however, the Bayesian brain hypothesis can be naturally extended to action selection, enabling a change to the loss function to include Bayesian surprise or entropy-based incentives. We again take a well-established paradigm, namely the two-step task (Daw et al. (2011); Figure 1C), to contrast such an action selection strategy with a classical RL approach. We focus par-

ticularly on the exploration-exploitation trade-off, relevant to many settings in which it is beneficial or even necessary (Sajid et al., 2021) to temporarily forego rewards in order to collect information that enables the formulation of a better action selection strategy. This tension between maximizing (short-term) reward and learning about the world is often treated as a problem of balancing these two ingredients and is of interdisciplinary interest, spanning multiple fields besides neuroscience (Wilson et al., 2021), including psychology (Cohen et al., 2007) and computer science (Osband et al., 2016; Sutton and Barto, 2018).

The loss function on action proposed by active inference implicates important characteristics relevant to the Bayesian brain hypothesis. First, expected free energy minimization requires the maintenance of probability distributions (or their sufficient statistics; Friston et al. (2011)) to derive the entropy or information gain terms. Second, these internal model-based quantities are hypothesized to influence action selection not only to obtain rewards but also to shape beliefs, thereby instantiating a method by which action contributes to the optimization of the generative model. Moreover, this issue concerns a crucial aspect of the framework, as omitting the information gain incentive (i.e. expected Bayesian surprise) can reduce active inference to produce behaviour that is hard to dissociate from a purely reward-maximizing RL agent. We investigated these ideas using similar model-comparison techniques as used to study perceptual learning with EEG. However, in this case we rely on action selection itself reading out the internal model, so as to make the use of additional readout functions redundant. However, from an experimental point of view, the process is likely to depend on two similar components. Namely, a generative model that is updated based on action outcomes and a loss function that is minimized via action selection, which consults learned statistics such as outcome probabilities and their uncertainty. Our focus is on using model comparison to infer on this loss function and study whether descriptions of human behaviour on the two-step task improve when a surprise-based incentive is considered.

The two-step task is a sequential decision making task introduced by Daw et al. (2011). On each trial, two stages need to be sequentially traversed via binary action selection with the goal of obtaining a reward in the final stage, although some versions require the avoiding of punishment instead (Lockwood et al., 2020). It was originally designed to disambiguate between model-free and model-based strategies. The underlying assumption is that model-based inference utilizes the structural knowledge of the probabilistic transitions between

the initial and final stage to seek out valuable states. A model-free strategy is insensitive to this knowledge and only relies on observed stimulus-action mapping. The task has seen widespread application and therefore has a broad literature available concerning task behaviour of humans and RL models (Wunderlich et al., 2012; Eppinger et al., 2013; Otto et al., 2013; Kool et al., 2016; Castro-Rodrigues et al., 2022). Furthermore, it was selected because the common RL model does not capture all aspects of human task behaviour despite its adoption (Feher da Silva and Hare, 2020), indicating the importance of considering alternate models of behaviour.

1.3.2 Modeling

For the study of learning, we differentiate between the probabilistic model and readout functions. The surprise quantities are used to model neural signals as these are hypothesized to encode probabilistic inference. Although the surprise quantities do not presume a specific neural implementation in an algorithmic sense, they do assume that the dynamics of brain function encode concepts of expectation and prediction, as well as their confidence and updates. A direct comparison of these surprise quantities aims to dissociate (in a spatiotemporal manner) between signals underlying their dynamics. Meanwhile, predictive surprise, Bayesian surprise, and confidence-corrected surprise have been defined as functions of the underlying model and they thereby read out this model. In this way, they may additionally be used to investigate the learning process itself. It is this second functionality that we further expand on in this section by introducing some characteristics of the generative model that we investigated. This dual-purpose also relates to the advantage of investigating both within a study; as learning dynamics estimate the parameters which surprise quantities read out, inference by experimentalists based on surprise quantities is expected to get more reliable as the generative model more accurately captures the neural representation. Vice versa, experimental inference on properties of the generative model may be biased when a readout function is used that does not correspond well to neural computations.

The roving paradigm features few discrete stimuli which enables a simple conjugate Bayesian learning model. A Dirichlet-Categorical model effectively tracks counts of the unique types of observations to infer their emission probabilities. It models the prior over the hidden state s using a Dirichlet distribution with $s_1, ..., s_M \sim Dir(\alpha_1, ..., \alpha_M)$ and the likelihood using the Categorical distribution $y \sim Cat(s_1, ..., s_M)$. As the hidden state s is

modeled as a discrete distribution over categories it is essentially a static representation, with s_t being shared across time for each observation. Nevertheless, changes to the environmental hidden state may be adapted to by incorporating exponential memory-decay, with the severity set by a parameter $\tau \in [0, 1]$. This results in earlier observations being increasingly discounted or forgotten, implying a limited timescale of information integration, for which evidence has been reported repeatedly (Ostwald et al., 2012; Rubin et al., 2016; Maheu et al., 2019). The forgetting rate sets the balance between being flexible to environmental change with increased risk of modeling noise (high forgetting) and being inflexible but resistant to noise (low forgetting). The appropriate level of forgetting is therefore dependent on the volatility of the current context.

When the environment is cyclical, forgetting and relearning can be a wasteful strategy. In response to a change in a friend's mood, rather than forgetting what you have learned about them, it is better to recognize the new mood and reactivate relevant previously learned knowledge. Volatility in this scenario concerns the frequency by which switches between discrete (hidden) states occur. A hierarchical model may infer such switches to construct and maintain separate beliefs for each environmental regime so that forgetting is unnecessary and all previous information can be integrated. Hidden Markov models (HMMs) describe a discrete hidden state s that evolves according to a Markov chain with transition probabilities $p(s_t|s_{t-1})$ (Rabiner and Juang, 1986; Ashwood et al., 2022). This can be described as a matrix $\mathbf{A} \in \mathbb{R}^{K\times K}$, with the number of hidden states usually $K > 1$. At each timepoint, an observation is sampled according to the emission probabilities $p(o_t|s_t)$. Comparing an HMM to a static-state alternative allows for a comparison about which structure (i.e., flat or hierarchical) the brain defaults to in order to estimate environmental statistics (Gallistel et al., 2014; Meyniel et al., 2016). This is possible because the distinction in the representation of the hidden state has important implications for the surprise readout functions. Recall that the sequence statistics occasionally switched between two possible regimes. For the Dirichlet-Categorical model, Bayesian and confidence-corrected surprise are functions of the beliefs about event probabilities per se. Meanwhile for the HMM, these surprise quantities read out the latent level which assigns probabilities to the two discrete hidden environmental states. If surprise signalling is found to pertain to these beliefs, we propose this as evidence in favour of hierarchical inference.

Early studies often assumed that stimulus-identity frequency is the statistic of interest (Näätänen et al., 2007). However, this is not always an optimal choice as sequences often feature transitional structure, such as rain following cloudy rather than sunny weather or happiness predicting continued happiness rather than despair. To test which sequence statistics the brain appears most sensitive to, we may consider the following alternatives:

1. Stimulus probability (SP): $y_t = o_t$ for $t = 1, ..., T$, thus modeling the stimulus identities per se without describing any Markov dependencies.
2. Alternation probability (AP): $y_t = d_t$ for $t = 2, ..., T$ with $d_t = 1_{o_t \neq o_{t-1}}$ being 1 if the current observation o_t differs from o_{t-1}. By capturing alternation and repetition of stimulus identities, the model is sensitive to a limited form of first-order Markov dependence.
3. Transition probability (TP): $y_t = o_t$ for $t = 1, ..., T$ with a set of hidden parameters $\mathbf{s}_1^{(i)}$ for each transition from $o_{t-1} = i$ capturing first-order Markov dependencies (TP_1; Meyniel et al. (2016)). A model can be further extended to include $\mathbf{s}_2^{(j)}$ for each transition from $o_{t-2} = j$ (TP_2). As the hidden regimes which govern sequence statistics differ in terms of their second-order transition probabilities this allows a TP_2 model to capture all dependencies. The regimes may still be dissociated by doing only AP or TP_1 inference, but not by solely relying on SP.

The Dirichlet-Categorical model and HMM can thus estimate different sequence statistics. Whereas an HMM integrates the full history of observations to infer occasional switches between sets of sequence statistics, the Dirichlet-Categorical model with a static hidden state has its horizon of information integration set by a free parameter. The resulting estimated statistics may in turn be used to compute different surprise quantities across time, enabling model comparison analyses to provide insight into their relative fit to the EEG signals.

For the study of action, one may also differentiate between two components: a probabilistic model and the loss function according to which actions are selected. However, as the computation of expected free energy relies on (statistics of) probability distributions, the learning dynamics for active inference agents necessarily differ from those used by the scalar-based RL model that is commonly used in the two-step task. We were therefore only able to compare the two approaches in totality, using supplementary analyses and simulations to clarify contributions of probabilistic learning and a surprise-based loss function to relative model performances.

Generally, a popular and straightforward model-free reinforcement learning strategy involves mapping expected return to state-action pairs $Q(s, a)$ through updating Q values by a prediction error δ, modulated by a learning rate α:

$$Q_{MF}(s, a) = Q_{MF}(s, a) + \alpha\delta \tag{1.11}$$

$$\delta = o_t - Q_{MF}(s, a) \tag{1.12}$$

For the two-step task, the State-action-reward-state-action (SARSA(λ)) algorithm for Markov decision processes can be used (Rummery and Niranjan, 1994; Sutton and Barto, 2018). This introduces an eligibility parameter λ that determines the effect of the final-stage prediction error on the initial-stage action values. It enables a learning process that does not depend on any structural knowledge of the environment. In order to evaluate the alternative strategy of exploiting information about the probabilistic mapping of initial-stage action a_j from the initial-stage state s_A to the final-stage states $\{s_B, s_C\}$ and consequently to final-stage actions a_2, a model-based approach is formulated:

$$Q_{MB}(s_A, a_j) = p(s_B|s_A, a_j) \max_{a_2 \in \mathcal{A}_B} Q_{MF}(s_B, a_2) + p(s_C|s_A, a_j) \max_{a_2 \in \mathcal{A}_C} Q_{MF}(s_C, a_2) \tag{1.13}$$

where $\mathcal{A}_B$ and $\mathcal{A}_C$ denote the sets of available actions in the respective final-stage states. These model-free and model-based Q-value estimates may be combined for initial-stage actions by running both algorithms in parallel and weighting their contribution using a parameter w:

$$Q_{net}(s_A, a_j) = wQ_{MB}(s_A, a_j) + (1 - w)Q_{MF}(s_A, a_j) \tag{1.14}$$

Finally, a softmax operator is applied to transform Q-values into a probability distribution over actions:

$$p(a_{p,t} = a|s_{p,t}) = \frac{\exp(\beta_p Q_{net}(s_{p,t}, a))}{\sum_{a'} \exp(\beta_p Q_{net}(s_{p,t}, a'))} \tag{1.15}$$

with the choice randomness controlled by β_p at stage $p \in \{1, 2\}$.

In contrast, active inference agents maintain a generative model of the task and act to minimize expected free energy via a loss function which includes Bayesian surprise. The

model includes the estimated transition probabilities (θ_1) and outcome probabilities (θ_2), which are together denoted as θ, and takes the following form:

$$p(o_t, s_{2,t}|s_{1,t}, \theta) = p(o_t|s_{2,t}, \theta)p(s_{2,t}|s_{1,t}, \theta)p(\theta) \tag{1.16}$$

Note that we focused on outcome probabilities θ_2 here for the computation of the expected free energy (equation 1.9). First, these probabilities drifted for the duration of the experiment and thus required continual learning, while the transition probabilities were fixed. Further, it was communicated to participants that transition probabilities corresponded to one of two mirrored structures, which means initial-stage actions provide equal amounts of information about these probabilities. Action-selection is consequently only sensitive to information-gain discrepancy regarding outcome probabilities.

To the extent that the resulting information gain is valued relative to reward depends on the prior preference distribution. This distribution captures the relative attractiveness of the different observations, with desired outcomes being assigned higher probabilities. The prior preferences over action outcomes are restricted to a Bernoulli distribution implying $o_t = 1$ is preferred over $o_t = 0$ (Marković et al., 2021):

$$P(o_t|C) = \frac{1}{Z(\lambda)} e^{o_t \lambda} e^{-(1-o_t)\lambda} \tag{1.17}$$

where λ denotes the precision of the prior preferences. In case of zero precision ($\lambda = 0$), there are no preferences, which leads to the singular objective of intrinsic value maximization corresponding to pure information gathering about the outcome probabilities. For higher values of λ, increasingly more weight is assigned to realizing prior preferences, which becomes prioritized over information gain. In sum, λ is a modulatory parameter governing explorative and exploitative behaviour.

The active inference model implemented here is model-based so as to link the expected free energy minimization in the initial-stage of the task to the observations in the final-stage. For the initial-stage actions, the estimated transition probabilities between stages, θ_1, are used to weigh the expected free energy (eq. 1.9) associated with final-stage actions:

$$G(a_j) = p(s_B|s_A, a_j, \theta_1) \sum_{a_2 \in \mathcal{A}_B} G(a_2) + p(s_C|s_A, a_j, \theta_1) \sum_{a_2 \in \mathcal{A}_C} G(a_2) \tag{1.18}$$

where $\mathcal{A}_B$ and $\mathcal{A}_C$ are the sets of available actions in the two corresponding final-stage states. This setup enables an analysis of the contribution of expected Bayesian surprise to initial-stage action selection per se as well as through comparison with the hybrid reinforcement learning model.

To estimate probability distributions for the drifting outcome probabilities θ_2, we chose for a simple surprise-based learning rule by adapting an algorithm proposed by Liakoni et al. (2021). Briefly, to ensure that learning can remain flexible so as to adapt beliefs to the drifting outcome probabilities of the two-step task, prior information is forgotten in proportion to which it is incompatible with new observations. Specifically, predictive surprise of the current observation determines the trial-specific rate of forgetting, implemented as a decay on concentration parameters of a Beta distribution. This allows for the accumulation of information during periods with steady outcome probabilities, while enabling adaptation to more volatile changes. Such surprise-mediated learning was compared with alternatives featuring static forgetting in their ability to capture human behaviour.

1.4 Aim of the thesis

Learning and action, despite traditionally being segregated, may be approached from a common perspective using probabilistic inference. Specifically, information-theoretic concepts such as surprise appear to be useful theoretical as well as experimental tools and have enabled formulations of surprise minimization that are hypothesized to underlie the distinct functions of perception, learning, and action. From an experimental point of view, quantities of surprise are well suited to be combined with model selection techniques. The experimental work of the thesis studied both learning (concerning surprise of current observations) and action (surprise of future observations) in human participants using tasks which are well-established in their respective domains. In regards to the former, we modeled neural signals as measured by EEG to infer both on important properties of the generative model and surprise computations relating to probabilistic belief inadequacy and updating. Using the roving-stimulus paradigm, we investigated learning of environmental statistics in the somatosensory system. This modality was chosen for two primary reasons. First, somesthesis is relatively unexplored in the context of Bayesian learning, while it is generally believed that a similar computational description should be applicable across domains, highlighting the importance of addressing this gap. Second, the spatial separation of the primary and

secondary somatosensory cortices allows for effective use of source reconstruction methods, in contrast to the more densely clustered organisation of the visual and auditory systems. In terms of the generative model, we compared a hierarchical model in the form of a hidden Markov model to a simpler, flat Dirichlet-Categorical model, investigated which sequence statistics are estimated, and inspected the time-horizon of information integration. We read out the learned statistics using surprise functions, enabling both inference on the underlying model as well as generating insight into the use and role of surprise computation in the brain. Specifically, these surprise comparisons allowed for testing whether puzzlement surprise signalling is sensitive to belief confidence and its dissociation from belief updating dynamics. This modeling framework was subsequently extended to a tri-modal roving paradigm in which stimulus sequences also featured cross-modal dependencies. Here we intended to use Bayesian learners and surprise functions to investigate whether findings extend to the auditory and visual modalities and whether mismatch computation on the roving paradigm is multi-modal as well as probabilistic. We hypothesized for surprise computation of a Bayesian learner to accurately model neural signals during sequence learning. Specifically, we expected for mismatch-related computation to include separate belief inadequacy and belief updating signals based on a model estimating stimulus transition probabilities. Given the general and modality-independent nature of probabilistic inference, we hypothesized for these results to generalize across different senses.

Surprise minimization is not only a useful description for perceptual learning but also action selection, which we investigated by analysing multiple behavioural datasets of the widely-adopted two-step task. The paradigm has been largely studied using traditional reinforcement learning algorithms estimating reward via single scalars, while research has noted that these models fail to fully capture important characteristics of human task behaviour. We compared a common implementation of this approach to a probabilistic learning model which uses an active inference-based action selection criterion. This strategy computes the expected Bayesian surprise resulting from possible actions, aiding predictive surprise minimization across longer horizons, thereby naturally providing information gain incentives as part of expected free energy minimization. In this context, we consider the hypothesis that human decisions trade off maximizing short-term reward with minimizing uncertainty in their belief distributions. This intriguing framework still severely lacks empirical validation using human data, especially in regards to information-gathering. To address this, we com-

bine model-comparison methods with a model-agnostic approach to investigate to which extent the considered models capture human behaviour and the contribution of information gain hereto. We expected for active inference-based agents to provide superior descriptions of model-based behaviour compared to a traditional reinforcement learning approach. Specifically, this improvement was hypothesized to result from the information gathering incentive, which is lacking in the pure reward-maximization formulation of RL. In the next chapter, I will describe the leveraged model comparison methods in more detail. This will be followed by a brief overview of the results.

Chapter 2

Summary of empirical studies

2.1 Model fitting and comparison

For the investigation of the described hypotheses, we relied predominantly on model comparison. Following a model fitting procedure to data, this methodology functions as statistical testing by enabling a comparison between two or more competing models that make different assumptions about the data generation process. The data might constitute a timeseries of neural signals as recorded by EEG or behavioural data on a decision making task. The quantity of interest is the model evidence $p(y|m)$ for model m and data y, as we want to be able to compare models in their ability to explain data obtained from participants. To ameliorate the problem of overfitting, the goodness of model fit needs to be traded-off against its complexity:

$$p(y|m) = \text{accuracy}(m) - \text{complexity}(m) \tag{2.1}$$

Various quantities have been proposed to this end, including the Akaike information criterion (AIC), Bayesian information criterion (BIC), and the free energy (Penny et al., 2010).

For the analysis of EEG signals, the models were fit using a variational inference algorithm for multiple linear regression (Penny et al., 2003, 2005; Flandin and Penny, 2007). Regressors, functioning as predictors for the EEG data y, were generated without subject-specific parameters using a combination of a learning model and surprise functions.

$$p(y, \beta, \lambda) = p(y|\beta, \lambda)p(\beta)p(\lambda) \tag{2.2}$$

where β denotes the regression weights and λ the observation noise precisions. The prior variational distributions were selected via a simulation study aimed to select the probabilistic model that minimizes Type II error under the constraint of minimizing Type I error. Upon convergence, the algorithm provides an approximation of the posterior parameter distribution $p(\beta, \lambda|y)$ and yields the free energy F as an approximation to the (log) model evidence. In this case, the model complexity can be rewritten as the KL divergence between the prior and approximate posterior parameter distributions. For models $m = 1...M$ and subjects $s = 1...S$, the overall model evidence given the total data set Y can be obtained

as follows:

$$\log p(Y|m) = \sum_{s=1}^{S} \log p(y_s|m), \text{ using} \tag{2.3}$$

$$F_m \approx \log p(Y|m) \approx \sum_{s=1}^{S} F_{s,m} \tag{2.4}$$

enabling direct comparisons between models by computing differences in free energy (e.g. $F_1 - F_2$ for models $m = \{1, 2\}$), corresponding to log Bayes factors (Penny et al., 2010).

For the modeling of behavioural data, various subject-specific parameters had to be fitted, for which we used a constrained minimization algorithm (L-BFGS-B). The problem of local minima was ameliorated by performing multiple iterations using random initialization, using only the iteration that yielded the highest log likelihood for model comparison. To this end, the Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were used:

$$BIC := k \ln(n) - 2 \ln(\hat{L}) \tag{2.5}$$

$$AIC := 2k - 2 \ln(\hat{L}) \tag{2.6}$$

with k being the number of free parameters in the model, n the amount of trials, and $\hat{L}$ denoting the maximized value of the subject- and model-specific log likelihood function. Thus, these criteria differ from the free energy by characterising model complexity as a function of model parameters and, for BIC, the number of data points. Fixed-effects analyses may be computed similarly by directly comparing sums across subjects.

However, fixed effects analyses implicitly assume that all subjects use the same model. They are also sensitive to outliers as an extreme value can significantly bias the sum in equation 2.3. In response, a random effects approach was chosen for group level inference which has the log model evidences as the only input (Stephan et al., 2009). This procedure uses a generative model for the data Y and models r_m as the frequency of model m in the population. A Dirichlet prior over r_m is used, $p(r_1, ..., s_M|\alpha_1, ..., \alpha_M) = Dir(\alpha_1, ..., \alpha_M)$, where $\alpha_1, ..., \alpha_M$ correspond to the unobserved counts of model occurrences in the population. These α parameters are then optimised to convergence, providing several statistics for model comparison. First, these parameters can be used to compute the expected frequencies $\langle r_m \rangle$, corresponding to the expected likelihood of obtaining model m for any randomly

selected subject:

$$\langle r_m \rangle = \frac{\alpha_m}{\alpha_1 + ... + \alpha_M} \tag{2.7}$$

Second, exceedance probabilities express the probability that model m is more likely than any other tested model given the group data:

$$\phi_m = p(r_m > r_j | Y; \alpha) \text{ for all } j \in \{1, ..., M | j \neq m\} \tag{2.8}$$

which have also been extended to correct for differences in model evidences resulting from chance (*protected exceedance probabilities*; Rigoux et al. (2014)). This method also enables so-called family-level analysis, in which models sharing certain features are grouped into families to infer on diverging features between families (Penny et al., 2010). In this case, rather than specifying uniform prior α_0 parameters across models, they are set equal across families instead.

2.2 Study 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.

Previous studies have investigated perceptual inference and learning by violating statistical regularities and have described such processes using surprise functions. Single-trial EEG signals may be modeled in this manner without the need for behavioural output. The information-theoretic quantities of predictive surprise (Kolossa et al., 2013, 2015; Kopp et al., 2016; Maheu et al., 2019; Modirshanechi et al., 2019; Mousavi et al., 2020), Bayesian surprise (Ostwald et al., 2012; Kolossa et al., 2015; Mars et al., 2008; Seer et al., 2016; Mousavi et al., 2020), and confidence-corrected surprise (Modirshanechi et al., 2019) have been previously applied to EEG signals. However, these studies have predominantly relied on studying single EEG components, rarely include the somatosensory domain, and, crucially, often do not feature direct comparisons of competing formulations of surprise signals. Here, we implement a stimulus-roving paradigm, in which trains of repeated electrical stimulation alternate between two intensities. This well-established task thereby allows for the exploration of mismatch responses independent of physical properties of stimuli, which we leverage to do a comprehensive model-comparison study. Specifically, we adopt a step-wise

analysis in which we infer on the underlying learning model and subsequently provide a spatiotemporal account of surprise signatures in both sensor and source space.

By contrasting neural responses to repeated and deviating stimuli, we identify several mismatch responses across peristimulus time (57, 119, 361ms), each with different linear dependencies on the recent stimulus history. By modeling single-trial signals in sensor space, we provide evidence for the use of a non-hierarchical learning model that learns first-order transition probabilities between stimuli using a local time horizon of integration. Next, this model was used to infer on surprise computations, with early (65-200ms) dynamics reflecting confidence-corrected and Bayesian surprise and indications for later (275-375ms) signals to encode predictive and Bayesian surprise. To perform this analysis in source space, source reconstruction was performed for early (0-200ms) signals, identifying two dipoles in both primary (S1) and secondary (S2) somatosensory cortices, which were consequently used to project single trial EEG data onto. This procedure was supported by finding sensible mismatch responses using the dipole projections, with the early (57ms) mismatch response attributed to S1 and the following (119ms) mismatch response found to result from differences in both S1 and S2 activation. Model comparison in source space identified the same probabilistic model as was found in sensor space and ascribed strong evidence to bilateral S2 activation to reflect confidence-corrected surprise from 70ms and suggested S1 to reflect Bayesian surprise around 140ms. By interpreting these two quantities as uncertainty-sensitive signals of model inadequacy and model updating respectively, these results suggest a possible interaction in the somatosensory system that may contribute to the probabilistic learning of environmental statistics.

2.3 Study 2

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

In study 1, we combined somatosensory stimulation in a roving stimulus paradigm with EEG to provide evidence for multiple distinct mismatch responses across peri-stimulus time with unique dependencies on stimulus histories. Further, we showed that Bayesian learning models combined with surprise readout functions capture observed neural dynamics well.

However, mismatch signalling and its representation as surprise-based computation arising from probabilistic inference has mainly been investigated in uni-modal settings (Näätänen et al., 1978; Czigler et al., 2006; Lieder et al., 2013; Ostwald et al., 2012; Stefanics et al., 2014; Naeije et al., 2018). Nevertheless, given that these findings span audition, vision, and somatosensation, it is suggested for these principles to apply across modalities. Here we implemented an EEG roving paradigm with simultaneous auditory, somatosensory, and visual stimulus presentation with cross-modal dependencies underlying stimulus-sequence generation.

We found uni-modal mismatch responses between 100-200ms post-stimulus for all three modalities. Using source localization, the responses were found to originate from the respective sensory cortices and shared common frontal sources. A later, frontally-generated mismatch response (300-350ms) appeared to encode cross-modal mismatch information by displaying a sensitivity to stimulus predictability conditional on multi-modal stimulus identities. As per the results of Study 1, we hypothesized a Bayesian learner as a Categorical-Dirichlet model estimating uni-modal transition probabilities (uni-modal model) or additionally uni-modal alternation probabilities conditional on the stimulus identities of other modalities (uni- and cross-modal model), which generated either predictive surprise, Bayesian surprise, or confidence-corrected surprise. Given that the initial analyses revealed patterns that may also have been generated by a linear change-detection process which counts stimulus repetitions, we compared such a model of single-trial responses to the family of Bayesian learners. Bayesian model comparisons preferred the family widely across the spatiotemporal domain, suggesting that probabilistic inference-based accounts are more apt models of neural signals on a roving paradigm than a traditional change-detection view. Next, the Bayesian learner modeling cross-modal dependencies was found to outperform pure uni-modal learning at some electrodes for both the early ($\sim$125ms) and later ($\sim$330ms) mismatch windows. Finally, surprise functions of Bayesian learners were compared as models of EEG signals. Moderate evidence was found for confidence-corrected surprise encoding across many electrodes for the early time window. Meanwhile, the late time window showed minor indications for Bayesian surprise. In conclusion, models based on probabilistic learning were found to fit the data better than a change-detection account. Comparisons of surprise computation per se were not conclusive, yet indicated a similar pattern found in Study 1, with confidence-corrected surprise preceding Bayesian surprise..

2.4 Study 3

Gijsen, S., Grundei, M. & Blankenburg, F. Active inference and the two-step task. Scientific Reports 12, 17682 (2022).

The active inference framework, derived from the free energy principle, provides an integrative extension of surprise minimization to not only underlie perceptual learning but also action (Friston et al., 2006; Friston, 2012; Schwartenbeck et al., 2015; Smith et al., 2022). This probabilistic account posits information-seeking incentives to guide human decision making, corresponding to a sensitivity to information gain in the form of expected Bayesian surprise regarding environmental statistics. This can lead to directed exploration behaviour, in which action selection is actively steered toward more uncertain options in order to facilitate learning. In order to investigate these ideas empirically, we exploit the two-step task, which requires the traversal of two stages via binary action selection so as to maximize reward, with reward probabilities drifting over time (Daw et al., 2011). Despite the task having been widely investigated using traditional reinforcement learning accounts implementing scalar reward learning in model-free and model-based fashion (Voon et al., 2015; Wyckmans et al., 2019; Castro-Rodrigues et al., 2022), it has been shown for these models to not fully capture human behaviour (da Silva and Hare, 2020). We contrast such an account with probabilistic surprise-based learning and active inference in its ability to describe human action selection strategies. To this end, we combined computational modeling and model-agnostic regression analyses of influences on action selection using four previously published datasets (da Silva and Hare, 2020; Kool et al., 2016; Lockwood et al., 2020)

Regression analyses revealed considerable differences in action selection behaviour between datasets. By discerning the influence of knowledge about between-stage transitions on initial-stage actions, the extent to which behaviour was model-based can be inferred. Action selection was considerably more model-based on two of the datasets, with model comparisons assigning strong evidence in favour of active inference (compared to hybrid reinforcement learning) only for these datasets, while models performed similarly on the remaining datasets. The precision of prior preferences determined the reliance on information gain in action selection for active inference agents. As such, recovered subject-specific preci-

sion parameters were inspected, which followed a bimodal distribution in all datasets. This suggested that approximately half of all subjects were sensitive to the information gain in their action selection. For the datasets which were better described by active inference, correlational analyses indicated that greater information gain sensitivity was related to better model fits of the active inference model compared to hybrid reinforcement learning. Given that the two modeling frameworks differed not just with respect to action selection, but also in terms of their learning methods, this finding suggests a contribution of information gain sensitivity to model selection results.

To further investigate this topic, we leveraged an observation by da Silva and Hare (2020), who in their work noted considerable main-effects of transition type in their two-step task datasets. Common transitions, defined as those which lead participants from the initial-stage to the final-stage state in line with predictions, increased the frequency by which participants selected the other initial-stage action two and three trials later. Thus, subjects showed a tendency to periodically switch between first-stage actions independent of final-stage outcomes, with the authors suggesting this behaviour to potentially indicate directed exploration. The subject-specific main-effects of transition type were found to correlate with the precision parameters, suggesting that information gain sensitivity related to this behavioural phenomenon. To further investigate the role of this parameter, we simulated data using the set of subject parameters using stratification based on precision parameter values. Indeed, information sensitivity produced more pronounced transition effects. However, these effects were underestimated by the model, even for participants most sensitive to information gain. This suggests that probabilistic learning using active inference may improve models of human behaviour on the two-step task via its information gain incentives, but that it only captures exploration behaviour partially. A discussion of issues with the used paradigm for the current goals is followed by suggestions for future empirical validations of the active inference framework.

Chapter 3

Discussion

The aim of this thesis was to investigate the role of information-theoretic surprise quantities in learning and action by adopting a probabilistic inference framework. Multiple surprise quantities have been derived from the concept of Shannon's information, with originally an event's surprise being equated to the information it provides (Shannon, 1948). The computation uses an agent's current beliefs, which are assumed to be probabilistic in nature and therefore are reconcilable with the Bayesian brain hypothesis. Beyond pure information, it has been hypothesized that surprise signals may also also scaled with the commitment to a belief, corresponding to negative entropy under confidence-corrected surprise (Faraji et al., 2018). Alternatively, Bayesian surprise has been proposed to capture how data affects an agent and may also be interpreted as a belief updating signal (Itti and Baldi, 2009). These quantities closely relate to the ideas of a predictive brain, with more severe violations of probabilistic predictions resulting in greater levels of surprise. Further, it has been posited for the process of surprise minimization to achieve approximate Bayesian inference and to underlie perception, learning, and action. Despite the distinct aforementioned hypotheses about surprise, their role in human perceptual learning remains unsettled as direct comparisons are lacking, particularly for somatosensation. And while prominent frameworks such as active inference posit an important role for expected surprise of future outcomes to guide human action with important implications for exploration behaviour, empirical validation remains lacking. We contribute to the understanding of surprise across learning and action by adopting the roving-stimulus paradigm and the two-step task respectively.

Specifically, the roving-stimulus paradigm was used to induce mismatch computations during somatosensory sequence learning while recording neural dynamics using EEG. We found multiple mismatch responses across time and electrode-space with differing dependencies on the short-term past of the stimulation sequence. Using single-trial modeling, neural signals were best explained by a non-hierarchical Bayesian learning model integrating information locally to estimate (first-order) stimulus transition probabilities. Spatiotemporally distinct signatures of surprise computation were observed, with evidence for confidence-

corrected surprise originating in S2 preceding indications of Bayesian surprise dynamics in S1. In an extension, we used a tri-modal roving paradigm and replicated the suitability of transition-based probabilistic inference with confidence-corrected and Bayesian surprise signatures. Using an information-theoretic description of action selection based on active inference, we further investigated whether human behaviour on the two-step task was sensitive to expected surprise of future observations in a comparison with a traditional RL approach. Although multiple datasets did not strongly favour either model, datasets which displayed more model-based behaviour saw considerably better performance of the active inference model. The sensitivity to Bayesian surprise, amounting to an information gain incentive, appeared to contribute to better model fits, yet did not fully capture the observed exploration behaviour. In sum, we found evidence for signatures of sensitivity to surprise across the domains of learning and action. First, signatures of surprise computation regarding current observations were observed in EEG signals during perceptual learning which were informative of the underlying generative model. Second, the inclusion of expected surprise of future observations in action selection indicates a promising path toward better models of human decision behaviour.

3.1 Perceptual learning as probabilistic inference

We first provide a succinct and high-level description of the dynamics underlying the Bayesian learning models to aid the consequent interpretation and discussion of results. For brevity, we focus here on learning a single probability $\theta \in [0, 1]$ associated with observing one of two possible observations $y_t \sim \{0, 1\}$ at time t, corresponding to a roving-paradigm with only two stimuli or the outcome of a final-stage action on the two-step task. This corresponds to a Bernoulli trial with $p(y_t) = \theta^{y_t}(1-\theta)^{1-y_t}$. Modeling the prior as Beta-distributed provides a closed-form expression for the posterior and allows us to represent uncertainty about θ. The Beta-distribution has the additional benefit of easily interpretable concentration parameters, with α and β keeping count of the amount of $y = 0$ and $y = 1$ samples respectively. Note that the Dirichlet-Categorical (DC) model described in the introduction is a generalization of this Beta-Bernoulli model. Although these models are likely significantly simpler than those employed by the brain, they displays important characteristics of Bayesian learning and are well understood (Griffiths and Ghahramani, 2011).

Observations are integrated with prior beliefs, yielding a posterior estimate of θ, which in turn is carried forward and forms the new prior belief for the next trial. By initializing the Beta-distribution as $\mathcal{B}e(\theta; \alpha = 1, \beta = 1)$, equal probability to all values of θ is ascribed with a mean probability estimate of $E[\theta] = \frac{\alpha}{\alpha+\beta} = 0.5$. In Figure 2A, such a belief over θ is updated after observing $y = 1$ and $y = 0$ sequentially. Afterwards, the mean is still 0.5, but the precision of the distribution has increased: greater probability is assigned to regions around $\theta = 0.5$ than at the extremes. In other words, $\theta = 0.5$ has become a much more likely hypothesis than $\theta = 0.99$. Over time, iterative Bayesian learning will converge to the true θ with increasingly greater confidence as uncertainty is reduced.

When this process is repeated and prior beliefs come to encode more information about θ, the effect of further observations is diminished. In Figure 2B, again $y = 0$ and $y = 1$ are observed respectively, but only after 10 occurrences of both event types are already integrated. As a result, the two additional samples have a comparatively minor effect on beliefs due to precise prior knowledge. If θ is not assumed to be static across time, these dynamics are problematic because beliefs become highly inflexible and unable to adapt to changes in environmental statistics. Forgetting is a simple kinetic to ensure beliefs remain flexible, whereby information 'leaks' out over time. Effectively, information is integrated only over a limited time-horizon, which becomes shorter as forgetting becomes more severe. For a Beta-distributed prior, one solution is to shrink the concentration parameters over time (Figure 2C,E), thereby increasing the effect of new observations.

In studies 1 and 2, we applied the different surprise readout functions to the current beliefs at each time step. Predictive surprise depends on the posterior predictive distribution, for which we integrate over the possible values of θ. The quantity therefore does not directly depend on uncertainty or confidence and is only a function of the currently estimated $p(y)$. In contrast, confidence-corrected surprise additionally scales with the (negative) entropy of the distribution over θ. It can therefore be seen to slowly increase over time, as the distribution becomes more precise (Figure 2D). With forgetting, this effect is diminished or even absent, as belief entropy stays relatively high. Meanwhile, Bayesian surprise captures the change in the distribution over θ as a result of observing y_t, which is proportional to the unexpectedness of the event. This quantity will quickly decrease as each new observation has a smaller effect on beliefs, which can be mediated by leaking information out of the prior, for example through forgetting. For the hidden Markov model

(HMM), especially Bayesian surprise and confidence-corrected surprise differ due to the model's dynamic, rather than static, hidden state. However, in light of the results we here forego a comprehensive description of the HMM dynamics and refer to the appendix (Study 1).

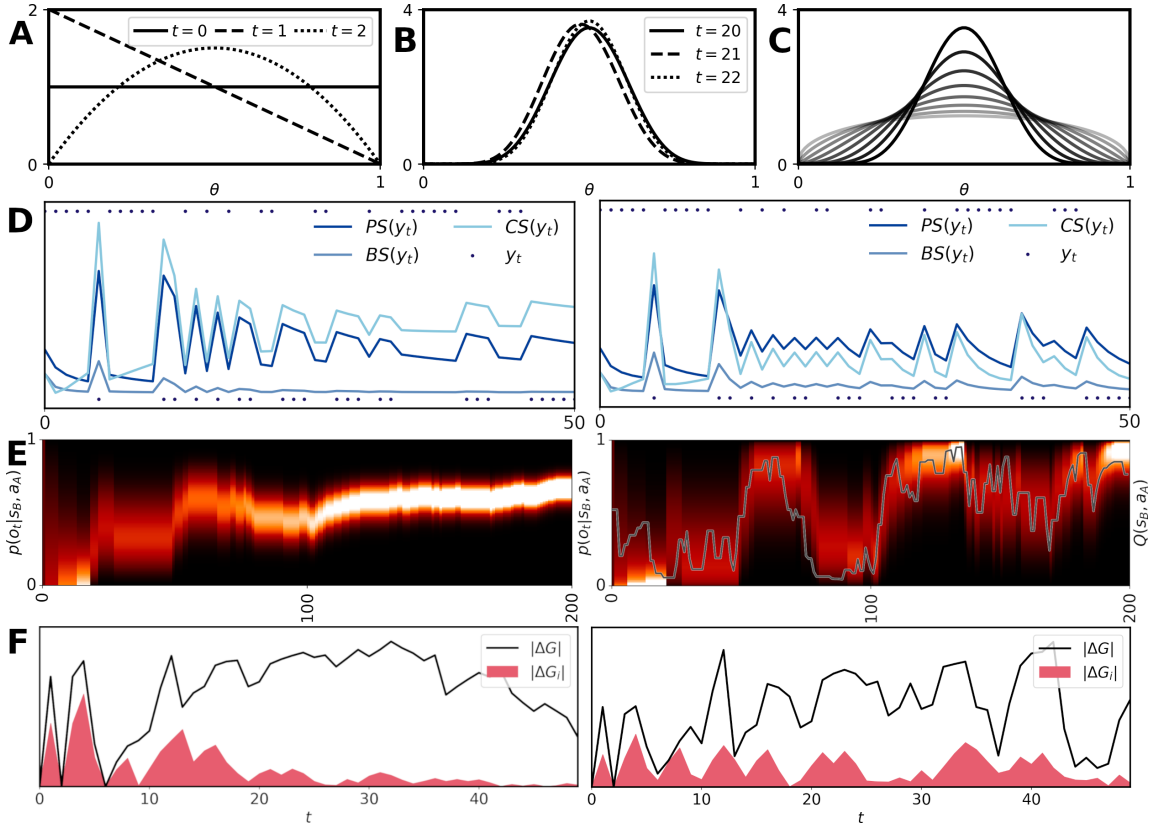


Figure 2: **Model dynamics.** A) The evolution of a uniform Beta-prior as a binary observation $y \sim \{0, 1\}$ is made for $y_{t=0} = 0$ and $y_{t=1} = 1$ sequentially. B) The same sequential observations are made given an informed Beta distribution concentrated at $\theta = 0.5$. C) A forgetting kinetic can reduce the precision of a distribution by leaking out information over time, indicated by progressively fainter lines. D) The surprise readout functions of predictive surprise $PS(y_t)$, Bayesian surprise $BS(y_t)$, and confidence-corrected surprise $CS(y_t)$ applied to the initial 50 time steps of beliefs of a Beta-Bernoulli model without (left-side; $\tau = 0$) and with (right-side; $\tau = 0.2$) forgetting. E) Heatmaps of the posterior probability of observing o_t over time estimated by a Beta-Bernoulli model, without (left-side) and with (right-side) surprise-based forgetting. Overlaid in grey is the scalar-value of the state-action pair $Q(s, a)$ as learned via prediction error reinforcement learning. F) The absolute difference in total expected free energy G between two alternative actions in black ($|\Delta G|$), with red shading indicating the proportion of this difference resulting from the intrinsic term ($|\Delta G_i|$). On the left side, the model features no forgetting and there is a continually reduced incentive for exploration. With forgetting (right-side), the incentive for uncertainty reduction stays a relevant contribution to action-selection.

Next, we turn to discuss our results. First, these surprise functions were used to infer on aspects of the underlying subject model. As hypothesized, we found evidence that EEG signals reflect surprise computation based on probabilistic inference, rather than linear change detection, matching previous findings in somatosensory learning (Ostwald et al., 2012). Recent work using uni-modal auditory and visual mismatch paradigms has similarly suggested probabilistic interpretations (Lieder et al., 2013; Stefanics et al., 2014) rather than traditional hypotheses including those based on neuronal adaptation (May et al., 1999; Jääskeläinen et al., 2004) and change detection (Schröger and Winkler, 1995). In the context of the used paradigm, we interpret this finding to support a probabilistic interpretation of general brain function. Namely, both studies featured simple tasks intended to maintain subject attention directed at the stimulation without requiring or explicitly instructing the learning of sequence statistics. We may therefore be reporting on a default system that is resorted to, with particularly the early EEG signals likely to correspond to implicit tracking of environmental statistics (Van Zuijen et al., 2006; Koelsch et al., 2016).

In the context of probabilistic inference, it remains an open question what sort of (reduced) representation of the stimulus statistics the brain estimates (Rubin et al., 2016). We present evidence for the estimation of transition probabilities, which subsumes both item frequency and alternation probability learning and is therefore the most general sequence statistic among the comparison set. This finding of (first-order) Markov dependencies in somatosensory learning constitutes an extension from previous reports in audition and vision (Meyniel et al., 2016; Maheu et al., 2019).

A further unresolved issue regarding the form of Bayesian perceptual learning in humans concerns the manner by which environmental volatility is handled (Behrens et al., 2007; Summerfield et al., 2011; Farashahi et al., 2017; Heilbron and Meyniel, 2019). Large abrupt changes to the sequence statistics were expected to elicit attention, which runs counter to the intended aim of studying non-conscious, implicit learning. We selected the two discrete states for stimulus sequence generation in study 1 with this in mind. Results of the post-experiment questionnaire consistently confirmed that regime switches were not perceived consciously. Although surprise about higher-level statistics has been reported in the literature (Gläscher et al., 2010; Iglesias et al., 2013), the similarity of regimes might explain why we did not find evidence in favour of surprise signalling based on beliefs pertaining to the regimes as modeled by an HMM. Indeed, surprise signals appeared to be based

on beliefs about stimulus transition probabilities per se as captured by the DC model. Our results thus suggest for implicit somatosensory learning to proceed non-hierarchically and to not capture changes in task statistics explicitly, although alternate schemes such as dynamic forgetting rates are left to be explored in the future. Furthermore, adaptation to changes in transition probabilities is still achieved by integrating only local, rather than global, information via exponential forgetting. Due to the irrelevance of sequence statistics to perform the instructed task, it is possible that a hierarchical model is not defaulted to due to its higher complexity and thus greater operational cost. Rather, hierarchical inference of sequence statistics may only be performed when appropriate and otherwise approximated using a flat model with forgetting. Although the degree to which sequence learning is hierarchical is unclear, local integration is commonly reported (Ostwald et al., 2012; Kolossa et al., 2013; Rubin et al., 2016; Maheu et al., 2019; Meyniel, 2020), possibly due to the dynamic character of the natural world. Recent findings suggest another possibility however, with simple models accurately approximating hierarchical inference by assuming certain forms of neural noise (Findling et al., 2021). Given resource constraints and selective pressures, such opportunities are likely to be exploited by organisms whenever possible.

3.1.1 Surprise signatures during learning

Using the model described above, we compared predictive surprise, Bayesian surprise, and confidence-corrected surprise as models for trial-by-trial EEG signal variation. In both studies, we found encoding of confidence-corrected surprise starting around 70ms poststimulus onset, followed by weaker evidence for Bayesian surprise around 140ms. In study 1, these effects were source-localized to originate in bilateral S2 and contra-lateral S1 respectively. The relatively stronger effects observed in study 1 might result from the uni-modal approach, whereas signals arising from different modalities are unlikely to synchronise spatiotemporally. This might be expected to have a lesser effect on higher-level comparisons such as family-analyses between Bayesian learning and change-detection or between uni-modal and multi-modal inference, which concern characteristics likely shared across modalities. Meanwhile, specific surprise computations would need to co-occur, or at least not interfere.

Confidence-corrected surprise signals a (probabilistic) mismatch of beliefs with the current observation and is scaled by the negative entropy of the belief distribution (Faraji et al., 2018). This quantity is therefore greater whenever expected surprise is low, corre-

sponding to a belief distribution with high precision. Confidence of (prior) beliefs has been shown to affect human perceptual learning (Meyniel et al., 2015; Meyniel and Dehaene, 2017; Meyniel, 2020) and decision making (Boldt et al., 2019; Heilbron and Meyniel, 2019). The processes of learning and confidence estimation have been seen to be tightly intertwined (Meyniel et al., 2015), which fits with a Bayesian learning account. In terms of our dipole results, sensory areas have previously been found to encode sensory uncertainty (van Bergen et al., 2015; van Bergen and Jehee, 2019; Walker et al., 2020), but under probabilistic inference may also encode uncertainty of beliefs about environmental statistics such as transition probabilities. Using fMRI, separate neural correlates of surprise and confidence of beliefs have been reported (Meyniel, 2020). Although the authors found evidence for sensory cortices encoding surprise but not confidence, BOLD signal was found to be universally affected by confidence. Given that we did not investigate for pure confidence signals, it remains unclear to what extent surprise and confidence are encoded separately.

Confidence, together with surprise, has been suggested to inform belief updating (Iglesias et al., 2013; Mathys et al., 2014; Meyniel et al., 2015). While surprise indicates the inadequacy of beliefs given observations and therefore a need to revise these beliefs, the magnitude of the update should be inversely related to the confidence associated with the beliefs. Although we did not explicitly investigate this relationship here, we suspect confidence to play a similar role in somatosensory learning, implied both in the Bayesian learning models and the encoding of confidence-corrected surprise (Faraji et al., 2018). The idea is also compatible with the temporal structure of our findings: since confidence-corrected surprise encodes both (predictive) surprise and confidence, it may inform subsequent belief updates represented by the observed Bayesian surprise dynamics, which has previously also been reported in somatosensory cortex around 140ms (Ostwald et al., 2012). An equivalent mechanism exists in many traditional reinforcement learning approaches, in which a point-estimate is updated based on the prediction error (Rescorla, 1972; Sutton and Barto, 2018). Indeed, a positive effect of prediction error on subsequent updates is a consistent finding in the literature and is considered to be mediated by the dopamine system (Montague et al., 1996; Schultz, 1998; O'Doherty et al., 2003; Seymour et al., 2004; Nassar et al., 2010; Rouhani et al., 2018; Rouhani and Niv, 2021). This substantial research field has often focused on learning in rewarding (or punishing) contexts, rather than sensory, valence-free learning. In the case of sensory learning, surprise and unsigned prediction errors are also

observed in sensory cortices as we do here (Kok et al., 2012; Meyniel, 2020), matching findings regarding mismatch responses in oddball paradigms (Den Ouden et al., 2009; Wacongne et al., 2011), including the somatosensory domain (Akatsuka et al., 2007; Ostwald et al., 2012). Although areas outside of sensory cortices have been implicated in perceptual model updating (O'Reilly et al., 2013; Meyniel, 2020), the roving paradigms we employed concerned low-level stimulus features of which the basic statistics may have been encoded in somatosensory cortex.

This insight might also help commensurate our findings with previous EEG work using information theoretic quantities as they have tended to focus on the later P300 component, with evidence reported for the encoding of predictive surprise (Mars et al., 2008; Kolossa et al., 2015; Kopp et al., 2016) as well as Bayesian surprise (Kolossa et al., 2015; Seer et al., 2016). Similarly, studies relying on delta-rule models find the P300 to signal prediction error (Nassar et al., 2019) or belief updating (Jepma et al., 2016, 2018). The multitude of studies indicating the P300 to signal these functions may question the current results indicating such signalling to occur earlier. However, the studies implementing these information-theoretic quantities often rely on explicit learning or decision making tasks, rather than studying implicit perceptual learning. As mentioned, post-hoc participant reports suggested learning to proceed mainly unconsciously, which might further support the interpretation that probabilistic inference was largely restricted to somatosensory cortices. Concurrently, by focusing on earlier sensory signals, we aimed to limit the influence of resource allocation and attentional orienting responses on analyses, as these functions have also been attributed to the P300 component (Kok, 2000; Kida et al., 2004; Chennu et al., 2013). Naturally, undocumented interactions with higher-level areas may still have occurred in study 1, which were indicated in study 2 where multi-modal dependencies were encoded primarily in (fronto-)central electrodes, potentially indicating an involvement of frontal cortex.

Recently, predictive surprise encoding was found in early auditory signals (60-130ms) over sensory areas using a sequence learning paradigm (Maheu et al., 2019). The authors interpreted these surprise signals to reflect a habituation process, based on the observation that they reflected stimulus probabilities without forgetting kinetics (Kandel and Tauc, 1965; Maheu et al., 2019). Given that we observed strong evidence in favour of transition probability estimation (rather than stimulus probability) and local (rather than global)

information integration, we suggest current results to reflect a perceptual learning process instead. Besides the difference in modalities, Maheu et al. (2019) used syllables as stimuli, potentially shifting surprise computation up the hierarchy to levels dedicated to language processing (Heilbron et al., 2022). However, a similar auditory paradigm was found to produce a predictive surprise response around 80-155ms using a transition probability model (Meyniel, 2020), but this was not compared to a stimulus probability model which might have shown it to be a similar response as reported in Maheu et al. (2019). In a visual paradigm dissociating surprise and belief updating, Visalli et al. (2021) also found early (70-140ms) predictive surprise at electrodes over sensory (occipital) cortex. Contrary to our design, stimulus identity was not orthogonal to the level of predictive surprise, leading the authors to interpret their finding as being driven by the physical properties of stimuli rather than informational content. In short, we are not the first to report early surprise signals, but provide evidence for a novel interpretation in terms of the encoding of surprise. Future somatosensory research could benefit from experimentally dissociating surprise and confidence of beliefs to clarify the nature of our observed sensory signals.

Overall, as direct comparisons between measures of surprise and model updating are rare, especially in somatosensory EEG data, it is possible that previously reported sensory error signals may have been of either nature. The currently presented results highlight the possibility that early somatosensory activity corresponds to both surprise signalling and model adjustment in a temporally coherent manner. This speaks to the value of information-theoretic quantities as descriptions of neural signals during perceptual learning.

3.2 Surprise computation for action selection

Study 3 investigated the role of expected future surprise in human behaviour by comparing a hybrid reinforcement learning model to a probabilistic model relying on active inference for action selection. Figure 2E includes a high-level comparison of the learning models for learning the probabilities of action outcomes, such as in the final-stage of the two-step task. The active inference agent relies on a probability distribution over the outcome probability, which becomes inflexible over time. Meanwhile, if a forgetting kinetic is included, for example by leaking prior information proportional to predictive surprise, the beliefs can be flexibly adapted to drifting task probabilities. In the latter case, the mean of the belief distribution can closely correspond to a scalar Q-value learned by the reinforcement learning

agent. While under RL the attractiveness of an action only depends on the outcome-dependent Q-value, active inference agents are sensitive to the uncertainty of beliefs as well as the expected outcome. The contribution of the intrinsic or information-gain term to the total expected free energy over time is displayed in Figure 2F. Although the impact of this term starts off high, the assumption of a stationary environment tends to reduce the term's relevance over time as uncertainty is resolved, resulting in prior preference realization as the primary determinant of behaviour, which can be approximated by prediction error-learning in RL. Forgetting presupposes a dynamic environment, which implies a lower-bound on belief entropy and promotes continuous information-gathering incentives. Environments with drifting statistics aid in the comparison of these two strategies by calling for forgetting kinetics (be it naively or via a surprise-based mechanism), leading to a sustained difference between models.

Questions regarding the participants' task model precede the manner in which participants learn task statistics and select actions. This results from their hierarchical relationship and is likely to be even more important here than when studying perceptual learning. An important aspect concerns whether participants rely on a task model at all, or whether they resort to model-free control instead. In their two experiments, da Silva and Hare (2020) set out specifically to provide intuitive explanations for all aspects of the two-step task. It is likely that this played an important role for the substantially increased model-based performance on these tasks compared to the original experiment by Daw et al. (2011) and later work (Kool et al., 2016; Lockwood et al., 2020), which we replicated to show only minor levels of model-based influence. It has been demonstrated that humans construct different internal models depending on the context and instructions (Green et al., 2010). This was recently shown convincingly for the two-step task specifically (Castro-Rodrigues et al., 2022), with information about the task structure resulting in considerably more participants resorting to model-based inference. Strikingly, this increase was greater than what was observed due to experience alone, despite the relatively low complexity of the task.

The sensitivity to expected Bayesian surprise in active inference, forming a crucial difference between the two compared model types, rests on model-based behaviour. As a result, the ability to differentiate between models may be reduced for participants who are not model-based or are not pursuing information gain. The models are also differentiated as a result of the active inference model assuming a probabilistic learning rule. However,

the task environment was not optimized to infer on such contributions to relative model fits. It is possible that a changepoint paradigm, in which statistics are occasionally resampled rather than drift continuously, may have aided in model differentiation. As discrete changepoints call for a temporary, substantial increase in belief uncertainty, they may have enabled another manner to inspect information gain sensitivity. Additionally, the consideration of different forms of model-free learning is an interesting extension to the current investigation, for which active inference offers intriguing options. For example, it has been described how agents may dynamically learn habits over time by adjusting the prior belief distribution over which policies will be selected (Friston et al., 2016). Such extensions may improve model-comparison analyses by enabling active inference to better fit behaviour of participants mainly engaging in model-free strategies. Due to behaviour being significantly better described by active inference only on two datasets (the 'Magic Carpet' and 'Spaceship' tasks), the upcoming section will focus on analyses using these experiments specifically unless stated otherwise.

Active inference prescribes a sensitivity to the expected Bayesian surprise associated with an action, corresponding to the information gain. In the two-step task specifically, a pursuit of information about the final-stage outcome probabilities may drive behaviour in addition to obtaining preferred outcomes. The sensitivity to information gain was determined by the precision of the prior preference distribution. We found that this precision parameter was correlated to relative model fits and regression-based transition effects. These relationships as well as simulation analyses indicated that expected information gain under active inference contributed to the model selection results. It further suggested potential for such an information incentive to underlie behaviour which the hybrid reinforcement learning approach is unable to generate, as this model relied solely on random exploration.

A sensitivity to expected Bayesian surprise in human behaviour in addition to random exploration would be in general agreement with previous work on the exploration-exploitation dilemma (Wilson et al., 2014; Gershman, 2018). The current implementation is based on information theory and active inference, which provides an intuitive framework to consider the trade-off between realizing preferences and learning by quantifying the amount of information potential actions are expected to provide. This approach has seen considerable support from the field of visual search, where eye movements have been successfully described as a process resolving uncertainty by considering the information gain associated

with visual locations (Itti and Baldi, 2009; Yang et al., 2016; Mirza et al., 2018). In more explicit decision making settings, expected information gain has been shown to accurately model human exploration behaviour (Nelson, 2005; Markant and Gureckis, 2012; Nelson et al., 2014; Tsividis et al., 2014). Although this information-based literature has considered costs of exploration Fu and Gray (2006), our study assesses the value of information in comparison to (reward) utility. Similar evidence for a sensitivity to expected information gain has been found in a bandit task by manipulating the availability of reward-information (Horvath et al., 2021), yet direct comparisons with alternative directed exploration strategies are rare for such contexts. And although theoretical and simulation-based arguments have been provided for the use of expected information gain to guide exploration (Little and Sommer, 2013), there are potential limitations. The computation of this quantity is infeasible for large state- or action-spaces and therefore approximations have been suggested, either in terms of the quantity itself (Marković et al., 2021), by imposing additional heuristics that constrain the amount of options considered (Smith et al., 2022), or using function approximators such as artificial neural networks (Ueltzhöffer, 2018; Fountas et al., 2020). Given that this is a novel research pursuit, it is unclear how efficient and accurate such approximations may get. The extent to which approximations by the brain are biased may relate to the observation made here that active inference only partially captured regression-based transition effects in the two-step task. An alternate explanation for this imperfect fit concerns the possibility for participants to use a different learning model to estimate outcome probabilities, as the expected information gain depends on currently held beliefs. This is, however, expected to only have marginal relevance due to the relative simplicity of the task limiting the divergence between different learning strategies.

It is also possible that minimizing expected free energy is not the best available description of directed exploration in humans. In the current two-step task context, this quantity may be thought of as including an information bonus to certain actions which makes them more attractive in proportion to the amount of potential information they can provide. Rather than the expected information gain, previous work has tested alternative bonuses to promote directed exploration while obtaining reward. These include a fixed 'all-or-nothing' bonus added to the least-explored option (Wilson et al., 2014), a bonus based on the variance of the current belief distribution (Daw et al., 2006; Frank et al., 2009), novelty (Krebs et al., 2009), or a penalty scaling linearly with the amount of times an action has been

visited (Dezza et al., 2017). The inconsistent results of this literature on whether action selection is biased toward uncertainty (Wilson et al., 2021) could be argued to relate to the use of such varied modeling approaches. However, the aforementioned information bonuses are expected to function relatively similar in simple tasks and inconsistencies may rather be due to the experimental context. After all, it appears that task instructions are important for determining the extent to which behaviour is model-based on the two-step task, which both directly (via model-derived information bonuses) and indirectly (via the amount of dedicated cognitive effort (Kool et al., 2016)) could impact directed exploration strategies. Further, not all of the exploration alternatives are formally derived from a framework, but can appear as ad-hoc heuristics instead. Meanwhile, the currently explored ideas are, via information theory, linked to the literature on active learning and curiosity and enable the investigation of a potentially unifying view of surprise minimization.

The field of curiosity is concerned with the endogenous 'question-and-answer' strategies humans and other animals appear to employ to learn about the world (Gottlieb et al., 2013; Gottlieb and Oudeyer, 2018). It suggests that incentives for information may be greatly beneficial in confronting large, complex environments with sparse rewards; key characteristics which reinforcement learning has grappled with since its conception (Sutton and Barto, 2018). In response, the field has explored intrinsic rewards to move beyond reward maximization, although these often take the form of handcrafted heuristics or inductive biases (Mohamed and Jimenez Rezende, 2015; Pathak et al., 2017), similar to the ad-hoc information bonuses described above. Meanwhile, curiosity natively emerges out of surprise minimization as implemented in active inference agents, describing behaviour that can naturally shift between exploration and exploitation. This has implications for environments without explicit rewards, where active inference has been suggested to continue to learn and display sensible behaviour (Friston et al., 2016; Sajid et al., 2021). Importantly, humans appear to effortlessly move between contexts with and without rewards, with potential implications for child development (Gopnik, 2012; Begus et al., 2016). Although active inference thus offers a compelling account with a potential for theoretical unification, it is rather modern RL which has achieved impressive performance in large state- or action-spaces such as in robotics (Zhu et al., 2020; Hua et al., 2021) and games (Berner et al., 2019; Schrittwieser et al., 2020). Moving forward, a symbiosis of the fields by combining engineering break-

throughs made in (Bayesian) RL with design concepts from active inference may continue to provide fertile ground for understanding action selection algorithms in the brain.

3.3 Role of surprise and uncertainty reduction across functional domains

In studies 1 and 2, analyses of EEG signals provided evidence for probabilistic inference in a perceptual learning paradigm and for information-theoretic surprise signatures in neural data. A potential role for such quantities in human action selection was investigated in study 3 using the active inference framework, with implications for a solution to the exploration-exploitation dilemma via expected information gain. Generally, our findings are situated within the theoretical proposition that surprise minimization is a core process underlying brain function spanning perception, learning, and action. This perspective implies both a widespread involvement of surprise computation across neural systems as well as a sensitivity to surprise of future outcomes, for which we here discuss different strands of available evidence. Regarding the role of surprise in learning, much of the reinforcement learning research is relevant as critical features are shared. For example, learning using delta-rules can approximate the mean of probability distributions, even for non-stationary environments (Wilson et al., 2013; Gershman and Niv, 2015). Further, unsigned prediction errors based on point-estimates tend to be highly correlated with predictive surprise (Sedley et al., 2016). The literature on surprise and belief updating may take a fully probabilistic approach or a traditional scalar-based reinforcement learning approach but in the absence of careful experimentation, results are unlikely to allow for a dissociation between these approaches.

Despite lacking theoretical specificity, the large body of literature using prediction errors and delta-rule learning is nevertheless incredibly valuable for the current work as it has broadly established the explanatory power of prediction and learning based on resulting errors. Here we leverage this literature and combine it with work which takes a probabilistic perspective insofar as the discussed mechanisms do not require either a probabilistic or scalar-based approach. First, there is considerable evidence that surprise interacts with memory systems (Sinclair and Barense, 2018; Rouhani and Niv, 2021), with surprise signals correlating with activity in memory-related temporal areas (Loued-Khenissi and Preuschoff, 2020) and the hippocampus (Rouhani and Niv, 2021). The memory-enhancing effect of surprise is suspected to be mediated by a different system than reward learning (Rouhani

et al., 2018). It has been proposed that a contribution of surprise to memory proceeds by structuring events into latent clusters, with high levels of surprise suggesting the need for memory formation rather than modification due to the incompatibility of new information with current beliefs (Gershman et al., 2017). This relationship appears related to the role of surprise in allocating attention and orienting responses by way of determining what in the environment is worth the opportunity cost of having resources deployed towards (Itti and Baldi, 2009; Chennu et al., 2013). Further, surprise has been found to effect pupil size, an effect which appears to be related to the arousal system and to be mediated by the locus coeruleus (Preuschoff et al., 2011; Joshi and Gold, 2020). Pupil metrics have even been shown to be informative of trial-by-trail learning (Nassar et al., 2012). These findings have been synthesized in an information theoretic framework suggesting various pupil dynamics to correspond to information processing (Zénon, 2019). These examples serve to indicate the potential widespread applications of surprise to brain function, indicating it may contribute to important systems beyond learning per se.

Nevertheless, research has mainly focused on the role of surprise in learning from outcomes (Meyniel et al., 2016), states (Gläscher et al., 2010), and rewards (Daw et al., 2011). It has been suggested for adaptation to result from surprise exerting a neuromodulatory effect on plasticity (Gerstner et al., 2018; Barry and Gerstner, 2022), although the exact implementation remains unverified. Beyond the previously described relationship between surprise and belief updating, more sophisticated functions have been investigated. This work builds on the well-documented phenomenon that not only surprise but also sensory uncertainty is encoded neurally (Strange et al., 2005; Bestmann et al., 2008; Bach and Dolan, 2012; McGuire et al., 2014; van Bergen and Jehee, 2019) and that this corresponds to subjectively reported confidence (Geurts et al., 2022) and affects action selection (Bestmann et al., 2008; van Bergen and Jehee, 2019). In more cognitive settings, error signals have been shown to modulate learning in a context dependent manner. Similar to the question of memory segmentation, certain incoming data requires a fundamentally different response. During learning, highly surprising events may indicate a significant change in the environment, thereby signalling a need to considerable alter beliefs. In contrast, a rare outlier event uninformative of underlying dynamics should be ignored, as adaptation to it would effectively be modeling noise or overfitting. Using such experimental conditions, Nassar et al. (2019) described the EEG P300 as a surprise signal that predicted greater learning

in the change-point condition and reduced learning in the outlier condition. This is in line with modeling results suggesting that humans use surprise to perform outlier detection and determine when not to learn (Liakoni et al., 2022), although this would require this characteristic of the environment itself to be learned or taught. Similarly, it has recently been shown that quantities based on predictive surprise as well as confidence-corrected surprise can be used to efficiently infer to what degree beliefs should be updated in environments with possible discrete changepoints (Faraji et al., 2018; Liakoni et al., 2021). This recent work connects to the concepts of ambiguity and risk from decision-making literature (Ellsberg, 1961). Ambiguity refers to uncertainty about the environment that is reducible by collecting information, while risk is irreducible uncertainty. It has been shown for humans to be sensitive to the reducibility of uncertainty in their BOLD response using fMRI (Huettel et al., 2006) and behaviour, for example by investigating which surprising events participants do or do not use to update beliefs (Kobayashi and Hsu, 2017). This implies not only a representation of belief uncertainty, but even beliefs about uncertainty, which can be described by using hierarchical probabilistic inference (e.g. Bach et al. (2011); Boundy-Singer et al. (2022)). The apparent consideration of such higher-order concepts in human behaviour indicates that scalar-based reinforcement learning accounts are likely insufficient given the omnipresence of uncertainty.

The phenomenon that the functional role of surprise is mediated by uncertainty highlights that surprise itself is only part of the story. These examples indicate that uncertainty can not only guide action selection by biasing behaviour for maximizing information, it can also guide belief updating by considering factors in the current context such as expected information and the reducibility of uncertainty. As discussed, if the brain is described to minimize informational surprise, a consideration of uncertainty is assumed under active inference, which is operationalized as entropy or expected information. Indeed, our findings suggest a sensitivity of perceptual surprise and model-based behaviour to the entropy of beliefs. Interestingly, one way in which surprise and uncertainty appear intertwined are their overlapping neural systems, together with reward processing. Despite midbrain dopaminergic neurons initially becoming almost synonymous with signalling reward prediction errors (Schultz, 1998), they have been shown in rodents to be required for value-neutral learning (Sharpe et al., 2017) and to respond to sensory features (Takahashi et al., 2017). Likewise, using fMRI in humans these neurons have been suggested to respond to the sensory identity

of rewards and the content of sensory prediction errors (Howard and Kahnt, 2018; Stalnaker et al., 2019). Strinkingly, midbrain dopamine neurons have even been shown to signal the expectation of information (Bromberg-Martin and Hikosaka, 2009) and uncertainty (Fiorillo et al., 2003). These observations have lead to the hypothesis that the dopamine system might prescribe value to information itself (Bromberg-Martin et al., 2010; Bromberg-Martin and Hikosaka, 2011; Kidd and Hayden, 2015), with important implications for exploration and curiosity. This in turn may relate to findings which indicate surprise may be pleasurable in activities such as sports viewing (Antony et al., 2021) and listening to music, where enjoyment depended on an interaction between predictive surprise and belief entropy (Cheung et al., 2019). In the active inference literature, the aforementioned dopamine findings are thought to result because dopamine may encode precision or confidence about potential actions (Friston et al., 2014). Specifically, using a biologically plausible variational update scheme for approximate Bayesian inference (Friston et al., 2013), changes in precision can be shown to correspond to changes in expected value (i.e. reward in traditional conditioning paradigms). This means that observations which increase the value of an action also increase precision. Dopaminergic findings on reward prediction error as well as information-based processing may both result from precision updating instead (Friston et al., 2015).

Such an view on the centrality of information may, in a relatively natural manner, be extended to other difficult questions in neuroscience. Zenon et al. (2019) consider the costs of cognition in terms of the amount of information needed to update beliefs and propose this quantity corresponds to subjective effort. The prior preferences under active inference are expectations, which action serves to fulfill. While lower hierarchical levels aim to conform to the environment, the highest levels encode long-term goals and motivations with high precision selected for by evolution. This means that surprise minimization aims to realize these states, lest our expectations are violated. Additionally, information theory tells us that very precise distributions, such as the expectation to feel satiated, are very costly to change. Given this prohibitive cost, a surprising state such as being hungry is therefore suggested to be resolved by eating, rather than changing expectations. This framework on information-based costs has also been shown to map well to the literature on task switching and paradigms such as the Stroop task and random dot motion tasks (Cooper et al., 2015; Zenon et al., 2019). In particular, the application of an information perspective to reaction times has a long history, with reaction times depending linearly on

the entropy of the response choice (Hick, 1952; Hyman, 1953). By dissociating predictive surprise and Bayesian surprise, recent results suggest that these interact to increase reaction times, indicating that not only belief updating (as suggested by Zenon and colleagues) but also surprise may be important (Visalli et al., 2019, 2021). However, although participants may be informed that certain observations are task-irrelevant (and thus should not be used for belief updating), this does not guarantee lower perceptual levels which are not task-sensitive to adhere to these instructions. A complete dissociation between surprise and belief updating is therefore difficult to ensure and continues to constitute an issue, especially for (lower-level) perceptual learning research. Nevertheless, the assumption of an inferring brain, with a limited informational capacity and effort to be a function of informational encoding, appears a suitable model for a variety of phenomena in line with a bounded rationality process (Tishby et al., 2000).

3.4 Open questions, future directions, and conclusions

This thesis details the use of information theoretic surprise measures to prescribe probabilistic inference to early perceptual signals and exploration behaviour under active inference. However, various important questions remain unanswered. First, it is unclear how our results on confidence-corrected surprise as a model for early somatosensory EEG signals relates to previous work on confidence and surprise. As this quantity was only recently introduced, not much comparable empirical work has yet been published (Faraji et al., 2018). Although confidence and predictive surprise have been found to interact in sensory cortex (Cheung et al., 2019), correlates are often investigated separately (Meyniel and Dehaene, 2017; Meyniel, 2020) and it will be interesting to see further adoption of confidence-corrected surprise. Meanwhile, our work would have benefited from similar contrasts between confidence and surprise to provide a better understanding of what is driving the currently presented results and to facilitate their interpretation in light of the existing literature. Further, while we find evidence for signals of prediction violation and belief updating, we did not directly infer on the effect that confidence-corrected surprise signalling might have exerted on subsequent updating. By including this mechanism explicitly in a model, this relationship could be further clarified (please see Faraji et al. (2018) for a simulation-based example). Alternatively, a somatosensory perceptual learning task with behavioural prompts reading out beliefs might be used to explicitly measure trial-by-trial

updating. However, this could significantly alter the studied system and instead probe the P300 generating system as in similar (but not somatosensory) studies (Kolossa et al., 2015; Kopp et al., 2016; Seer et al., 2016).

From a methodological standpoint, it bears mentioning that studies 1 and 2 include large amounts of Bayesian model comparisons, without applying any corrections. Although these analyses do not constitute statistical tests per se, it would be highly beneficial for future method developments to account for high exceedance probabilities arising due to chance from a high number of model comparisons. A lack of such methodological sophistication ultimately limits the conclusively of our results. Further, these studies included simple tasks to maintain the subjects' attention on the stimulation, which required judgements about the identity of stimuli. As such, it is possible that EEG signals of implicit learning are contaminated by decision-making processes, although we aimed to limit this by focusing on early dynamics up to 200ms post-stimulus.

While we attempt one of the first empirical validations of expected free energy minimization in human behaviour by focusing on directed exploration, it remains unclear how alternative action selection loss functions would compare. The comparison with the standard reinforcement learning model as used in the two-step task literature was considered important and showed promise of active inference to capture additional behaviour, yet was not contrasted with other directed exploration strategies. Marković et al. (2021) provide important simulation-based evidence that active inference can outperform Bayesian reinforcement learning alternatives as a model for dynamical bandit tasks. Directly comparing models is likely to be important going forward to understand whether certain behavioural effects are uniquely explained by a model like active inference. However, as models become more similar, precise experimental manipulations are necessary to ensure models are recoverable. The two-step task is unlikely to meet these requirements.

Next, given that reinforcement learning models share key characteristics with probabilistic models and direct comparisons are often lacking or not possible due to experimental limitations, much evidence is available which is compatible with both approaches. For example, there is considerable evidence showing the representation of uncertainty both neurally (Strange et al., 2005; Bach et al., 2011; Vilares et al., 2012) and behaviourally (Heilbron and Meyniel, 2019; Boundy-Singer et al., 2022; Geurts et al., 2022), yet behavioural evidence uniquely in favour of Bayesian reinforcement learning or a free energy minimization approach

is scarce. Future empirical research might consider including direct comparisons between these frameworks and focus on important ways in which they differ. An example qualitative comparison was recently performed in simulation, utilizing dynamic environments and the omission of rewards to stress certain model characteristics, such as the ability to generate interesting behaviours despite flat prior preferences or no value function (Sajid et al., 2021). Further, a dynamic process under active inference by which model-based behaviour transitions into habitual control is something left unexplored in human behaviour. Habits have been discussed to be obtained by becoming embedded in the prior beliefs over actions through repeated execution (Friston et al., 2016) or as Bayesian model averaging over simple and more complex models (FitzGerald et al., 2014).

Although care was taken to select similar data sets, remaining heterogeneity such as considerable variation in sample sizes complicate a clear resolution of result discrepancies between data sets. Finally, analyses were likely complicated by the task choice in two ways. First, it has been shown that participants have little influence on the average obtained reward by engaging in model-based control (Kool et al., 2016). This likely explains why additional exploration behaviour did not yield more reward, which might have reduced the incentive for information gathering strategies. Second, reward and information were correlated as subjects only learned about the consequences of chosen actions. In so far as subjects pursue reward, information tends to be higher for more rewarding actions as they are sampled more often. Alternate paradigms to decouple reward and information may be fruitful extensions.

Across all studies, our employed models are likely to be simplifications compared to neural algorithms, indicated by the uncertainty about how such models may be scaled beyond small state and action spaces as commonly studied in neuroscience. In addition, many other candidate model architectures are possible. This is important to consider as model comparison only allows for valid conclusions in regard to the included models, which represent a fraction of total possibilities. However, the compared models capture key properties of different learning and action strategies, thereby being highly useful in establishing relative plausibility in a formal manner. Over time, hypothesis space may thus be explored and illuminated.

To operate efficiently in a dynamic and ambiguous world, it is thought crucial to maintain a generative model which includes uncertainty of beliefs. Our studies suggest for sur-

prise, indicating a mismatch between model and reality, and belief entropy, an informational measure of uncertainty, to be encoded in neural signals during perceptual learning and that these quantities may be used in action selection. By relying on information theory, defined in probabilistic terms, and using Bayesian learning models, these results lend credence to the idea that the brain maintains a probabilistic representation of sensory information in the form of probability distributions. These beliefs may be used for prediction, with surprise of current observations providing a key indicator for the need for model adjustment, in line with early hypotheses of learning (Rescorla, 1972; Sutton and Barto, 2018). Through a formalized treatment of probabilistic uncertainty, we report both a neural and behavioural sensitivity to uncertainty. Regarding the former, this provides evidence for more sophisticated belief-based models of perceptual learning over scalar-based accounts (Parr and Friston, 2018). With respect to the latter, the reduction of distributional uncertainty might form an epistemic incentive. This is in accordance with the interpretation that sensory information is valued as a function of its precision because imprecise information allows for a lesser reduction in uncertainty (Parr and Friston, 2017). Such a consideration of the primacy of information gathering (or uncertainty reduction) speaks to recent information-theoretic and free energy frameworks elaborating on surprise and uncertainty research. Rather than information- or novelty-based heuristics being added to schemes maximizing reward-based utility, information gain itself might be assigned utility too. This implies that both exploration and exploitation are two facets of the same objective, as assumed under active inference (Pezzulo and Friston, 2019).

We contribute first to the understanding of perceptual learning, particularly in the understudied domain of somatosensation. We investigated multiple prominent surprise functions as well as various key characteristics of the learning model in the same study. The work thereby constitutes an important improvement over the existing literature by better accounting for interactions between learning and readout functions. Further, we contribute to the empirical investigation of active inference and the dynamics of its epistemic incentives via a contrasting anaylsis with an RL approach. As a consequence, we additionally provide insight into the shortcomings of a popular RL model in capturing human behaviour on the influential two-step task. Altogether, our results are broadly in line with the hypothesis that the brain aims to minimize surprise (or free energy as a proxy) about current observations by adjusting model parameters and surprise of future outcomes by considering the expected

surprise (i.e. entropy or uncertainty) of potential actions. Despite these important advances, considerable progress is still required to understand the nature of probabilistic information processing in the brain.

BIBLIOGRAPHY

Akatsuka, K., T. Wasaka, H. Nakata, T. Kida, and R. Kakigi (2007). The effect of stimulus probability on the somatosensory mismatch field. *Experimental brain research 181*(4), 607–614.

Amado, C., P. Hermann, P. Kovács, M. Grotheer, Z. Vidnyánszky, and G. Kovács (2016). The contribution of surprise to the prediction based modulation of fmri responses. *Neuropsychologia 84*, 105–112.

Andric, M., B. Davis, and U. Hasson (2017). Visual cortex signals a mismatch between regularity of auditory and visual streams. *NeuroImage 157*, 648–659.

Antony, J. W., T. H. Hartshorne, K. Pomeroy, T. M. Gureckis, U. Hasson, S. D. McDougle, and K. A. Norman (2021). Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *Neuron 109*(2), 377–390.

Ashwood, Z. C., N. A. Roy, I. R. Stone, A. E. Urai, A. K. Churchland, A. Pouget, and J. W. Pillow (2022). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience 25*(2), 201–212.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review 61*(3), 183.

Auksztulewicz, R. and K. Friston (2016). Repetition suppression and its contextual determinants in predictive coding. *cortex 80*, 125–140.

Bach, D. R. and R. J. Dolan (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature reviews neuroscience 13*(8), 572–586.

Bach, D. R., O. Hulme, W. D. Penny, and R. J. Dolan (2011). The known unknowns: neural representation of second-order uncertainty, and ambiguity. *Journal of Neuroscience 31*(13), 4811–4820.

Baldeweg, T., A. Klugman, J. Gruzelier, and S. R. Hirsch (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophrenia research 69*(2-3), 203–217.

Barlow, H. B. et al. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication 1*(01).

Barry, M. and W. Gerstner (2022). Fast adaptation to rule switching using neuronal surprise. *bioRxiv*.

Bastos, A. M., W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, and K. J. Friston (2012). Canonical microcircuits for predictive coding. *Neuron 76*(4), 695–711.

Begus, K., T. Gliga, and V. Southgate (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proceedings of the National Academy of Sciences 113*(44), 12397–12402.

Behrens, T. E., M. W. Woolrich, M. E. Walton, and M. F. Rushworth (2007). Learning the value of information in an uncertain world. *Nature neuroscience 10*(9), 1214–1221.

Bellemare, M. G., W. Dabney, and R. Munos (2017). A distributional perspective on reinforcement learning. In *International Conference on Machine Learning*, pp. 449–458. PMLR.

Berner, C., G. Brockman, B. Chan, V. Cheung, P. Debiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse, et al. (2019). Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*.

Besle, J., A. Fort, and M.-H. Giard (2005). Is the auditory sensory memory sensitive to visual information? *Experimental Brain Research 166*(3), 337–344.

Bestmann, S., L. M. Harrison, F. Blankenburg, R. B. Mars, P. Haggard, K. J. Friston, and J. C. Rothwell (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Current Biology 18*(10), 775–780.

Binder, K., D. Heermann, L. Roelofs, A. J. Mallinckrodt, and S. McKay (1993). Monte carlo simulation in statistical physics. *Computers in Physics 7*(2), 156–157.

Bishop, C. M. and N. M. Nasrabadi (2006). *Pattern recognition and machine learning*, Volume 4. Springer.

Bland, A. R. and A. Schaefer (2012). Different varieties of uncertainty in human decision-making. *Frontiers in neuroscience 6*, 85.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association 112*(518), 859–877.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of mathematical psychology 76*, 198–211.

Boldt, A., C. Blundell, and B. De Martino (2019). Confidence modulates exploration and exploitation in value-based learning. *Neuroscience of consciousness 2019*(1), niz004.

Botvinick, M. and M. Toussaint (2012). Planning as inference. *Trends in cognitive sciences 16*(10), 485–488.

Boundy-Singer, Z. M., C. M. Ziemba, and R. L. Goris (2022). Confidence reflects a noisy decision reliability estimate. *Nature Human Behaviour*, 1–13.

Bromberg-Martin, E. S. and O. Hikosaka (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron 63*(1), 119–126.

Bromberg-Martin, E. S. and O. Hikosaka (2011). Lateral habenula neurons signal errors in the prediction of reward information. *Nature neuroscience 14*(9), 1209–1216.

Bromberg-Martin, E. S., M. Matsumoto, and O. Hikosaka (2010). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron 68*(5), 815–834.

Butler, J. S., J. J. Foxe, I. C. Fiebelkorn, M. R. Mercier, and S. Molholm (2012). Multisensory representation of frequency across audition and touch: high density electrical mapping reveals early sensory-perceptual coupling. *Journal of Neuroscience 32*(44), 15338–15344.

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

Castro-Rodrigues, P., T. Akam, I. Snorasson, M. Camacho, V. Paixão, A. Maia, J. B. Barahona-Corrêa, P. Dayan, H. B. Simpson, R. M. Costa, et al. (2022). Explicit knowledge of task structure is a primary determinant of human model-based action. *Nature Human Behaviour*, 1–16.

Chennu, S., V. Noreika, D. Gueorguiev, A. Blenkmann, S. Kochen, A. Ibánez, A. M. Owen, and T. A. Bekinschtein (2013). Expectation and attention in hierarchical auditory prediction. *Journal of Neuroscience 33*(27), 11194–11205.

Cheung, V. K., P. M. Harrison, L. Meyer, M. T. Pearce, J.-D. Haynes, and S. Koelsch (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology 29*(23), 4084–4092.

Cohen, J. D., S. M. McClure, and A. J. Yu (2007). Should i stay or should i go? how the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences 362*(1481), 933–942.

Cooper, P. S., P. M. Garrett, J. L. Rennie, and F. Karayanidis (2015). Task uncertainty can account for mixing and switch costs in task-switching. *PloS one 10*(6), e0131556.

Cowan, N., I. Winkler, W. Teder, and R. Näätänen (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (erp). *Journal of experimental psychology: Learning, Memory, and Cognition 19*(4), 909.

Czigler, I., I. Winkler, L. Pató, A. Várnagy, J. Weisz, and L. Balázs (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.

da Silva, C. F. and T. A. Hare (2020). Humans primarily use model-based inference in the two-stage task. *Nature Human Behaviour 4*(10), 1053–1066.

Dabney, W., Z. Kurth-Nelson, N. Uchida, C. K. Starkweather, D. Hassabis, R. Munos, and M. Botvinick (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature 577*(7792), 671–675.

Daw, N. D., A. C. Courville, and P. Dayan (2008). Semi-rational models of conditioning: The case of trial order. *The probabilistic mind*, 431–452.

Daw, N. D., S. J. Gershman, B. Seymour, P. Dayan, and R. J. Dolan (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron 69*(6), 1204–1215.

Daw, N. D., Y. Niv, and P. Dayan (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience 8*(12), 1704–1711.

Daw, N. D., J. P. O'doherty, P. Dayan, B. Seymour, and R. J. Dolan (2006). Cortical substrates for exploratory decisions in humans. *Nature 441*(7095), 876–879.

Dayan, P. and Y. Niv (2008). Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology 18*(2), 185–196.

Den Ouden, H. E., K. J. Friston, N. D. Daw, A. R. McIntosh, and K. E. Stephan (2009). A dual role for prediction error in associative learning. *Cerebral cortex 19*(5), 1175–1185.

Dezza, C., A. J. Yu, A. Cleeremans, and W. Alexander (2017). Learning the value of information and reward over time when solving exploration-exploitation problems. *Scientific reports 7*(1), 1–13.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, 643–669.

Eppinger, B., M. Walter, H. R. Heekeren, and S.-C. Li (2013). Of goals and habits: age-related and individual differences in goal-directed decision-making. *Frontiers in neuroscience 7*, 253.

Ernst, M. O. (2006). A bayesian view on multimodal. *Human body perception from the inside out*, 105.

Eugster, M. J., T. Ruotsalo, M. M. Spapé, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski (2016). Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific reports 6*(1), 1–10.

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

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

Feher da Silva, C. and T. A. Hare (2020). Humans primarily use model-based inference in the two-stage task. *Nature Human Behaviour 4*(10), 1053–1066.

Findling, C., N. Chopin, and E. Koechlin (2021). Imprecise neural computations as a source of adaptive behaviour in volatile environments. *Nature Human Behaviour 5*(1), 99–112.

Fiorillo, C. D., P. N. Tobler, and W. Schultz (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science 299*(5614), 1898–1902.

FitzGerald, T. H., R. J. Dolan, and K. J. Friston (2014). Model averaging, optimal inference, and habit formation. *Frontiers in human neuroscience 8*, 457.

FitzGerald, T. H., P. Schwartenbeck, M. Moutoussis, R. J. Dolan, and K. Friston (2015). Active inference, evidence accumulation, and the urn task. *Neural computation 27*(2), 306–328.

Flandin, G. and W. D. Penny (2007). Bayesian fmri data analysis with sparse spatial basis function priors. *NeuroImage 34*(3), 1108–1125.

Fountas, Z., N. Sajid, P. Mediano, and K. Friston (2020). Deep active inference agents using monte-carlo methods. *Advances in neural information processing systems 33*, 11662–11675.

Fouragnan, E., C. Retzler, and M. G. Philiastides (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fmri meta-analysis. *Human brain mapping 39*(7), 2887–2906.

Frank, M. J., B. B. Doll, J. Oas-Terpstra, and F. Moreno (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature neuroscience 12*(8), 1062–1068.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences 360*(1456), 815–836.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience 11*(2), 127–138.

Friston, K. (2012). A free energy principle for biological systems. *Entropy 14*(11), 2100–2121.

Friston, K., T. FitzGerald, F. Rigoli, P. Schwartenbeck, G. Pezzulo, et al. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews 68*, 862–879.

Friston, K., J. Kilner, and L. Harrison (2006). A free energy principle for the brain. *Journal of physiology-Paris 100*(1-3), 70–87.

Friston, K., J. Mattout, and J. Kilner (2011). Action understanding and active inference. *Biological cybernetics 104*(1), 137–160.

Friston, K., F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo (2015). Active inference and epistemic value. *Cognitive neuroscience 6*(4), 187–214.

Friston, K., P. Schwartenbeck, T. FitzGerald, M. Moutoussis, T. Behrens, and R. J. Dolan (2013). The anatomy of choice: active inference and agency. *Frontiers in human neuroscience 7*, 598.

Friston, K., P. Schwartenbeck, T. FitzGerald, M. Moutoussis, T. Behrens, and R. J. Dolan (2014). The anatomy of choice: dopamine and decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences 369*(1655), 20130481.

Fu, W.-T. and W. D. Gray (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology 52*(3), 195–242.

Gallistel, C. R., M. Krishan, Y. Liu, R. Miller, and P. E. Latham (2014). The perception of probability. *Psychological Review 121*(1), 96.

Garrison, J., B. Erdeniz, and J. Done (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews 37*(7), 1297–1310.

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition 173*, 34–42.

Gershman, S. J. (2019). What does the free energy principle tell us about the brain? *arXiv preprint arXiv:1901.07945*.

Gershman, S. J., M.-H. Monfils, K. A. Norman, and Y. Niv (2017). The computational nature of memory modification. *Elife 6*, e23763.

Gershman, S. J. and Y. Niv (2015). Novelty and inductive generalization in human reinforcement learning. *Topics in cognitive science 7*(3), 391–415.

Gerstner, W., M. Lehmann, V. Liakoni, D. Corneil, and J. Brea (2018). Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in neural circuits 12*, 53.

Geurts, L. S., J. R. Cooke, R. S. van Bergen, and J. F. Jehee (2022). Subjective confidence reflects representation of bayesian probability in cortex. *Nature Human Behaviour 6*(2), 294–305.

Girshick, A. R., M. S. Landy, and E. P. Simoncelli (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience 14*(7), 926–932.

Gläscher, J., N. Daw, P. Dayan, and J. P. O'Doherty (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron 66*(4), 585–595.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science 337*(6102), 1623–1627.

Gottlieb, J. and P.-Y. Oudeyer (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience 19*(12), 758–770.

Gottlieb, J., P.-Y. Oudeyer, M. Lopes, and A. Baranes (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences 17*(11), 585–593.

Green, C., C. Benson, D. Kersten, and P. Schrater (2010). Alterations in choice behavior by manipulations of world model. *Proceedings of the national academy of sciences 107*(37), 16401–16406.

Griffiths, T. L. and Z. Ghahramani (2011). The indian buffet process: An introduction and review. *Journal of Machine Learning Research 12*(4).

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.

Heilbron, M., K. Armeni, J.-M. Schoffelen, P. Hagoort, and F. P. De Lange (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences 119*(32), e2201968119.

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

Helmholtz, H. v. (1856). Treatise of physiological optics: Concerning the perceptions in general. *Classics in psychology*, 79–127.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of experimental psychology 4*(1), 11–26.

Hohwy, J. (2018). The predictive processing hypothesis. *The Oxford handbook of 4E cognition*, 129–145.

Horvath, L., S. Colcombe, M. Milham, S. Ray, P. Schwartenbeck, and D. Ostwald (2021). Human belief state-based exploration and exploitation in an information-selective symmetric reversal bandit task. *Computational Brain & Behavior 4*(4), 442–462.

Houthooft, R., X. Chen, Y. Duan, J. Schulman, F. De Turck, and P. Abbeel (2016). Vime: Variational information maximizing exploration. *Advances in neural information processing systems 29*.

Howard, J. D. and T. Kahnt (2018). Identity prediction errors in the human midbrain update reward-identity expectations in the orbitofrontal cortex. *Nature communications 9*(1), 1–11.

Hua, J., L. Zeng, G. Li, and Z. Ju (2021). Learning for a robot: Deep reinforcement learning, imitation learning, transfer learning. *Sensors 21*(4), 1278.

Huang, Y. and R. P. Rao (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science 2*(5), 580–593.

Huettel, S. A., C. J. Stowe, E. M. Gordon, B. T. Warner, and M. L. Platt (2006). Neural signatures of economic preferences for risk and ambiguity. *Neuron 49*(5), 765–775.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of experimental psychology 45*(3), 188.

Iglesias, S., C. Mathys, K. H. Brodersen, L. Kasper, M. Piccirelli, H. E. den Ouden, and K. E. Stephan (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron 80*(2), 519–530.

Itti, L. and P. Baldi (2009). Bayesian surprise attracts human attention. *Vision research 49*(10), 1295–1306.

Jääskeläinen, I. P., J. Ahveninen, G. Bonmassar, A. M. Dale, R. J. Ilmoniemi, S. Levänen, F.-H. Lin, P. May, J. Melcher, S. Stufflebeam, et al. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences 101*(17), 6809–6814.

Jazayeri, M. and M. N. Shadlen (2010). Temporal context calibrates interval timing. *Nature neuroscience 13*(8), 1020–1026.

Jepma, M., S. B. Brown, P. R. Murphy, S. C. Koelewijn, B. de Vries, A. M. van den Maagdenberg, and S. Nieuwenhuis (2018). Noradrenergic and cholinergic modulation of belief updating. *Journal of Cognitive Neuroscience 30*(12), 1803–1820.

Jepma, M., P. R. Murphy, M. R. Nassar, M. Rangel-Gomez, M. Meeter, and S. Nieuwenhuis (2016). Catecholaminergic regulation of learning rate in a dynamic environment. *PLoS computational biology 12*(10), e1005171.

Joshi, S. and J. I. Gold (2020). Pupil size as a window on neural substrates of cognition. *Trends in cognitive sciences 24*(6), 466–480.

Kaelbling, L. P., M. L. Littman, and A. R. Cassandra (1998). Planning and acting in partially observable stochastic domains. *Artificial intelligence 101*(1-2), 99–134.

Kandel, E. R. and L. Tauc (1965). Heterosynaptic facilitation in neurones of the abdominal ganglion of aplysia depilans. *The Journal of Physiology 181*(1), 1.

Kangassalo, L., M. Spapé, N. Ravaja, and T. Ruotsalo (2020). Information gain modulates brain activity evoked by reading. *Scientific reports 10*(1), 1–10.

Kida, T., Y. Nishihira, A. Hatta, T. Wasaka, T. Tazoe, Y. Sakajiri, H. Nakata, T. Kaneda, K. Kuroiwa, S. Akiyama, et al. (2004). Resource allocation and somatosensory p300 amplitude during dual task: effects of tracking speed and predictability of tracking direction. *Clinical Neurophysiology 115*(11), 2616–2628.

Kidd, C. and B. Y. Hayden (2015). The psychology and neuroscience of curiosity. *Neuron 88*(3), 449–460.

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

Klayman, J. and Y.-W. Ha (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review 94*(2), 211.

Klyubin, A. S., D. Polani, and C. L. Nehaniv (2004). Organization of the information flow in the perception-action loop of evolved agents. In *Proceedings. 2004 NASA/DoD Conference on Evolvable Hardware, 2004.*, pp. 177–180. IEEE.

Klyubin, A. S., D. Polani, and C. L. Nehaniv (2008). Keep your options open: An information-based driving principle for sensorimotor systems. *PloS one 3*(12), e4018.

Knill, D. C. and A. Pouget (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences 27*(12), 712–719.

Kobayashi, K. and M. Hsu (2017). Neural mechanisms of updating under reducible and irreducible uncertainty. *Journal of Neuroscience 37*(29), 6972–6982.

Koelsch, S., T. Busch, S. Jentschke, and M. Rohrmeier (2016). Under the hood of statistical learning: A statistical mmn reflects the magnitude of transitional probabilities in auditory sequences. *Scientific reports 6*(1), 1–11.

Kok, A. (2000). Age-related changes in involuntary and voluntary attention as reflected in components of the event-related potential (erp). *Biological psychology 54*(1-3), 107–143.

Kok, P., L. J. Bains, T. van Mourik, D. G. Norris, and F. P. de Lange (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology 26*(3), 371–376.

Kok, P., D. Rahnev, J. F. Jehee, H. C. Lau, and F. P. De Lange (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral cortex 22*(9), 2197–2206.

Kolossa, A., T. Fingscheidt, K. Wessel, and B. Kopp (2013). A model-based approach to trial-by-trial p300 amplitude fluctuations. *Frontiers in human neuroscience 6*, 359.

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

Kool, W., F. A. Cushman, and S. J. Gershman (2016). When does model-based control pay off? *PLoS computational biology 12*(8), e1005090.

Kopp, B., C. Seer, F. Lange, A. Kluytmans, A. Kolossa, T. Fingscheidt, and H. Hoijtink (2016). P300 amplitude variations, prior probabilities, and likelihoods: A bayesian erp study. *Cognitive, Affective, & Behavioral Neuroscience 16*(5), 911–928.

Krebs, R. M., B. H. Schott, H. Schütze, and E. Düzel (2009). The novelty exploration bonus and its attentional modulation. *Neuropsychologia 47*(11), 2272–2281.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The annals of mathematical statistics 22*(1), 79–86.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current opinion in neurobiology 11*(4), 475–480.

Laughlin, S. B., R. R. de Ruyter van Steveninck, and J. C. Anderson (1998). The metabolic cost of neural information. *Nature neuroscience 1*(1), 36–41.

Lawrence, S. J., E. Formisano, L. Muckli, and F. P. de Lange (2019). Laminar fmri: Applications for cognitive neuroscience. *Neuroimage 197*, 785–791.

Lee, D., H. Seo, and M. W. Jung (2012). Neural basis of reinforcement learning and decision making. *Annual review of neuroscience 35*, 287.

Lee, T. S. and D. Mumford (2003). Hierarchical bayesian inference in the visual cortex. *JOSA A 20*(7), 1434–1448.

Liakoni, V., M. P. Lehmann, A. Modirshanechi, J. Brea, A. Lutti, W. Gerstner, and K. Preuschoff (2022). Brain signals of a surprise-actor-critic model: Evidence for multiple learning modules in human decision making. *NeuroImage 246*, 118780.

Liakoni, V., A. Modirshanechi, W. Gerstner, and J. Brea (2021). Learning in volatile environments with the bayes factor surprise. *Neural Computation 33*(2), 269–340.

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

Lieder, F., K. E. Stephan, J. Daunizeau, M. I. Garrido, and K. J. Friston (2013). A neurocomputational model of the mismatch negativity. *PLoS computational biology 9*(11), e1003288.

Little, D. Y. and F. T. Sommer (2013). Learning and exploration in action-perception loops. *Frontiers in neural circuits 7*, 37.

Lockwood, P. L., M. C. Klein-Flügge, A. Abdurahman, and M. J. Crockett (2020). Model-free decision making is prioritized when learning to avoid harming others. *Proceedings of the National Academy of Sciences 117*(44), 27719–27730.

Loued-Khenissi, L. and K. Preuschoff (2020). Information theoretic characterization of uncertainty distinguishes surprise from accuracy signals in the brain. *Frontiers in artificial intelligence 3*, 5.

Lowet, A. S., Q. Zheng, S. Matias, J. Drugowitsch, and N. Uchida (2020). Distributional reinforcement learning in the brain. *Trends in Neurosciences 43*(12), 980–997.

Lungarella, M. and O. Sporns (2006). Mapping information flow in sensorimotor networks. *PLoS computational biology 2*(10), e144.

Maheu, M., S. Dehaene, and F. Meyniel (2019). Brain signatures of a multiscale process of sequence learning in humans. *elife 8*, e41541.

Mareschal, I., A. J. Calder, and C. W. Clifford (2013). Humans have an expectation that gaze is directed toward them. *Current Biology 23*(8), 717–721.

Markant, D. and T. Gureckis (2012). Does the utility of information influence sampling behavior? In *Proceedings of the annual meeting of the cognitive science society*, Volume 34.

Marković, D., H. Stojić, S. Schwöbel, and S. J. Kiebel (2021). An empirical evaluation of active inference in multi-armed bandits. *Neural Networks 144*, 229–246.

Mars, R. B., S. Debener, T. E. Gladwin, L. M. Harrison, P. Haggard, J. C. Rothwell, and S. Bestmann (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *Journal of Neuroscience 28*(47), 12539–12545.

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

May, P., H. Tiitinen, R. J. Ilmoniemi, G. Nyman, J. G. Taylor, and R. Näätänen (1999). Frequency change detection in human auditory cortex. *Journal of computational neuroscience 6*(2), 99–120.

McGuire, J. T., M. R. Nassar, J. I. Gold, and J. W. Kable (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron 84*(4), 870–881.

Meyniel, F. (2020). Brain dynamics for confidence-weighted learning. *PLoS computational biology 16*(6), e1007935.

Meyniel, F. and S. Dehaene (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences 114*(19), E3859–E3868.

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

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

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

Mirza, M. B., R. A. Adams, C. Mathys, and K. J. Friston (2018). Human visual exploration reduces uncertainty about the sensed world. *PloS one 13*(1), e0190429.

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

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

Mohamed, S. and D. Jimenez Rezende (2015). Variational information maximisation for intrinsically motivated reinforcement learning. *Advances in neural information processing systems 28*.

Montague, P. R., P. Dayan, and T. J. Sejnowski (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of neuroscience 16*(5), 1936–1947.

Mousavi, Z., M. M. Kiani, and H. Aghajan (2020). Brain signatures of surprise in eeg and meg data. *bioRxiv*.

Muckli, L., F. De Martino, L. Vizioli, L. S. Petro, F. W. Smith, K. Ugurbil, R. Goebel, and E. Yacoub (2015). Contextual feedback to superficial layers of v1. *Current Biology 25*(20), 2690–2695.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics 66*(3), 241–251.

Näätänen, R., A. W. Gaillard, and S. Mäntysalo (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica 42*(4), 313–329.

Näätänen, R., P. Paavilainen, T. Rinne, and K. Alho (2007). The mismatch negativity (mmn) in basic research of central auditory processing: a review. *Clinical neurophysiology 118*(12), 2544–2590.

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

Nassar, M. R., R. Bruckner, and M. J. Frank (2019). Statistical context dictates the relationship between feedback-related eeg signals and learning. *Elife 8*, e46975.

Nassar, M. R., K. M. Rumsey, R. C. Wilson, K. Parikh, B. Heasly, and J. I. Gold (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience 15*(7), 1040–1046.

Nassar, M. R., R. C. Wilson, B. Heasly, and J. I. Gold (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience 30*(37), 12366–12378.

Nelson, J. D. (2005). Finding useful questions: on bayesian diagnosticity, probability, impact, and information gain. *Psychological review 112*(4), 979.

Nelson, J. D., B. Divjak, G. Gudmundsdottir, L. F. Martignon, and B. Meder (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition 130*(1), 74–80.

Nevo, I. and I. Erev (2012). On surprise, change, and the effect of recent outcomes. *Frontiers in psychology 3*, 24.

Osband, I., C. Blundell, A. Pritzel, and B. Van Roy (2016). Deep exploration via bootstrapped dqn. *Advances in neural information processing systems 29*.

Ostwald, D., E. Kirilina, L. Starke, and F. Blankenburg (2014). A tutorial on variational bayes for latent linear stochastic time-series models. *Journal of Mathematical Psychology 60*, 1–19.

Ostwald, D., B. Spitzer, M. Guggenmos, T. T. Schmidt, S. J. Kiebel, and F. Blankenburg (2012). Evidence for neural encoding of bayesian surprise in human somatosensation. *NeuroImage 62*(1), 177–188.

Otto, A. R., C. M. Raio, A. Chiang, E. A. Phelps, and N. D. Daw (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences 110*(52), 20941–20946.

O'Doherty, J., P. Dayan, K. Friston, H. Critchley, R. Dolan, et al. (2003). Temporal difference learning model accounts for responses in human ventral striatum and orbitofrontal cortex during pavlovian appetitive learning. *Neuron 38*, 329–337.

O'Reilly, J. X., U. Schüffelgen, S. F. Cuell, T. E. Behrens, R. B. Mars, and M. F. Rushworth (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences 110*(38), E3660–E3669.

Parr, T. and K. J. Friston (2017). Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface 14*(136), 20170376.

Parr, T. and K. J. Friston (2018). The anatomy of inference: generative models and brain structure. *Frontiers in computational neuroscience*, 90.

Pathak, D., P. Agrawal, A. A. Efros, and T. Darrell (2017). Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning*, pp. 2778–2787. PMLR.

Penny, W., S. Kiebel, and K. Friston (2003). Variational bayesian inference for fmri time series. *NeuroImage 19*(3), 727–741.

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

Penny, W. D., N. J. Trujillo-Barreto, and K. J. Friston (2005). Bayesian fmri time series analysis with spatial priors. *NeuroImage 24*(2), 350–362.

Pezzulo, G. and K. J. Friston (2019). The value of uncertainty: An active inference perspective. *Behavioral and Brain Sciences 42*.

Preuschoff, K., B. M. 't Hart, and W. Einhäuser (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in neuroscience 5*, 115.

Rabiner, L. and B. Juang (1986). An introduction to hidden markov models. *ieee assp magazine 3*(1), 4–16.

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

Rescorla, R. A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, 64–99.

Riefer, P. S., R. Prior, N. Blair, G. Pavey, and B. C. Love (2017). Coherency-maximizing exploration in the supermarket. *Nature human behaviour 1*(1), 1–4.

Rigoux, L., K. E. Stephan, K. J. Friston, and J. Daunizeau (2014). Bayesian model selection for group studies—revisited. *Neuroimage 84*, 971–985.

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

Rouhani, N., K. A. Norman, and Y. Niv (2018). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition 44*(9), 1430.

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

Rummery, G. A. and M. Niranjan (1994). *On-line Q-learning using connectionist systems*, Volume 37. University of Cambridge, Department of Engineering Cambridge, UK.

Sajid, N., P. J. Ball, T. Parr, and K. J. Friston (2021). Active inference: demystified and compared. *Neural computation 33*(3), 674–712.

Sanborn, A. N. and N. Chater (2016). Bayesian brains without probabilities. *Trends in cognitive sciences 20*(12), 883–893.

Sanborn, A. N. and R. Silva (2013). Constraining bridges between levels of analysis: A computational justification for locally bayesian learning. *Journal of Mathematical Psychology 57*(3-4), 94–106.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE transactions on autonomous mental development 2*(3), 230–247.

Schrittwieser, J., I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, et al. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature 588*(7839), 604–609.

Schröger, E. and I. Winkler (1995). Presentation rate and magnitude of stimulus deviance effects on human pre-attentive change detection. *Neuroscience letters 193*(3), 185–188.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of neurophysiology 80*(1), 1–27.

Schultz, W., W. R. Stauffer, and A. Lak (2017). The phasic dopamine signal maturing: from reward via behavioural activation to formal economic utility. *Current opinion in neurobiology 43*, 139–148.

Schulz, E., R. Bhui, B. C. Love, B. Brier, M. T. Todd, and S. J. Gershman (2019). Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings of the National Academy of Sciences 116*(28), 13903–13908.

Schwartenbeck, P., T. FitzGerald, R. Dolan, and K. Friston (2013). Exploration, novelty, surprise, and free energy minimization. *Frontiers in psychology*, 710.

Schwartenbeck, P., T. H. FitzGerald, C. Mathys, R. Dolan, M. Kronbichler, and K. Friston (2015). Evidence for surprise minimization over value maximization in choice behavior. *Scientific reports 5*(1), 1–14.

Schwartenbeck, P., T. H. FitzGerald, C. Mathys, M. Kronbichler, and K. Friston (2015). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cerebral cortex 25*(10), 3434–3445.

Schwartenbeck, P., J. Passecker, T. U. Hauser, T. H. FitzGerald, M. Kronbichler, and K. J. Friston (2019). Computational mechanisms of curiosity and goal-directed exploration. *Elife 8*.

Sedley, W., P. E. Gander, S. Kumar, C. K. Kovach, H. Oya, H. Kawasaki, M. A. Howard III, and T. D. Griffiths (2016). Neural signatures of perceptual inference. *elife 5*, e11476.

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

Seymour, B., J. P. O'Doherty, P. Dayan, M. Koltzenburg, A. K. Jones, R. J. Dolan, K. J. Friston, and R. S. Frackowiak (2004). Temporal difference models describe higher-order learning in humans. *Nature 429*(6992), 664–667.

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

Sharpe, M. J., C. Y. Chang, M. A. Liu, H. M. Batchelor, L. E. Mueller, J. L. Jones, Y. Niv, and G. Schoenbaum (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience 20*(5), 735–742.

Silver, D., S. Singh, D. Precup, and R. S. Sutton (2021). Reward is enough. *Artificial Intelligence 299*, 103535.

Sinclair, A. H. and M. D. Barense (2018). Surprise and destabilize: prediction error influences episodic memory reconsolidation. *Learning & memory 25*(8), 369–381.

Siu, C. R. and K. M. Murphy (2018). The development of human visual cortex and clinical implications. *Eye and brain 10*, 25.

Smith, R., K. J. Friston, and C. J. Whyte (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology 107*, 102632.

Smith, R., N. Kirlic, J. L. Stewart, J. Touthang, R. Kuplicki, S. S. Khalsa, J. Feinstein, M. P. Paulus, and R. L. Aupperle (2021). Greater decision uncertainty characterizes a transdiagnostic patient sample during approach-avoidance conflict: a computational modelling approach. *Journal of Psychiatry and Neuroscience 46*(1), E74–E87.

Smith, R., N. Kirlic, J. L. Stewart, J. Touthang, R. Kuplicki, T. J. McDermott, S. Taylor, S. S. Khalsa, M. P. Paulus, and R. L. Aupperle (2021). Long-term stability of computational parameters during approach-avoidance conflict in a transdiagnostic psychiatric patient sample. *Scientific reports 11*(1), 1–13.

Smith, R., R. Kuplicki, J. Feinstein, K. L. Forthman, J. L. Stewart, M. P. Paulus, T. . Investigators, and S. S. Khalsa (2020). A bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *PLoS computational biology 16*(12), e1008484.

Smith, R., M. J. Ramstead, and A. Kiefer (2022). Why bayesian brains perform poorly on explicit probabilistic reasoning problems.

Smith, R., P. Schwartenbeck, J. L. Stewart, R. Kuplicki, H. Ekhtiari, M. P. Paulus, T. . Investigators, et al. (2020). Imprecise action selection in substance use disorder: Evidence for active learning impairments when solving the explore-exploit dilemma. *Drug and alcohol dependence 215*, 108208.

Smith, R., S. Taylor, J. L. Stewart, S. M. Guinjoan, M. Ironside, N. Kirlic, H. Ekhtiari, E. J. White, H. Zheng, R. Kuplicki, et al. (2022). Slower learning rates from negative outcomes in

substance use disorder over a 1-year period and their potential predictive utility. *Computational Psychiatry 6*(1).

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and cognition 112*, 92–97.

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

Stalnaker, T. A., J. D. Howard, Y. K. Takahashi, S. J. Gershman, T. Kahnt, and G. Schoenbaum (2019). Dopamine neuron ensembles signal the content of sensory prediction errors. *Elife 8*.

Stefanics, G., J. Kremláček, and I. Czigler (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in human neuroscience 8*, 666.

Stephan, K. E., W. D. Penny, J. Daunizeau, R. J. Moran, and K. J. Friston (2009). Bayesian model selection for group studies. *Neuroimage 46*(4), 1004–1017.

Sterzer, P., R. A. Adams, P. Fletcher, C. Frith, S. M. Lawrie, L. Muckli, P. Petrovic, P. Uhlhaas, M. Voss, and P. R. Corlett (2018). The predictive coding account of psychosis. *Biological psychiatry 84*(9), 634–643.

Stocker, A. A. and E. P. Simoncelli (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience 9*(4), 578–585.

Strange, B. A., A. Duggins, W. Penny, R. J. Dolan, and K. J. Friston (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Networks 18*(3), 225–230.

Summerfield, C., T. E. Behrens, and E. Koechlin (2011). Perceptual classification in a rapidly changing environment. *Neuron 71*(4), 725–736.

Sun, Y., F. Gomez, and J. Schmidhuber (2011). Planning to be surprised: Optimal bayesian exploration in dynamic environments. In *International conference on artificial general intelligence*, pp. 41–51. Springer.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement learning: An introduction*. MIT press.

Takahashi, Y. K., H. M. Batchelor, B. Liu, A. Khanna, M. Morales, and G. Schoenbaum (2017). Dopamine neurons respond to errors in the prediction of sensory features of expected rewards. *Neuron 95*(6), 1395–1405.

Tishby, N., F. C. Pereira, and W. Bialek (2000). The information bottleneck method. *arXiv preprint physics/0004057*.

Tishby, N. and D. Polani (2011). Information theory of decisions and actions. In *Perception-action cycle*, pp. 601–636. Springer.

Tsividis, P., S. Gershman, J. Tenenbaum, and L. Schulz (2014). Information selection in noisy environments with large action spaces. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Volume 36.

Ueltzhöffer, K. (2018). Deep active inference. *Biological cybernetics 112*(6), 547–573.

van Bergen, R. S. and J. F. Jehee (2019). Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *Journal of Neuroscience 39*(41), 8164–8176.

van Bergen, R. S., W. Ji Ma, M. S. Pratte, and J. F. Jehee (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature neuroscience 18*(12), 1728–1730.

Van Boxtel, J. J. and H. Lu (2013). A predictive coding perspective on autism spectrum disorders.

Van Zuijen, T. L., V. L. Simoens, P. Paavilainen, R. Näätänen, and M. Tervaniemi (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. *Journal of cognitive neuroscience 18*(8), 1292–1303.

Vilares, I., J. D. Howard, H. L. Fernandes, J. A. Gottfried, and K. P. Kording (2012). Differential representations of prior and likelihood uncertainty in the human brain. *Current Biology 22*(18), 1641–1648.

Visalli, A., M. Capizzi, E. Ambrosini, B. Kopp, and A. Vallesi (2021). Electroencephalographic correlates of temporal bayesian belief updating and surprise. *NeuroImage 231*, 117867.

Visalli, A., M. Capizzi, E. Ambrosini, I. Mazzonetto, and A. Vallesi (2019). Bayesian modeling of temporal expectations in the human brain. *Neuroimage 202*, 116097.

Voon, V., K. Baek, J. Enander, Y. Worbe, L. Morris, N. Harrison, T. Robbins, C. Rück, and N. Daw (2015). Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. *Translational psychiatry 5*(11), e670–e670.

Wacongne, C., E. Labyt, V. van Wassenhove, T. Bekinschtein, L. Naccache, and S. Dehaene (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences 108*(51), 20754–20759.

Walker, E. Y., R. J. Cotton, W. J. Ma, and A. S. Tolias (2020). A neural basis of probabilistic computation in visual cortex. *Nature Neuroscience 23*(1), 122–129.

Walsh, K. S., D. P. McGovern, A. Clark, and R. G. O'Connell (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the new York Academy of Sciences 1464*(1), 242–268.

Watanabe, E., A. Kitaoka, K. Sakamoto, M. Yasugi, and K. Tanaka (2018). Illusory motion reproduced by deep neural networks trained for prediction. *Frontiers in psychology*, 345.

Wei, X.-X. and A. A. Stocker (2015). A bayesian observer model constrained by efficient coding can explain'anti-bayesian'percepts. *Nature neuroscience 18*(10), 1509–1517.

Weilnhammer, V., H. Stuke, G. Hesselmann, P. Sterzer, and K. Schmack (2017). A predictive coding account of bistable perception-a model-based fmri study. *PLoS computational biology 13*(5), e1005536.

Wiener, N. (2019). *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press.

Wilson, R. C., E. Bonawitz, V. D. Costa, and R. B. Ebitz (2021). Balancing exploration and exploitation with information and randomization. *Current opinion in behavioral sciences 38*, 49–56.

Wilson, R. C., A. Geana, J. M. White, E. A. Ludvig, and J. D. Cohen (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General 143*(6), 2074.

Wilson, R. C., M. R. Nassar, and J. I. Gold (2013). A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS computational biology 9*(7), e1003150.

Winkler, I. and I. Czigler (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.

Wolpert, D. H. (2012). What the no free lunch theorems really mean; how to improve search algorithms. In *Santa Fe Institute*, Volume 7, pp. 1–13.

Wunderlich, K., P. Smittenaar, and R. J. Dolan (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron 75*(3), 418–424.

Wyckmans, F., A. R. Otto, M. Sebold, N. Daw, A. Bechara, M. Saeremans, C. Kornreich, A. Chatard, N. Jaafari, and X. Noël (2019). Reduced model-based decision-making in gambling disorder. *Scientific reports 9*(1), 1–10.

Yang, S. C.-H., M. Lengyel, and D. M. Wolpert (2016). Active sensing in the categorization of visual patterns. *Elife 5*, e12215.

Zénon, A. (2019). Eye pupil signals information gain. *Proceedings of the Royal Society B 286*(1911), 20191593.

Zenon, A., O. Solopchuk, and G. Pezzulo (2019). An information-theoretic perspective on the costs of cognition. *Neuropsychologia 123*, 5–18.

Zhu, H., J. Yu, A. Gupta, D. Shah, K. Hartikainen, A. Singh, V. Kumar, and S. Levine (2020). The ingredients of real-world robotic reinforcement learning. *arXiv preprint arXiv:2004.12570*.

Appendix

Original work: Study 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.

* Shared authorship

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

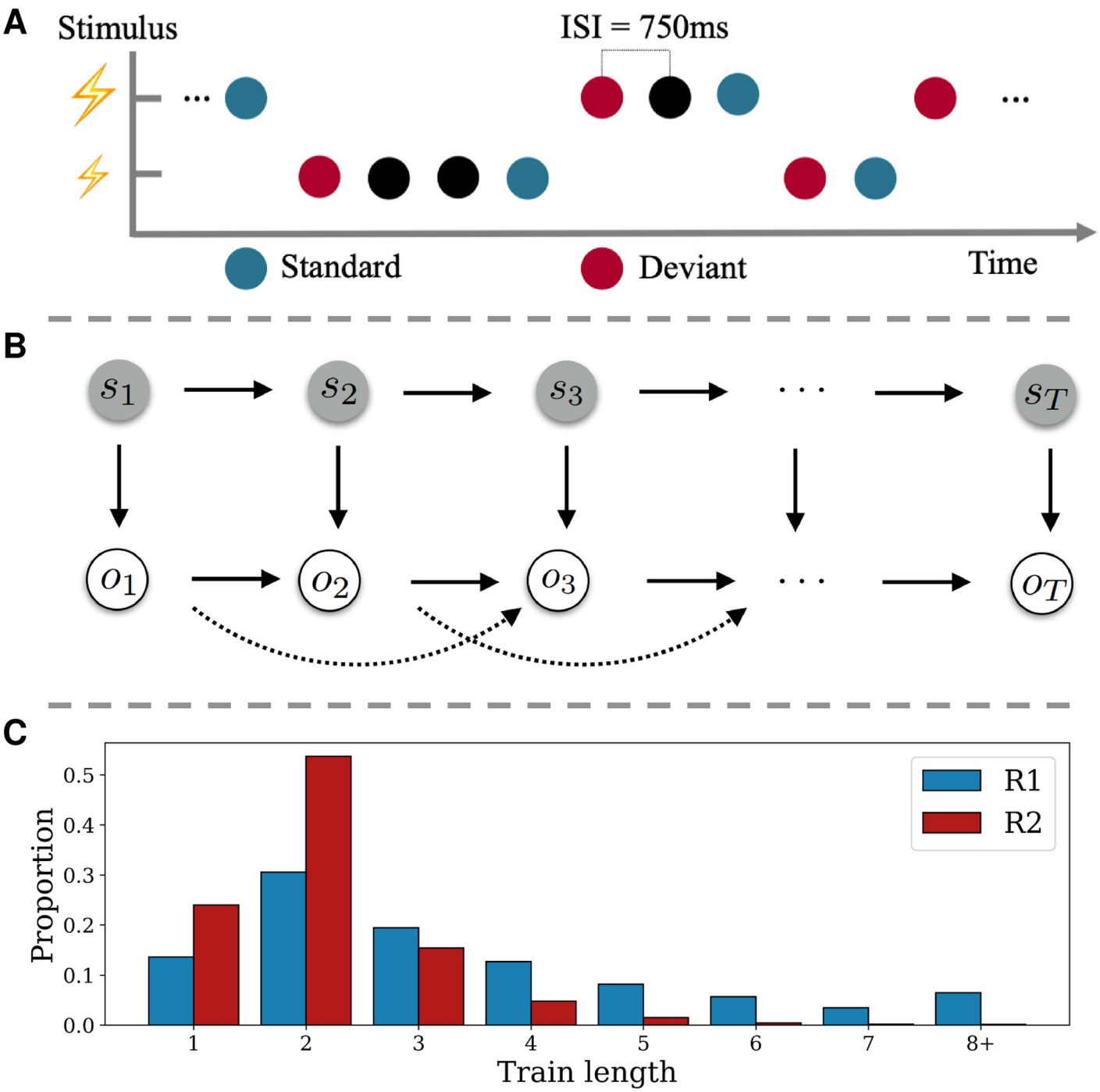


Fig 1. Experimental design and stimulus generation. A) Presentation of experimental stimuli using a roving-stimulus paradigm. Stimuli with two different intensities are presented. Their role as standard or deviant depends on their respective position within the presentation sequence. B) Graphical model 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

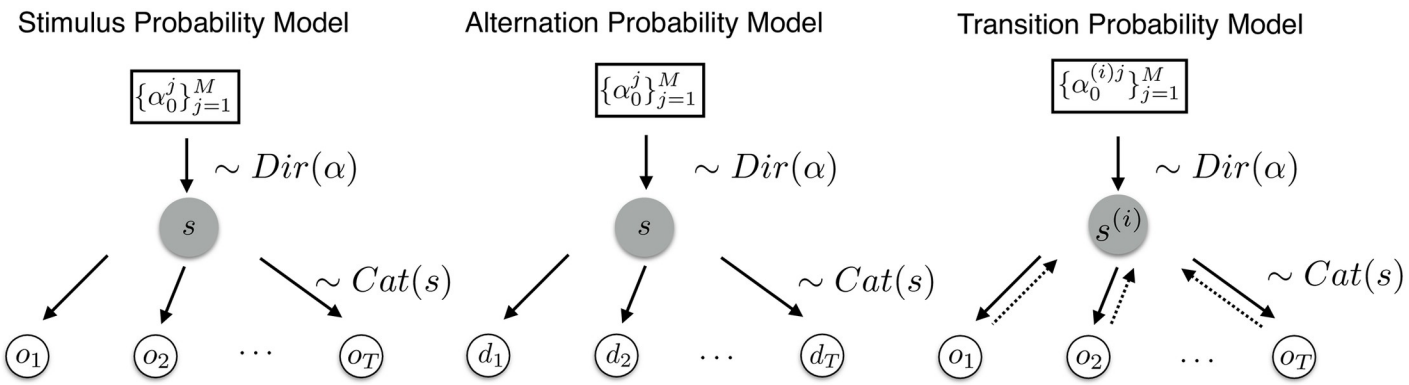


Fig 2. Dirichlet-Categorical model as a graphical model. Left: The stimulus probability model which tracks the hidden state vector determining the sampling process of the raw observations. Middle: The alternation probability model which infers the hidden state distribution based on alternations of the 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.

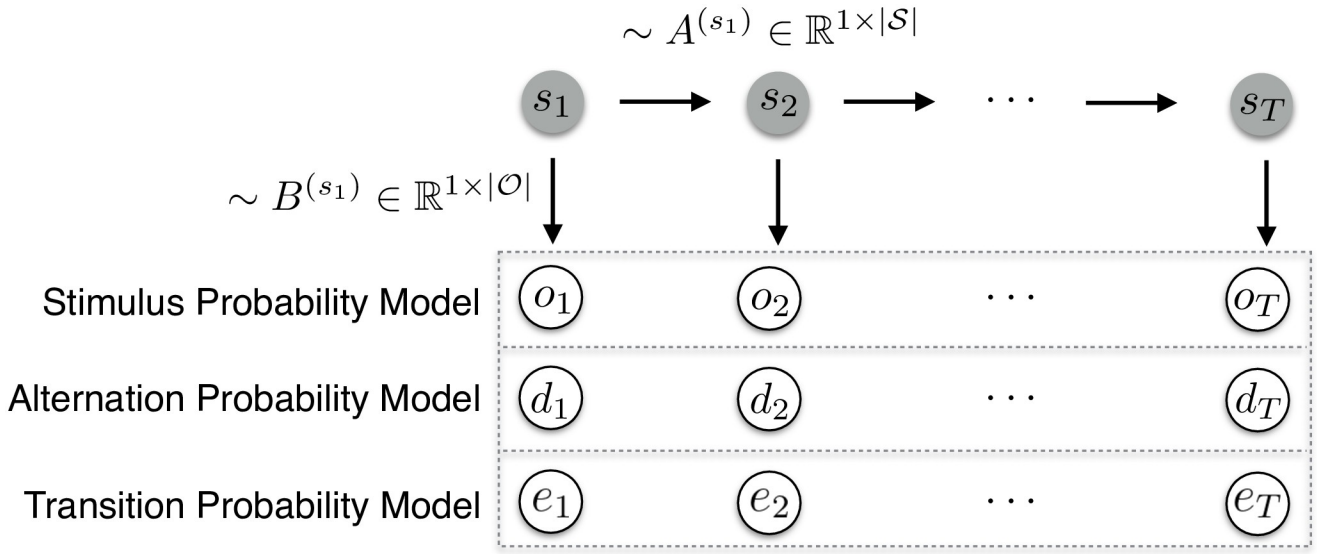


Fig 3. Hidden Markov model as a graphical model. Upper row depicts the evolution of states s_t according to the transition matrix $A^{(s_t)}$. The states emit observational data (dotted rectangle) according to the probabilities specified in stochastic matrix $B^{(s_t)}$ which depends on the type of inference. 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

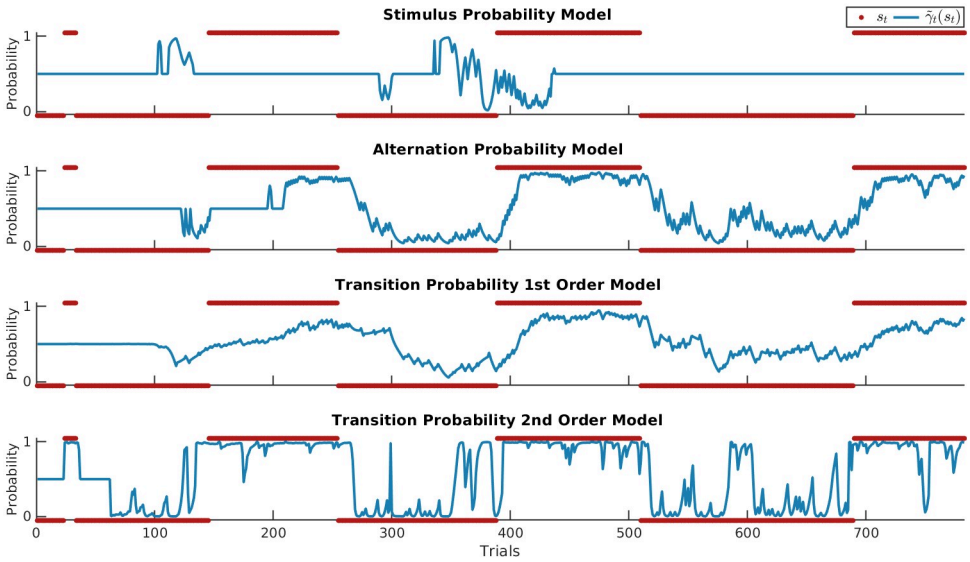


Fig 4. Posterior probabilities of the HMM. Comparison of the filtering posterior $\hat{\gamma}_t(s_t) = p(s_t|o_1, \ldots, o_t)$ of the different HMM inference models for an example sequence. The true, but unknown regimes of the data generation process are plotted in red. Note that, as the regimes were 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

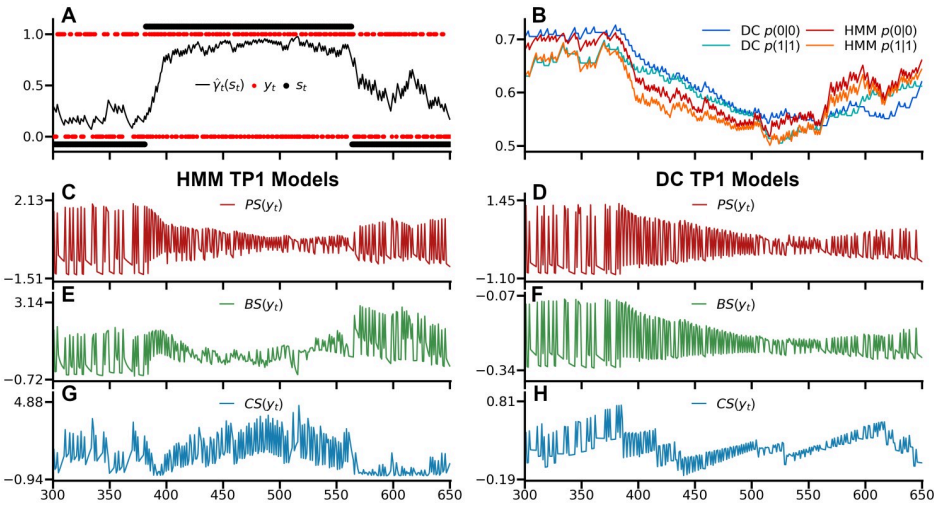


Fig 5. Surprise readouts. A) Example sequence with o_t in red, s_t in black with $s_t = 0$ for the slow-switching regime and $s_t = 1$ for the fast switching regime, and the HMM filtering posterior $\hat{\gamma}_t(s_t)$ in between. The rare catch-trials are not plotted to facilitate a direct comparison 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

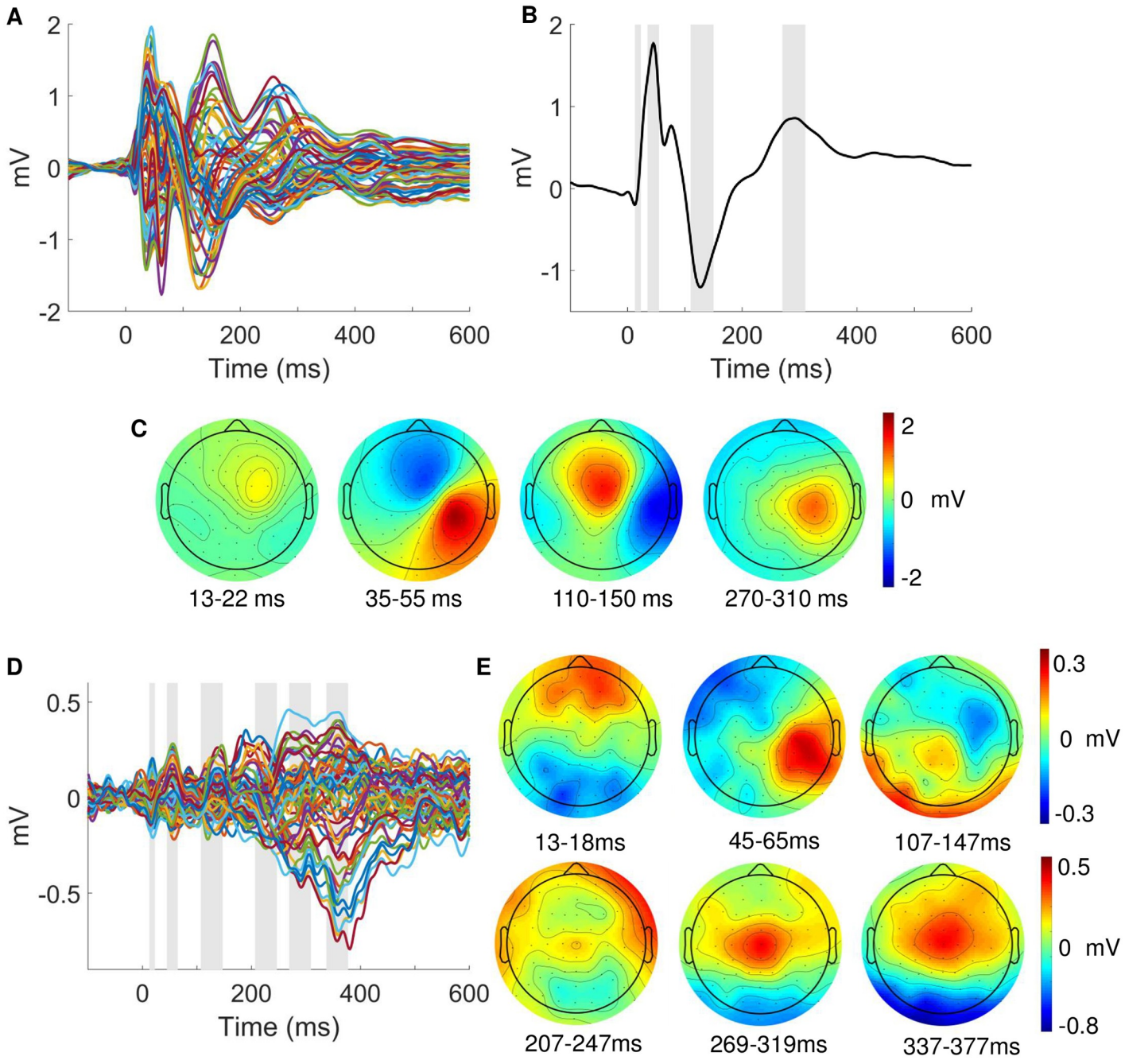


Fig 6. Event-related potentials. (A) Grand average SEP of all 64 electrodes. (B) Average SEP across electrodes C4, C6, CP4, CP6 (contralateral to stimulation). Grey bars indicate time windows around the standard somatosensory ERP components (13-23ms; 35-55ms; 110-150ms; 270-310ms). (C) ERP scalp topographies 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

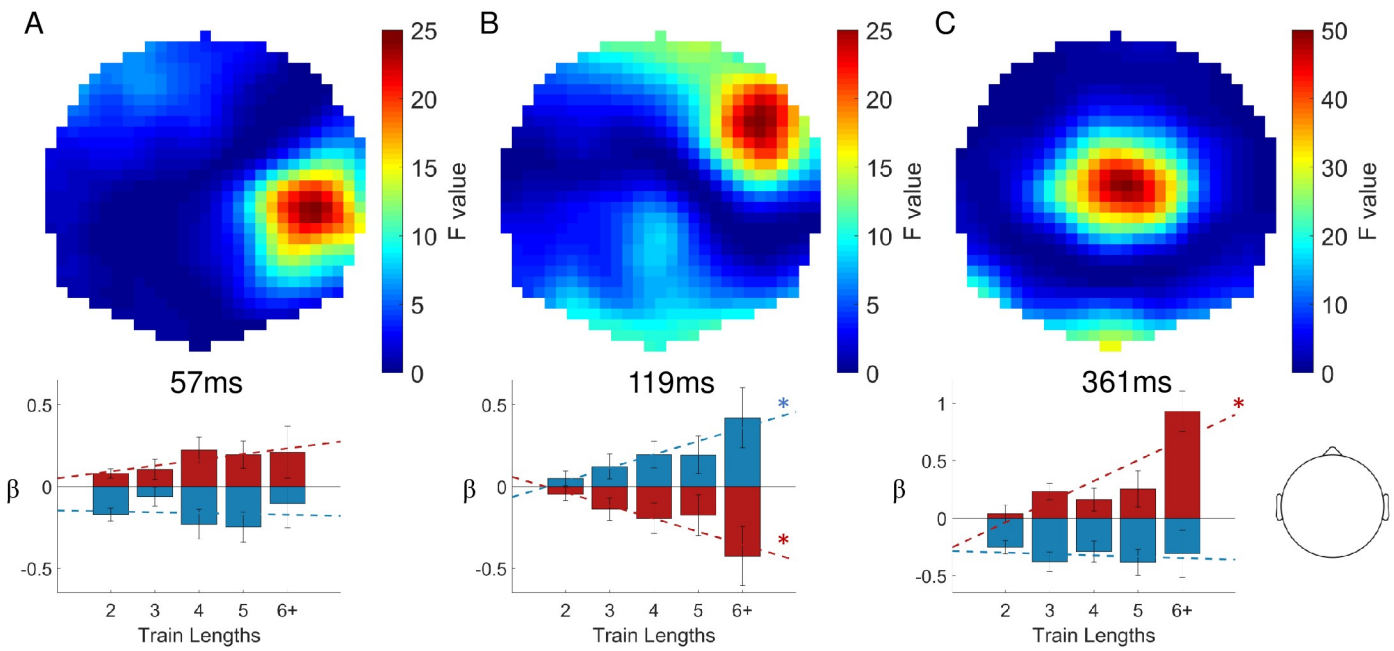


Fig 7. Statistical parametric maps of mismatch responses. Top row: Topographical F maps resulting from contrasting standard and deviant conditions averaged across the times of significant clusters: 57ms (A), 119ms (B) and 361ms (C). Bottom row: Corresponding beta parameter estimates of the significant 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.

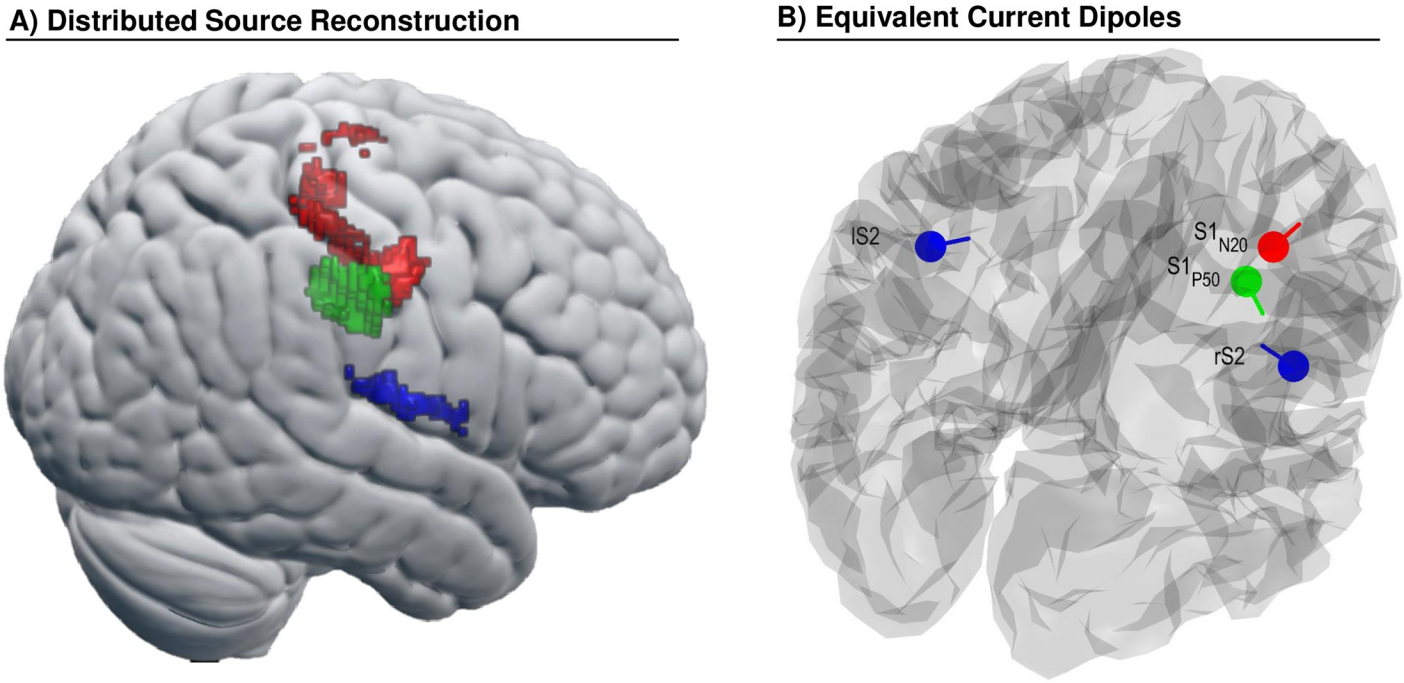
https://doi.org/10.1371/journal.pcbi.1008068.g007

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,

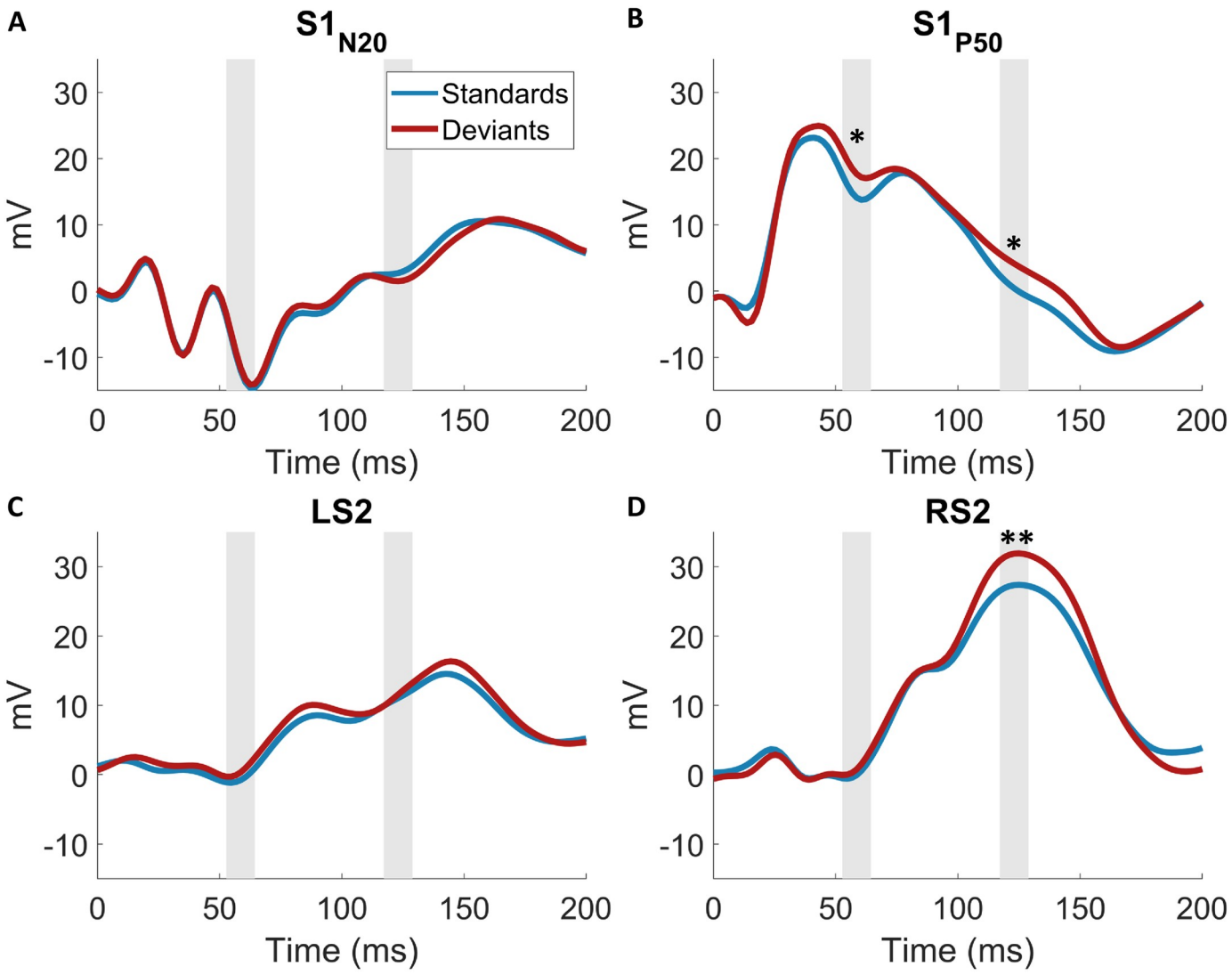


Fig 9. Grand average waveforms of EEG dipole projections. Standards and deviants were contrasted within time windows of interest informed by the GLM in the results section. $^{*}p < 0.05$; $^{**}p < 0.01$; Bonferroni corrected.

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).

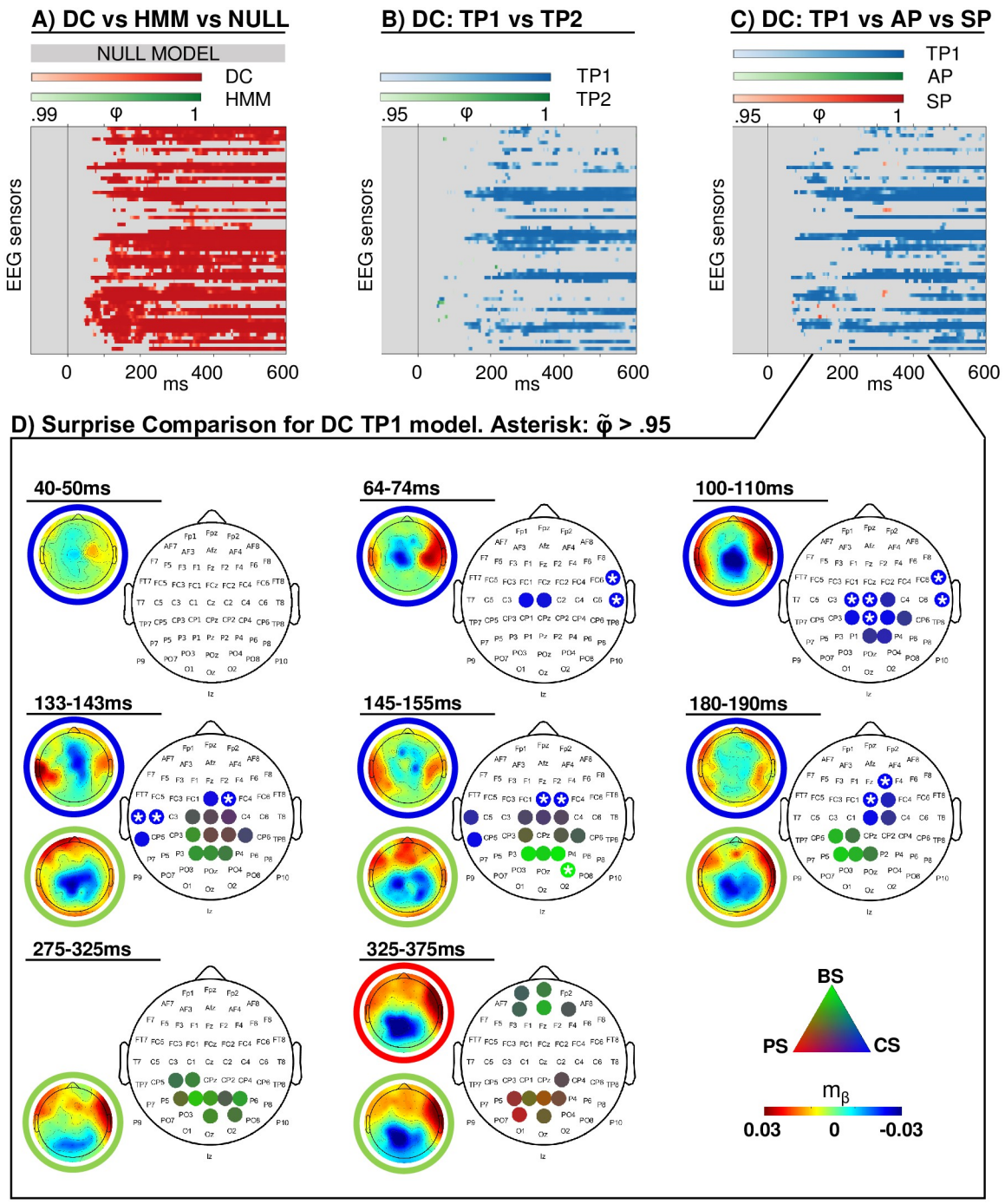


Fig 10. Modeling results. Exceedance probabilities (φ) resulting from the random-effects family-wise model comparison. (A) Dirichlet-Categorical (DC) model, Hidden Markov Model (HMM) and null model family comparison, thresholded at $\varphi > 0.99$ and applied for data reduction at all further levels. (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

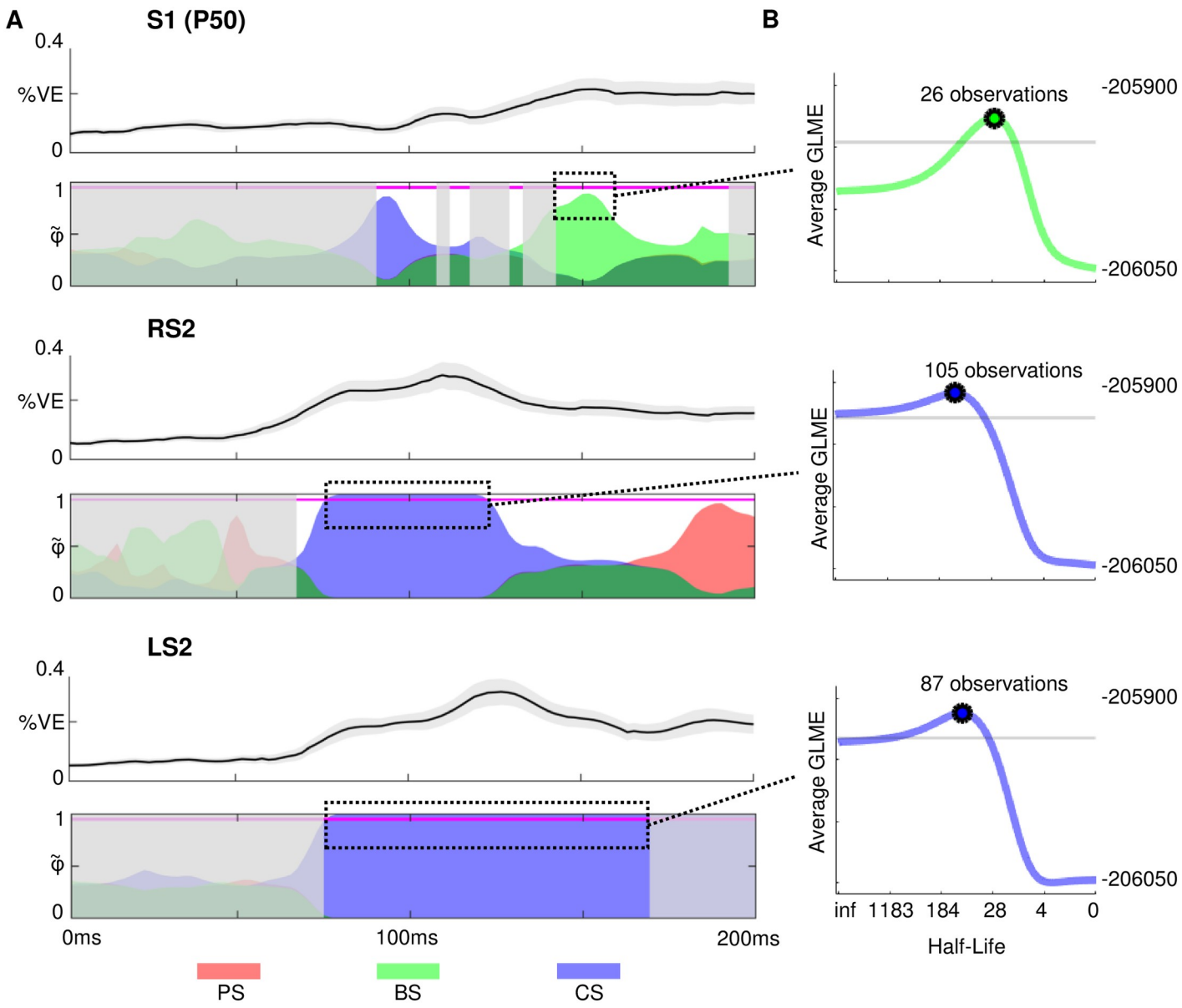


Fig 11. Modeling results in source space with best fitting forgetting-parameter values. Red: Predictive surprise (PS), Green: Bayesian surprise (BS), Blue: Confidence-corrected surprise (CS) A) Colored areas depict protected exceedance probabilities ($\tilde{\varphi}$) of the surprise readout functions 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 work: Study 2

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

HUMAN BRAIN MAPPING

EEG mismatch responses in a multi-modal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation and vision.

Journal:	*Human Brain Mapping*
Manuscript ID	HBM-22-1048
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	14-Oct-2022
Complete List of Authors:	Grundei, Miro; Freie Universität Berlin, Neurocomputation and Neuroimaging Unit; Humboldt-Universität zu Berlin, Berlin School of Mind and Brain Schröder, Pia; Freie Universität Berlin, Neurocomputation and Neuroimaging Unit Gijsen, Sam; Freie Universität Berlin, Neurocomputation and Neuroimaging Unit; Humboldt-Universität zu Berlin, Berlin School of Mind and Brain Blankenburg, Felix; Freie Universität Berlin, Neurocomputation and Neuroimaging Unit; Humboldt-Universität zu Berlin, Berlin School of Mind and Brain
Keywords:	Mismatch negativity, P3, cross-modal, multisensory, Bayesian inference, predictive processing, surprise

EEG mismatch responses in a multi-modal 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, 14195 Berlin, Germany

[2)] Berlin School of Mind and Brain, Humboldt Universität zu Berlin, 10117 Berlin, Germany

MG: m.grundei@fu-berlin.de
PS: pia.schroeder@fu-berlin.de
SG: sam.gijsen@fu-berlin.de
FB: felix.blankenburg@fu-berlin.de

Data availability:
The raw EEG data and the code used for analysis will be made freely available in an online repository upon publication of the article.

Funding:
This work was supported by Berlin School of Mind and Brain, Humboldt Universität zu Berlin (MG & 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 interests:
The authors declare that no competing interests exist.

Ethics approval:
The study was approved by the local ethics committee at the Freie Universität Berlin and complied with the Human Subjects Guidelines of the Declaration of Helsinki.

Abstract

The human brain is constantly subjected to a multi-modal stream of probabilistic sensory inputs. 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 tri-modal roving stimulus paradigm. The employed sequences consisted of high and low intensity stimuli in the auditory, somatosensory and visual modalities and were governed by uni-modal transition probabilities and cross-modal conditional dependencies. We found modality specific signatures of MMN (~100-200ms) in all three modalities, which were source localized to the respective sensory cortices and shared right lateralized pre-frontal sources. Additionally, we identified a cross-modal signature of mismatch processing in the P3a time range (~300-350ms), 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, the P3a responses across modalities were increased for mispredicted compared to predicted and unpredictable stimuli, suggesting sensitivity to cross-modal predictive information. Finally, model comparisons indicated that the observed single trial dynamics were best captured by Bayesian learning models tracking uni-modal stimulus transitions as well as cross-modal conditional dependencies.

Keywords:

Mismatch negativity, P3, cross-modal, multisensory, Bayesian inference, predictive processing, surprise

Introduction

Humans inhabit a highly structured environment governed by complex regularities. The brain is subjected to such environmental regularities by a multi-modal 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 (Gregory 1980; Bregman 1994; Friston 2005; Winkler 2009; Frost 2015; Dehaene 2015).

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; Näätänen 1978; Näätänen 2007) and the P3 (or P300; Sutton 1965; Squires 1975; Polich 2007). The MMN is defined as a negative EEG component resulting from subtraction of standard from deviant trials between ~100-200ms post-stimulus. 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 (Pazo-Alvarez 2003; Kimura 2011; Stefanics 2014) and, to a lesser extent, the somatosensory modality (Kekoni 1997; Hu 2013; Andersen 2019). The P3 is a later positive going component in response to novelty between 200-600ms around central electrodes, which has been described for the auditory, somatosensory and visual modalities and is known for its modality independent characteristics (Escera 2000; Friedman 2001; Knight 1998; Schroeger 1996; Polich 2007).

Despite being one of the most well-studied EEG components, the neuronal generation of the MMN remains subject of ongoing debate (Näätänen 2005; May 2010; Garrido 2009b). 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; May 2004; Jääskiläinen 2004; May 1999). In animals, SSA has been shown to result in response patterns similar to the MMN (Ulanovsky 2003; 2004) and simulation work suggests

that different types of MMN-like responses can be reproduced by pure adaptation models (May 2010). However, it remains unclear if the full range of MMN characteristics can be explained by adaptation alone (Garrido 2009b; Fitzgerald 2020; Wacongne 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; 1992; 2005). 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 mismatch responses 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 1999; Horvath 2004; Macdonald 2011) and unexpected stimulus omissions (Yabe 1997; Hughes 2001; Salisbury 2012 Wacongne 2011; Heilbron 2018). Similar characteristics have been reported for P3 MMRs (Duncan 2009; Prete 2022) and both MMN and P3 responses have been shown to increase for unexpected compared to expected deviants (Sussman 1998; 2005; Schroeger 2015). Insights concerning the predictive nature of mismatch responses 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 (Helmholtz 1856; Gregory 1980; Friston 2005). 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 (Friston 2005; 2010; Bastos 2012). 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 (Rao 1999; Winkler 2012; Friston 2005; 2010). 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 (Garrido 2008; 2009a; Auksztulewicz 2016). As such, the theory has the potential to unify previously opposing theories of MMN generation (Garrido 2008; 2009a; 2009b) while accounting for its key empirical features (Wacongne 2012; Heilbron 2018).

With regards to the proposed universal nature of predictive accounts of brain function, reports of comparable mismatch responses 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 2006; Kok 2014) and the somatosensory MMN (Naeije 2018; Andersen 2019), and modelling studies in all three modalities suggest that mismatch responses may reflect signatures of Bayesian learning (Lieder 2013; Maheu 2019; Stefanics 2018; Ostwald 2012; Gijsen 2021). While studies directly investigating mismatch signals in response to multi-modal 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 2006; 2008; Frost 2015) and cross-modal perceptual associations can influence statistical learning of sequence regularities (Andric 2008; Parmentier 2011), modulate mismatch responses (Besle 2005; Butler 2012; Zhao 2015; Kiat 2018; Friedel 2020) and influence subsequent uni-modal processing in various ways (Shams 2011). Recent advances in modelling Bayesian causal inference suggest that the main computational stages of multi-modal inference evolve along a multisensory hierarchy involving early sensory segregation followed by mid-latency sensory fusion and late Bayesian causal inference (Rohe 2015; 2019; Cao 2019). 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 mismatch responses in different modalities in a single experiment and to elucidate in how far they reflect local, uni-modal 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 their frequency of occurrence across the sequence is equal (Cowan 1993; Baldeweg 2004). 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 mismatch responses, 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 uni-modal transition probabilities as well as cross-modal conditional dependencies. This allowed us to test to what extent early and late mismatch responses 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 mismatch responses. Finally, we complemented our average-based analyses with single-trial modelling to investigate if signatures of uni-modal and cross-modal Bayesian inference can account for trial-to-trial fluctuations in the MMN and P3 amplitudes.

Materials and Methods

Participants underwent a novel multi-modal 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 the section *Probabilistic sequence generation*), resulting in different combinations of low and high stimuli across the three modalities in each trial.

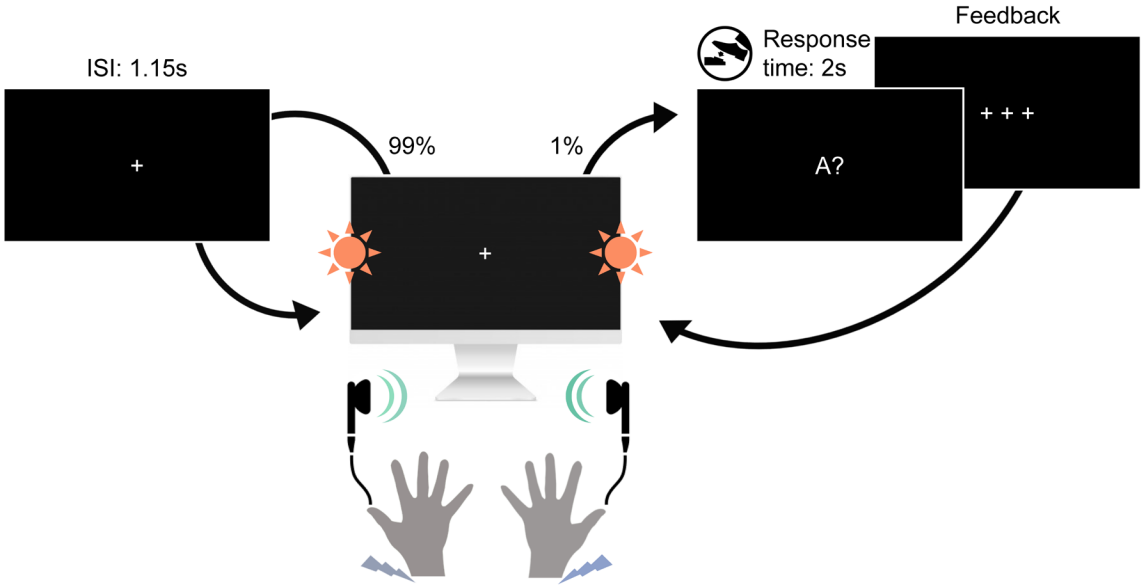


Figure 1: Experimental paradigm. Participants were seated in front of a screen and received sequences of simultaneously presented bilateral auditory beep stimuli (green), somatosensory electrical pulse stimuli (purple) and visual flash stimuli (orange) each at either low or high intensity. On consecutive 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 *Probabilistic sequence generation*). 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.

Participants

34 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.

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 1150ms (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 500Hz and 100ms 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: $81.43 \pm 1.22\,dB$; mean of the high intensity stimulus: $93.02 \pm 0.98\,dB$).

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 0.2ms 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: $3.97 \pm 0.84mA$) 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: $6.47 \pm 1.33mA$) was determined 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 100ms 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 60cm to the screen such that the LED's were placed within the visual field at a visual angle of about 67 degrees.

In each of 6 experimental runs of 11.5 minutes, 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.

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 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 6 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 6 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 $(p = 0.825)$ than changes $(p = 0.175)$ 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).

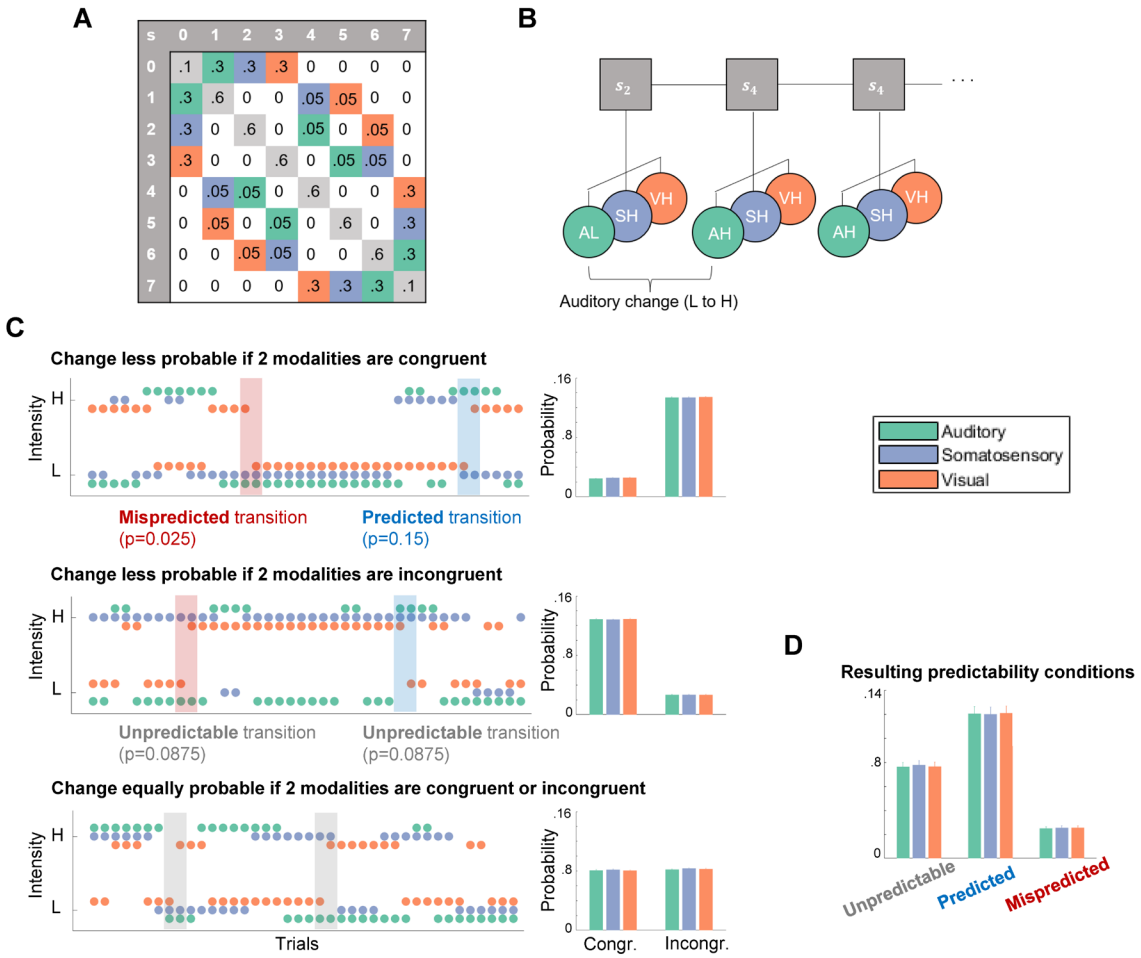


Figure 2: Probabilistic sequence generation. A) Schematic of state transition matrix. Colors depict transitions in the respective modality which were assigned specific transition probabilities: Green=auditory,

purple=somatosensory, orange=visual, light-gray=tri-modal repetition, white=multi-modal change (set to zero). 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=0.025) if the other two modalities are congruent (and likely if they are incongruent; p=0.15). Middle: Transition probabilities determine that for each modality a change is likely (p=0.15) if the other two modalities are congruent (and unlikely if they are incongruent; p=0.025). Bottom: Transition probabilities determine that for each modality a change is equally likely (p=0.0875) if the other two modalities are congruent or incongruent. D) Averaged empirical change probabilities for predictability conditions.

EEG data collection and preprocessing

Data was collected using a 64-channel active electrode EEG system (ActiveTwo, BioSemi, Amsterdam, Netherlands) at a sampling rate of 2048Hz, 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, Massachusetts, USA). First, the data was referenced against the average reference, high-pass filtered (0.01Hz), and downsampled to 512Hz. Subsequently, eye-blinks were corrected using a topographical confound approach (Berg 1994; Ille 2002) and epoched using a peri-stimulus time interval of -100 to 1050ms. All trials were visually inspected and artefactual 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 (45Hz) and the preprocessed EEG data was baseline corrected with respect to the pre-stimulus interval of -100 to -5ms. To use the general linear model (GLM) implementation of SPM, the electrode data of each participant was linearly interpolated into a 32x32 grid for each time point, resulting in one three-dimensional image (with dimensions 32x32x590) per trial. These images were then spatially smoothed with a 12x12mm full-width half-maximum (FWHM) Gaussian kernel to meet the requirements of random field theory, which the SPM software uses to control the family wise error rate.

Event-related responses and statistical analysis

First, to extract basic mismatch response 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 2007). Two time windows of interest were defined based on the literature (Duncan 2009) to search for earlier negative clusters between 50-300ms, corresponding to the mismatch negativity, and later positive clusters between 200-600ms, corresponding to the P3. Clusters were defined as adjacent electrodes with a cluster defining threshold of p_{fwe}<0.05.

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 mismatch responses (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 4 global standard and 4 global deviant regressors. 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.

Finally, a *P3-Conjunction* model was specified that consisted of 7 regressors: an intercept regressor and 6 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 2005).

Source localization

To investigate the most likely underlying neuronal sources for the mismatch negativity and P3 mismatch response 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 100-200ms. For the P3, the average MMR at 330ms 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 (BEM) 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 $p<0.05$ 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 2005).

Single-trial modelling 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 mismatch responses. To this end, we generated regressors from different Bayesian learning (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:

$$TLCD(o_t) = \begin{cases} -c_t & \text{if } d_t = 1 \\ c_{t-1} & \text{if } d_t = 0 \end{cases}$$

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 (Itti 2009; Baldi 2010) and is defined as the Kullback-Leibler (KL) divergence between the belief distribution prior and posterior to the update: $BS(y_t) = KL(p(s_{t-1}|y_{t-1}, \dots, y_1)||p(s_t|y_t, \dots, y_1))$. PS is based on Shannon's (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: $PS(y_t) = -\ln p(y_t|s_t) = -\ln\ p(y_t|y_{t-1}, \dots, y_1)$. CS additionally considers the commitment of the generative model and scales with the negative entropy of the prior distribution (Faraji 2018). It is defined as the KL divergence between the (informed) prior distribution at the current time step and a flat prior distribution $\hat{p}(s_t)$ updated with the most recent event y_t : $CS(y_t) = KL(p(s_t)||\hat{p}(s_t|y_t)))$.

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 mismatch responses (Gijsen 2021). Here, we will primarily provide a brief description of the specifics of two implementations of Dirichlet-Categorical observer models, a uni- 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 uni-modal Dirichlet-Categorical model considers a binary sequence with two possible stimulus identities (low and high) estimating transition probabilities with $y_t = o_t$ for $t = 1, \dots, T$ with a set of hidden parameters $s^{(i)}$ for each

possible transition from $o_{t-1} = i$. This uni-modal 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 $t = 2, \dots, T$) with a set of hidden parameters $s^{(i)}$ when other modalities are incongruent and $s^{(c)}$ when other modalities are congruent. Therefore, while the uni-modal 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.

Model fitting procedure

The technical details of the model fitting and subsequent Bayesian model selection 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 2007; Penny 2005; Penny 2003). The obtained model-evidence maps were subsequently subjected to the Bayesian model selection (BMS) procedure implemented in SPM12 (Stephan 2009) to draw inferences across participants with well-established handling of the accuracy-complexity trade-off (Woolrich 2012).

In total, 8 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 uni-modal regressors and one additionally including the cross-modal regressors. The purely uni-modal regression model will be called UM and the regression model including uni-modal and, additionally, cross-modal regressors will be called UCM. The design matrix of the TLCD regression model consisted of 4

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 4 regressors, an offset and the surprise responses of the uni-modal 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 only gets assigned higher model evidence than the reduced UM regression model only if the additional regressors contribute significantly to a better model fit (Stephan 2009).

To allow for the possibility of different timescales of stimulus integration (Maheu 2019; Ossmy 2013; Runyan 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.

Bayesian model comparison

The estimated model-evidence maps were used to evaluate the models' relative performance across participants via family-wise Bayesian model selection (Penny 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 mismatch responses, 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 Bayesian learning 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 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 post-stimulus 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 uni-modal (UM) and cross-modal (UCM) models, as well as the null model, to test which electrodes and time points showed influences of uni- versus multi-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 predictive surprise (PS), Bayesian surprise (BS) or confidence-corrected surprise (CS).

Results

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 $1.4 \pm 0.25s$.

Event-related potentials

Uni-modal mismatch responses

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.

Auditory MMRs

The MMN, as the classic mismatch response, 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 100-250ms (Näätänen 1978; Näätänen 1979; Näätänen 2007). Correspondingly, we find a

significant negative fronto-central auditory MMN cluster between 80-200ms (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 200-400ms, corresponding to a P3 modulation, as well as beyond 400ms 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 (Ulanovsky 2003; 2004; Cowan 1993; Baldeweg 2004). Therefore, we hypothesised a gradual increase of the auditory response to standard stimuli around the time of the MMN, known as repetition positivity (Baldeweg 2004; Baldeweg 2006; Haenschel 2005) as well as reciprocal negative modulation of the corresponding deviant response (Näätänen 2007; Bendixen 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 (94-200ms, cluster p_{fwe}<0.001). This effect was driven by a negative linear modulation of the deviant response (98-200ms, cluster p_{fwe}<0.001) as well as a repetition positivity effect on the standards (111-172ms, cluster p_{fwe}<0.001). Similarly, the later P3 MMR increased with standard repetitions (200-600ms, cluster p_{fwe}<0.001) and this effect was driven by an increase of deviant responses (200-600ms, cluster p_{fwe}<0.001) and a decrease of standard responses (205-359ms, cluster p_{fwe}<0.001). Given the temporal difference between standard (around 200-350ms) and deviant (200-600ms) train length effects, the parametric modulation of the late MMR beyond 350ms seems to be primarily driven by the increase in deviant responses.

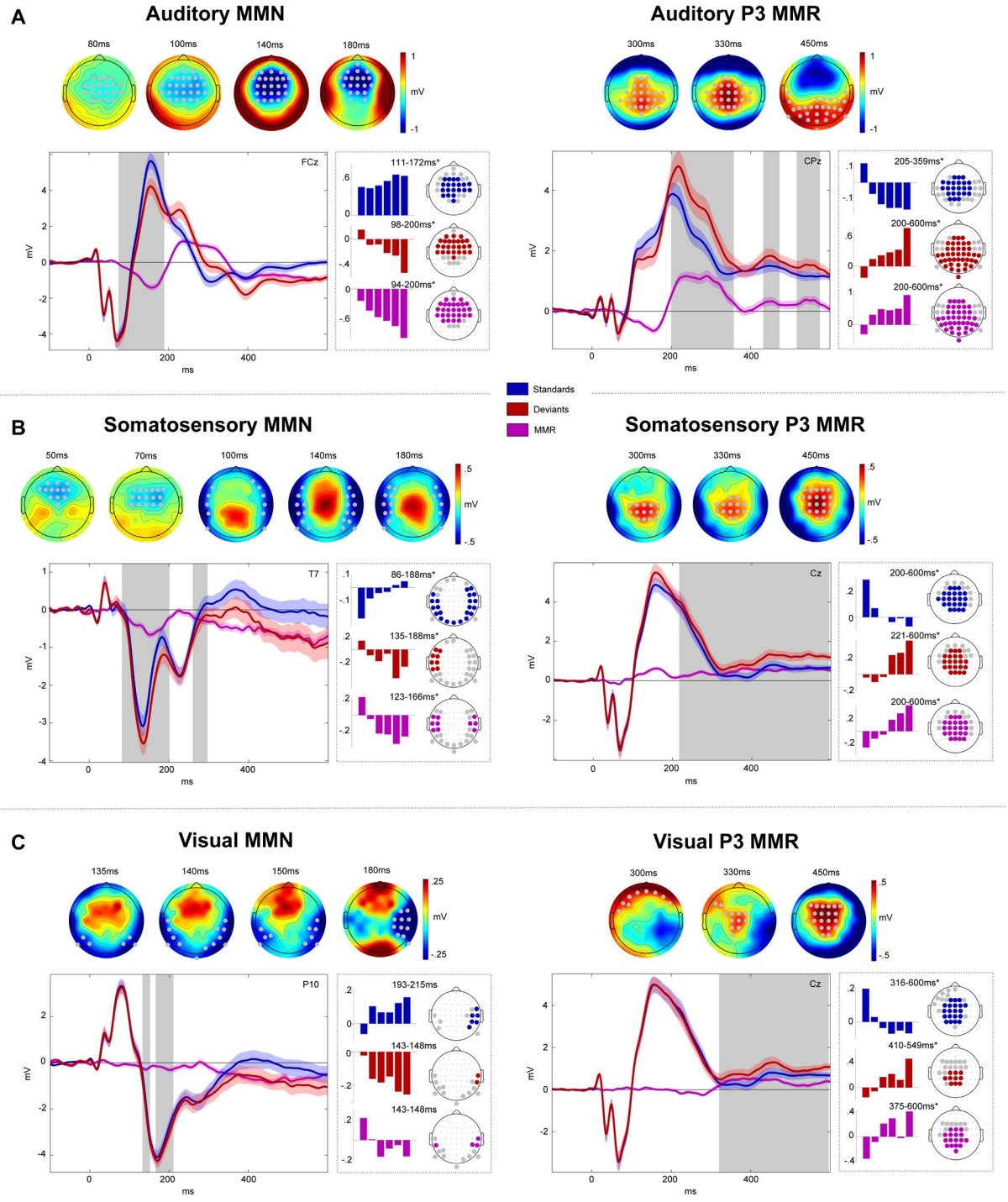


Figure 3. Mismatch responses. Panels A-C show MMRs of auditory (A), somatosensory (B) and visual (C) modalities. Within panels: Left: MMN. Right: P3 MMR. Gray dots (top) and gray boxes (bottom) indicate significant MMR electrodes and time points with p_{fwe}<0.05. Top row: MMR scalp topographies (deviants-standards). 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}<0.05).

Somatosensory MMRs

We hypothesized somatosensory MMRs to consist of early bilateral (fronto-) temporal negativities, resulting primarily from increased N140 components (Kekoni 1997), with a corresponding central positivity extending into a later central P3 component.

After an early mismatch effect starting at ~50ms at fronto-central electrodes, a more pronounced bilateral temporal cluster emerged that extended from ~90-190ms 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 (200-600ms) 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 (123-166ms, cluster p_{fwe}<0.05). This effect was driven by an increase of deviant negativity (135-188ms, cluster p_{fwe}<0.05) in combination with a positivisation of the standard (86-188ms, cluster p_{fwe}<0.05). Similarly, the later P3 MMR increases with repetition of standards (200-600ms, cluster p_{fwe}<0.05), mutually driven by increasing deviant responses (221-600ms, cluster p_{fwe} <0.05) and decreasing standard responses (200-600ms, cluster p_{fwe}<0.05).

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 ~130-200ms (figure 3C). In the later time window, we found a central positive component between ~300-600ms, 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}>0.05). 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 (143-148ms, peak p_{uncorr} <0.005), with opposite tendencies of deviant negative increase (143-148ms, peak p_{uncorr}<0.05) and standard decrease (193-215ms, peak p_{fwe}<0.05). 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 (375-600ms, cluster p_{fwe}<0.05), again constituted by an increase in deviant responses (410-549ms, cluster p_{fwe} <0.05) and concomitant decrease in standard responses (316-600ms, cluster p_{fwe}<0.05).

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 ~300ms (cluster p_{fwe}<0.05) that comprised anterior central effects around 300-350ms followed by more posterior effects from 400-600ms (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 (Wancongne 2011; Bekinschtein 2009) and to be modulated by stimulus predictability (Ritter 1999; Sussman 2003; Horvath 2008; Horvath 2012; Max 2015; Prete 2022). Indeed, within the common P3 cluster, both deviants (299-313ms, peak p_{fwe}<0.05) and standards (316-332ms, peak p_{fwe}<0.05) 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<0.001, Bonferroni corrected), *mispredicted>unpredictable* (t=14.76; p<0.001, Bonferroni corrected) and no significant difference between *unpredictable>predicted* conditions (t=0.01; p>0.05). 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; p<0.001, Bonferroni corrected), *mispredicted>unpredictable* (t=6.87; p<0.001, Bonferroni corrected) and *unpredictable>predicted* conditions (t=3.83; p<0.001, 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).

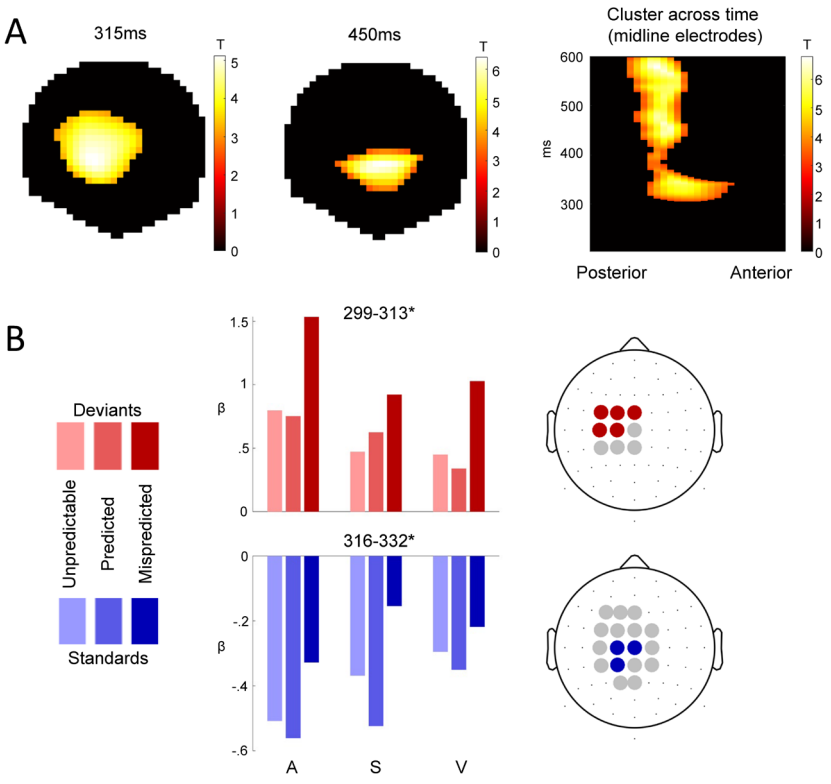


Figure 4. Cross-modal P3 effects. A) T-Maps of the conjunction of deviant>standard contrasts across the auditory, somatosensory and visual modalities. B) Beta estimates averaged across electrodes within significant clusters with peak p_{fwe}<0.05, resulting from two-way ANOVAs testing for differences between *unpredictable*, *predicted* and *mispredicted* deviants (red) and standards (blue).

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}<0.05; left cluster: peak t=6.20; right cluster: peak t=7.64) corresponding to auditory cortex and in inferior temporal areas (p_{fwe}<0.05; left cluster: peak t=5.63; right cluster: peak t=5.60). The somatosensory MMN shows highest source activation in postcentral gyrus (p_{fwe}<0.05; left cluster: peak t=5.22; right cluster: peak t=4.92) corresponding to primary somatosensory cortex. Similarly, the visual MMN shows highest source activation in the occipital cortex (p_{fwe}<0.05; 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}<0.001 (only shown in table 1) suggests additional activation of hierarchically higher sensory areas such as secondary somatosensory cortex for the sMMN (p_{uncorr}<0.001; 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 (p_{fwe}<0.05; cluster: peak t=3.15) and right middle frontal gyrus (p_{fwe}0.05; cluster: peak t=2.89). Additional significant common sources include frontal pole (p_{fwe}<0.05; left cluster: peak t=2.56; right cluster: peak t=2.28), left inferior temporal gyrus (p_{fwe}<0.05; cluster: peak t=2.52) and right inferior parietal lobe (p_{fwe}<0.05; cluster: peak t=2.85).

For the late P3 MMR a wide range of sources was expected to contribute to the EEG signal (Linden 2005; Sabeti 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}<0.05; cluster: peak t=4.34) and bilateral (pre-)frontal cortex (p_{fwe}<0.05; left inferior frontal gyrus cluster: peak t=3.57; left superior frontal gyrus cluster: peak t=3.13; left middle frontal gyrus cluster: peak t=2.87; left frontal pole cluster: peak t=3.31; right inferior frontal gyrus cluster: peak

t=3.45; right middle frontal gyrus 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}<0.05; cluster: peak t=3.21), left and right lateral occipital cortex (p_{fwe}<0.05; left cluster: peak t=2.99; right cluster: peak t=3.45) and right precentral gyrus (p_{fwe}<0.05; cluster: peak t=2.64).

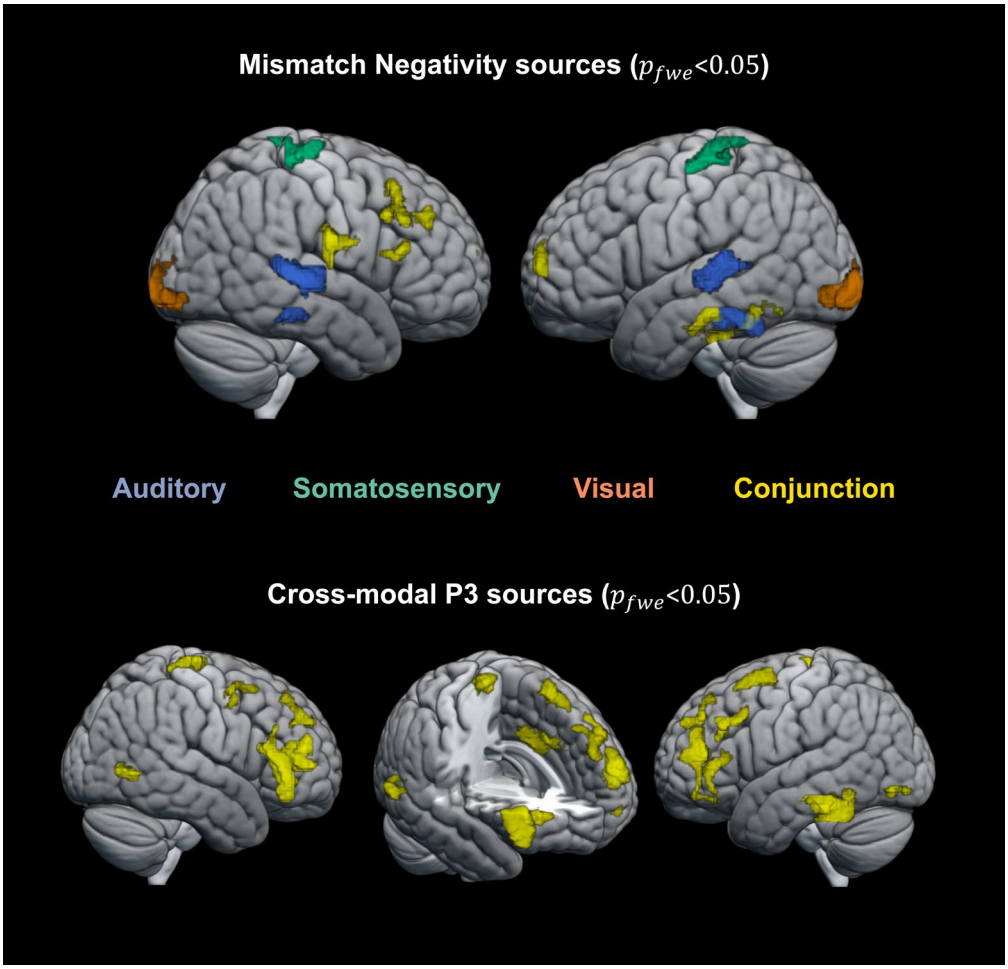


Figure 5. Source localisation. Top row: significant sources (p_{fwe}<0.05) for the auditory (purple), somatosensory (green) and visual (orange) MMNs as well as their conjunction (yellow). Bottom row: significant sources (p_{fwe}<0.05) for the conjunction (yellow) of the P3 MMR in the auditory, somatosensory and visual modalities.

Table 1. Source localization and cytoarchitectonic reference.

Contrast	hemi-sphere	cytoarchitecture (probability)	MNI coord. at cytoarch.	t statistics at cytoarch. (p-value)
aMMN	left	Auditory areas: TE 4 (67.3%) TE 3 (15.1%) TE 1 (50.7%)	 -52 -26 0 -59.8 -17.6 5.4 -50.3 -19.2 5.8	 6.2 (p_{fwe}<0.05) 4.81 (p_{fwe}<0.05) 4.05 (p_{uncorr}<0.001)

aMMN	right	Auditory areas: TE 4 (54.1%) TE 3 (33.6%) TE 1 (61.9%)	 -56 -26 0 -64.2 -16.4 5 53 -10.1 3.8	 7.64 (p_{fwe}<0.05) 5.85 (p_{fwe}<0.05) 3.53 (p_{uncorr}<0.001)
sMMN	left	Somatosensory areas: 3b [S1] (31.4%) OP4 [S2] (38.6%) OP1 [S2] (16%)	 -15.7 -33.7 68.6 -65 -14.8 20.1	 4.8 (p_{fwe}<0.05) 3.72 (p_{uncorr}<0.001)
sMMN	right	Somatosensory areas: 3b [S1] (40.3%) OP4 [S2] (prob. 46.9%) OP1 [S2] (prob. 9.1%)	 13.3 -33.7 68.2 66.5 -10.6 20.9	 4.76 (p_{fwe}<0.05) 3.51 (p_{uncorr}<0.001)
vMMN	left	Visual areas: hOc1 [V1] (84.6%) hOc2 [V2] (11.1%) hOc4v [V4] (51%) hOc3v [V3] (24.6%) FG4 (89.5%)	 -10 -100 0 -30 -87.9 -11.7 -43.5 -49.4 -13.3	 6.18 (p_{fwe}<0.05) 5.51 (p_{fwe}<0.05) 3.83 (p_{uncorr}<0.001)
vMMN	right	Visual areas: hOc1 [V1] (86.4%) hOc2 [V2] (11%) hOc3 [V3] (43.7%) hOc4 [V4] (23.8%) FG4 (59.8%)	 20 -100 -4 34.4 -88.7 -7.7 50.3 -42.5 -18.8	 5.17 (p_{fwe}<0.05) 4.7 (p_{fwe}<0.05) 3.82 (p_{uncorr}<0.001)
MMN con-junction	left	Frontal pole (28%) Inferior temporal gyrus (49%)	-14 62 8 -50 -24 -28	2.56 (p_{fwe}<0.05) 2.52 (p_{fwe}<0.05)
MMN con-junction	right	Middle frontal gyrus (36%) Inferior frontal gyrus (53%) Frontal pole (65%) Inferior parietal lobe (46.8%)	42 28 30 53 27.6 17.3 30 42 32 62 -14 24	2.89 (p_{fwe}<0.05) 2.96 (p_{fwe}<0.05) 2.28 (p_{fwe}<0.05) 2.85 (p_{fwe}<0.05)
P3 con-junction	(left)	Anterior cingulate gyrus (51%)	-6 14 36	4.34 (p_{fwe}<0.05)
P3 con-junction	left	Frontal pole (74%) Inferior frontal gyrus (35%) Middle frontal gyrus (78%) Superior frontal gyrus (43%) Inferior temporal gyrus (34%) Lateral occipital [hOc4la] (81%)	-34 44 20 52 26 18 -44 30 32 -22 12 60 -48 -44 -24 -46 -80 -8	3.31 (p_{fwe}<0.05) 2.74 (p_{fwe}<0.05) 2.87 (p_{fwe}<0.05) 3.13 (p_{fwe}<0.05) 3.21 (p_{fwe}<0.05) 2.99 (p_{fwe}<0.05)

P3 con-junction	right	Frontal pole (86%) Inferior frontal gyrus (39.5%) Middle frontal gyrus (38%) Precentral gyrus [4a] (19%) Lateral occipital cortex [hOc5] (55%)	45 42 10 52 26 18 36 2 54 6 -32 64 56 -62 0	3.57 (p_{fwe}<0.05) 3.45 (p_{fwe}<0.05) 3.0 (p_{fwe}<0.05) 2.64 (p_{fwe}<0.05) 3.45 (p_{fwe}<0.05)

Single-trial modelling

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 8 different models capturing various learning strategies were fit to the single-trial EEG data and compared via family-wise Bayesian model selection. A summary of the modelling 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 Bayesian model selection which is schematically depicted in figure 6A. The first comparison concerned the train length dependent change detection model (TLCD), a model family containing all Bayesian learning 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 $\varphi > 0.95$ from ~70ms 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 focussed 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 multi-modal configurations beyond uni-modal transition probability inference. Within the electrodes and time-points with sufficient evidence for Bayesian learning signatures (as established by the first model comparison), a comparison of a purely uni-modal 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 uni-modal models are highlighted in figure 6B and C. The UCM family outperforms the UM family at central and fronto-central electrodes at ~100-400ms (with $\varphi > 0.95$).

Inspection of the beta estimates of auditory, somatosensory and visual regressors of the UCM regression models shows that the beta maps of the uni-modal 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 confidence-corrected surprise, 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, Bayesian surprise 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 20 stimuli. The same range of stimulus integration was found for the UCM regression model. Overall, confidence-corrected surprise models tended to have higher

integration windows (~10-20 stimuli) compared to Bayesian surprise models (~5-10 stimuli).

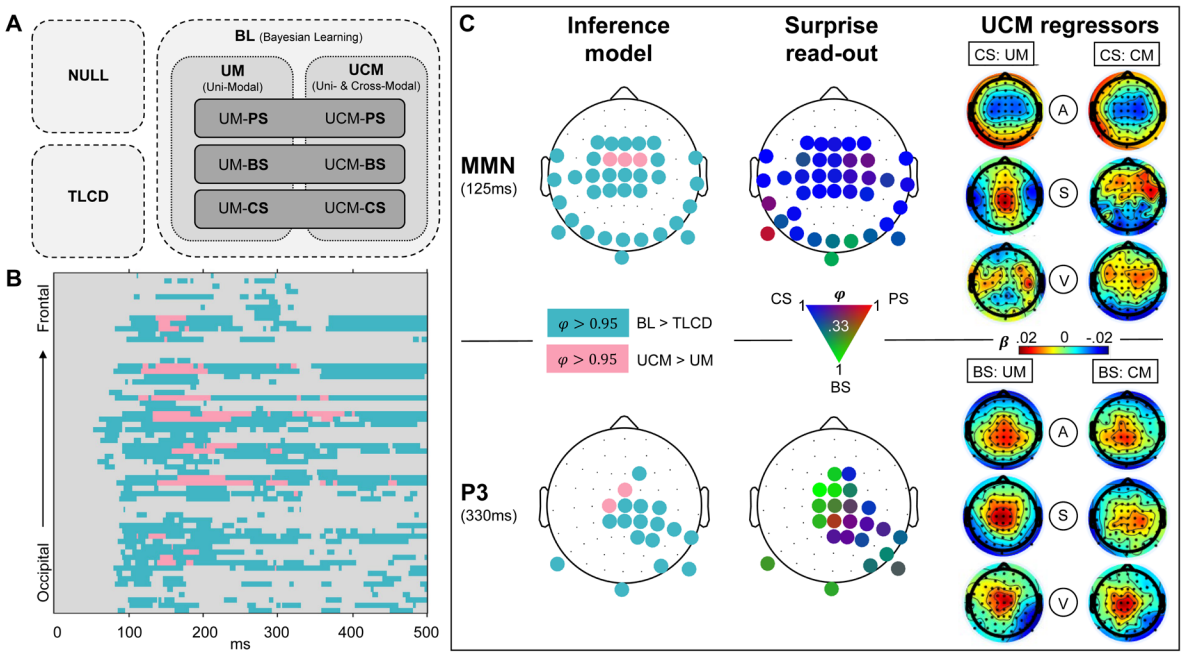


Figure 6. Modelling results. A) Schematic overview of models. Model comparison 1 (light-gray box, dashed contour): Null model family (NULL), train length dependent change detection model family (TLCD) and Bayesian learning model family (BL). Comparison 2 (gray box, dotted contour): Uni-modal regression 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 post-stimulus 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 3-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=uni-modal) for CS read-out models (top) and BS read-out models (bottom).

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 100-200ms 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 visual MMN, 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 Bayesian learning models, tracking transitions between observations, captured the observed dynamics of single-trial responses to the roving stimulus sequences better than a static model reflecting train length dependent change detection. Moreover, a BL model which additionally captured cross-modal conditional dependence of stimulus alternation outperformed a purely uni-modal 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 confidence-corrected surprise whereas later P3 MMRs seem to reflect dynamics of Bayesian surprise.

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 (Näätänen 1978; Winkler 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 Näätänen (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 (Näätänen 2007) and might speak against pure N1 adaptation (as suggested by May 2004; Jääskiläinen 2004; May 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 1997; Spackman 2007; 2010; Shen 2018) 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 (Shinozaki 1998; Akatsuka 2005). However, some evidence appears to converge on a temporally centred negativity with corresponding central positivity as the primary sMMN around 140ms (Ostwald 2012; Gijsen 2021).

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 visual MMN (vMMN) at occipital electrodes extending to temporal electrodes within a time window of 100-200ms post stimulus, with corresponding (fronto-)central positivity. This observation is in line with previous research reporting posterior (Urakawa 2010; Kimura 2010; Clery 2013) and temporal (Hesenfeld 2003; Kuldkepp 2013) patterns of vMMN with corresponding central positivity (Czigler 2006; Cleary 2013; File 2017).

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 inferior frontal gyrus (IFG) and middle frontal gyrus (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 (STG) as well as right IFG as underlying the aMMN (Opitz 2002; Molholm 2005; Näätänen 2005; Garrido 2008; Garrido 2009a) 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 2007a; 2007b; Spackman 2010; Butler 2012; Ostwald 2012; Naeije 2016; 2018; Andersen 2019; Gijsen 2021) in combination with involvement of (inferior) frontal regions (Huang 2005; Ostwald 2012; Allen 2016; Fardo 2017; Downar 2000).

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 (Yucel 2007; Kimura 2010; 2011; 2012; Urakawa 2010) and have particularly highlighted the IFG as a frontal generator of the vMMN (Downar 2000; Hedge 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 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 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 2009b).

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 (Sams 1983; Näätänen 1992; Imada 1993; Javitt 1998).

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 visual MMN. 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 (2004; 2006; 2007; Haenschel 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 2005). With regards to its functional role, it has been argued to reflect auditory sensory memory trace formation (Baldeweg 2004; Costa-Faidella 2011a; Costa-Faidella 2011b). 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 2013), potentially because a memory trace of the standard stimulus identity must be re-established 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 2004; Haenschel 2005; Costa-Faidella 2011a; Costa-Faidella 2011b). Importantly, although some evidence exists to suggest an additional role for train length dependent deviant modulation also in roving paradigms (Cowan 1993; Haenschel 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 100ms 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 2007), only few other studies have reported sensitivity of the sMMN to stimulus repetition. Interestingly, in our previous study on

somatosensory MMRs (Gijsen 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. Grill-Spector 2006; Clifford 2007) it is rarely directly reported in visual MMN studies (but see Kremláček 2016). Overall, the visual MMN 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 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 *Limitations* for a discussion of alternative explanations regarding the current study).

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 2009b). 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 (Naeije 2018; Andersen 2019) and visual (Czigler 2006; Kok 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 2011; Stefanics 2014), signaling prediction error instead of basic change detection (Stefanics 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 (Aukstulewicz 2016; Baldeweg 2006, 2007; Costa-Faidella 2011a; Costa-Faidella 2011b). 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 (Friston 2005; 2009; Aukstulewicz 2016). In line with these hypotheses, Garrido and colleagues (2008; 2009a) used dynamic causal modelling (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 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.

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 300-350ms in central electrodes, followed by a slightly more posterior cluster extending from 400-600ms. Particularly the earlier cluster may correspond to the well-known P3a response, which peaks at around 300ms after change-onset at (fronto-)central electrodes and is thought to be elicited regardless of

sensory modality (Escera 2000; Friedman 2001; Knight 1998; Schroeger 1996; Polich 2007).

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; Schroeger 2015). While the P3a has been initially related to attentional switches to task-irrelevant but salient stimulus features (Escera 2000; Friedman 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 (Wacongne 2011; Schroeger 2015). Similar to the MMN, P3 responses are known to be modulated by stimulus probability (Duncan-Johnson 1977) and can be elicited by unexpected stimulus repetitions (Squires 1976; Duncan 2009) and omissions of predicted sound stimuli (Sutton 1967; Prete 2022), 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 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 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 (Ritter 1999; Sussman 2003; Horvath 2012). Similarly, explicit top-down knowledge of sequence regularities has been shown to reduce the P3, while leaving the MMN unaffected (Max 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 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 2021). Furthermore, computational studies indicate that P3 responses reflect specific quantities of unexpectedness as well as updates to a prior belief (Jepma 2017; Kolossa 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 (Wacongne 2011; Bekinschtein 2009; Winkler 2005). Our findings of a sensitivity of the P3 response to cross-modal predictive information carried by the multi-modal 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 multi-modal context information.

The observed pattern suggests influences of precision weighting on prediction errors (Friston 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 multi-modal 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 multi-modal 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 supra-modal processing of stimulus transitions and deviance detection (Downar 2000; Huang 2005). Similarly, a fronto-parietal attention network (Corbetta 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 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 supra-modally to facilitate processing of information from multi-modal events (Macaluso 2005; 2010). Thus, the predictive information in the multi-modal sequences presented in the current study may be processed in such a fronto-parietal network to aid the perception of multi-modal stimulus streams. Future research would benefit from studies further investigating such multi-modal probabilistic sequences with higher spatial resolution to inform these proposed interpretations.

Modelling 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 train length dependent change detection (TLCD) model to families of Bayesian learning (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 uni-modal transitions, outperformed a purely uni-modal 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 (Ghazanfanar 2006; Driver 2008) and that processing multi-modal information might be appropriately captured by Bayesian inference (Kording 2007; Shams 2022). Interestingly, however, an earlier cross-modal effect was found prior to 300ms 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 modelling 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 2008; 2009a, 2009b). 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 confidence-corrected surprise (CS) in earlier mismatch signatures prior to 200ms, while the late clusters tend to reflect Bayesian surprise (BS). This is well in line with our previous study performed in the somatosensory modality (Gijsen 2021) and other studies have similarly reported a reflection of BS in P3 mismatch responses (Ostwald 2012; Kolossa 2015; Mars 2008; Seer 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 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 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.

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 organisation of the visual cortex (Sereno 1995; Horton 1991), a “far peripheral” placement (i.e. >60 degrees; Strassburger 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 Bayesian learning 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).

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 uni-modal and multi-modal sensory input streams.

Acknowledgement

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

References

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.

An, H., Ho Kei, S., Auksztulewicz, R., & Schnupp, J. W. (2021). Do auditory mismatch responses differ between acoustic features?. *Frontiers in Human Neuroscience*, *15*, 613903.

Andric, M., Davis, B., & Hasson, U. (2017). Visual cortex signals a mismatch between regularity of auditory and visual streams. *NeuroImage*, *157*, 648-659.

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: Past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160047.

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *cortex*, *80*, 125-140.

Baldeweg T, Williams JD, Gruzelier JH. Differential changes in frontal and sub-temporal components of mismatch negativity. Int J Psychophysiol. 1999;33:143–148.

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.

Baldeweg, T., Wong, D., & Stephan, K. E. (2006). Nicotinic modulation of human auditory sensory memory: evidence from mismatch negativity potentials. *International Journal of Psychophysiology*, *59*(1), 49-58.

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: a predictive coding perspective. *Journal of Psychophysiology*, *21*(3-4), 204-213.

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*, *106*(5), 1672-1677.

Berg P, Scherg M. A multiple source approach to the correction of eye artifacts. Electroencephalogra- phy and clinical neurophysiology. 1994; 90(3):229–241. https://doi.org/10.1016/0013-4694(94)90094- 9 PMID: 7511504

Besle, J., Fort, A., & Giard, M. H. (2005). Is the auditory sensory memory sensitive to visual information?. *Experimental Brain Research*, *166*(3), 337-344.

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), 2-2.

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.

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. *Journal of Neuroscience*, *32*(44), 15338-15344.

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

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.

Costa-Faidella, J., Baldeweg, T., Grimm, S., & Escera, C. (2011a). Interactions between “what” and “when” in the auditory system: temporal predictability enhances repetition suppression. *Journal of Neuroscience*, *31*(50), 18590-18597.

Costa-Faidella, J., Grimm, S., Slabu, L., Diaz-Santaella, F., & Escera, C. (2011b). 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., & Näätänen, 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.

Czigler, I., Winkler, I., Pató, L., Várnagy, A., Weisz, J., & Balázs, 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.

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

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

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*(5), 456-467.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., ... & 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.

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.

Escera, C., Alho, K., Schroger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. Audiol Neurootol, 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.

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.

Flandin, G., & Penny, W. D. (2007). Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage*, *34*(3), 1108-1125.

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. Neurosci Biobehav Rev, 25(4), 355-373. https://doi.org/10.1016/s0149-7634(01)00019-7

Friedel, E. B., 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.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815-836.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society B: Biological sciences*, *364*(1521), 1211-1221.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127-138.

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.

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., Stephan, K. E., Baldeweg, T., & Friston, K. J. (2009a). Repetition suppression and plasticity in the human brain. *Neuroimage*, *48*(1), 269-279.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009b). The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol, 120(3), 453-463. https://doi.org/10.1016/j.clinph.2008.11.029

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 181-197.

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.

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. *Journal of Neuroscience*, *25*(45), 10494-10501.

Hedge, C., Stothart, G., Jones, J. T., Frías, P. R., Magee, K. L., & Brooks, J. C. (2015). A frontal attention mechanism in the visual mismatch negativity. *Behavioural brain research*, *293*, 173-181.

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

Heslenfeld, D. J. (2003). Detection of change: event-related potential and fMRI findings.

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.

Horváth, J., & Winkler, I. (2004). How the human auditory system treats repetition amongst change. *Neuroscience letters*, *368*(2), 157-161.

Horváth, J., Müller, D., Weise, A., & Schröger, E. (2010). Omission mismatch negativity builds up late. *Neuroreport*, *21*(7), 537-541.

Horvath, J., & Bendixen, A. (2012). Preventing distraction by probabilistic cueing. Int J Psychophysiol, 83(3), 342-347. https://doi.org/10.1016/j.ijpsycho.2011.11.019

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.

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), 1073-1089.

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.

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.

Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW. Human posterior auditory cortex gates novel sounds to consciousness. Proc. Natl. Acad. Sci. USA. 2004;101:6809–6814.

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/Evoked Potentials Section*, *108*(2), 143-153.

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.

Kiat, J. E. (2018). Assessing cross-modal target transition effects with a visual-auditory oddball. *International Journal of Psychophysiology*, *129*, 58-66.

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

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.

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.

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

Kimura, M. (2012). Visual mismatch negativity and unintentional temporal-context-based prediction in vision. *International Journal of Psychophysiology*, *83*(2), 144-155.

Knight, R. T., & Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. J Clin Neurophysiol, 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.

Kolossa, A., Fingscheidt, T., Wessel, K., & Kopp, B. (2013). A model-based approach to trial-by-trial P300 amplitude fluctuations. *Frontiers in human neuroscience*, *6*, 359.

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

Körding, 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.

Kremláček, J., Kreegipuu, K., Tales, A., Astikainen, P., Poldver, N., Näätänen, R., & Stefanics, G. (2016). Visual mismatch negativity (vMMN): A review and meta-analysis of studies in psychiatric and neurological disorders. *Cortex*, *80*, 76-112.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *JOSA A*, *20*(7), 1434-1448.

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.

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.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in neurosciences*, *28*(5), 264-271.

Macaluso, E. (2010). Orienting of spatial attention and the interplay between the senses. *Cortex*, *46*(3), 282-297.

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.

Male, A. G., O'Shea, R. P., Schröger, E., Müller, 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.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177-190.

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. *Journal of Neuroscience*, *28*(47), 12539-12545.

Max, C., Widmann, A., Schroger, E., & Sussman, E. (2015). Effects of explicit knowledge and predictability on auditory distraction and target performance. Int J Psychophysiol, 98(2 Pt 1), 174-181. https://doi.org/10.1016/j.ijpsycho.2015.09.006

May P, Tiitinen H, Ilmoniemi RJ, Nyman G, Taylor JG, Näätänen R. Frequency change detection in human auditory cortex. J. Comput. Neurosci. 1999;6:99–120.

May PJ, Tiitinen H. The MMN is a derivative of the auditory N100 response. Neurol Clin Neurophysiol. 2004 Nov 30;2004:20. PMID: 16012601.

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

Maheu, M., Dehaene, S., & Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. *elife*, *8*, e41541.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, *108*(48), 19401-19406.

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. *Journal of Neuroscience*, *36*(37), 9572-9579.

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.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica*, *42*(4), 313-329.

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

Näätänen, R. (2003). Mismatch negativity: clinical research and possible applications. *International Journal of Psychophysiology*, *48*(2), 179-188.

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.

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.

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.

Parmentier, F. B., Elsley, J. V., Andrés, P., & Barceló, F. (2011). Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition*, *119*(3), 374-380.

Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: a review. *Biological psychology*, *63*(3), 199-236.

Penny, W., Kiebel, S., & Friston, K. (2003). Variational Bayesian inference for fMRI time series. *NeuroImage*, *19*(3), 727-741.

Penny, W. D., Trujillo-Barreto, N. J., & Friston, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, *24*(2), 350-362.

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.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol, 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. Eur J Neurosci. 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.

Recasens, M., & Uhlhaas, P. J. (2017). Test–retest reliability of the magnetic mismatch negativity response to sound duration and omission deviants. *Neuroimage*, *157*, 184-195.

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.

Rohe, T., Ehlis, A. C., & Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. Nat Commun, 10(1), 1907. https://doi.org/10.1038/s41467-019-09664-2

Sabeti, M., Katebi, S. D., & Rastgar, K. (2015). Source localization algorithms to find attention and memory circuits in the brain. *Journal of King Saud University-Computer and Information Sciences*, *27*(3), 334-343.

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.

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): emitted MMN to violations of an auditory gestalt. *Psychophysiology*, *49*(4), 544-548.

Sams, M., Alho, K., & Näätänen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological psychology*, *17*(1), 41-58.

Schroger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. J Cogn Neurosci, 8(6), 527-539. https://doi.org/10.1162/jocn.1996.8.6.527

Schröger, E., & Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport*, *9*(15), 3355-3358.

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.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., ... & Tootell, R. B. H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889-893.

Shams, L., Wozny, D. R., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in psychology*, *2*, 264.

Shams, L., & Beierholm, U. (2022). Bayesian causal inference: a unifying neuroscience theory. *Neuroscience & Biobehavioral Reviews*, 104619.

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

Shinozaki, N., Yabe, H., Sato, Y., Hiruma, T., Sutoh, T., Matsuoka, T., & Kaneko, S. (2003). Spectrotemporal window of integration of auditory information in the human brain. *Cognitive Brain Research*, *17*(3), 563-571.

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in human neuroscience*, *8*, 666.

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.

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of vision*, *11*(5), 13-13.

Sussman, E., Ritter, W., & Vaughan Jr, H. G. (1998). Predictability of stimulus deviance and the mismatch negativity. *Neuroreport*, *9*(18), 4167-4170.

Sussman, E., Winkler, I., & Schroger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. Psychon Bull Rev, 10(3), 630-637. https://doi.org/10.3758/bf03196525

Sussman, E. S., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. *Neuroreport*, *16*(13), 1519-1523.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187-1188.

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.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature neuroscience*, *6*(4), 391-398.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience*, *24*(46), 10440-10453.

Urakawa, T., Inui, K., Yamashiro, K., & Kakigi, R. (2010). Cortical dynamics of the visual change detection process. *Psychophysiology*, *47*(5), 905-912.

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.

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. Proc Natl Acad Sci U S A, 108(51), 20754-20759. https://doi.org/10.1073/pnas.1117807108

Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. J Neurosci, 32(11), 3665-3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Winkler, I., Takegata, R., & Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Brain Res Cogn Brain Res, 25(1), 291-299. https://doi.org/10.1016/j.cogbrainres.2005.06.005

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

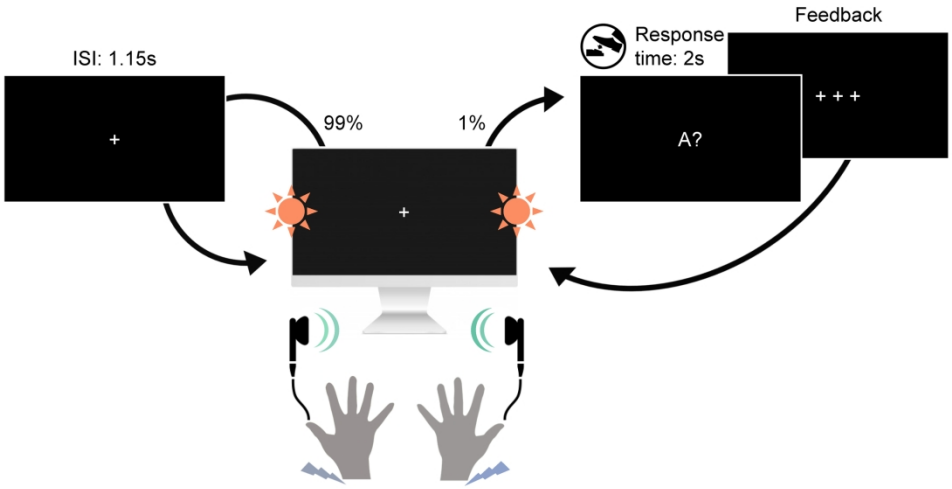
Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci, 13(12), 532-540. https://doi.org/10.1016/j.tics.2009.09.003

Woolrich, M. W. (2012). Bayesian inference in FMRI. *NeuroImage*, *62*(2), 801-810.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *Neuroreport*, *8*(8), 1971-1974.

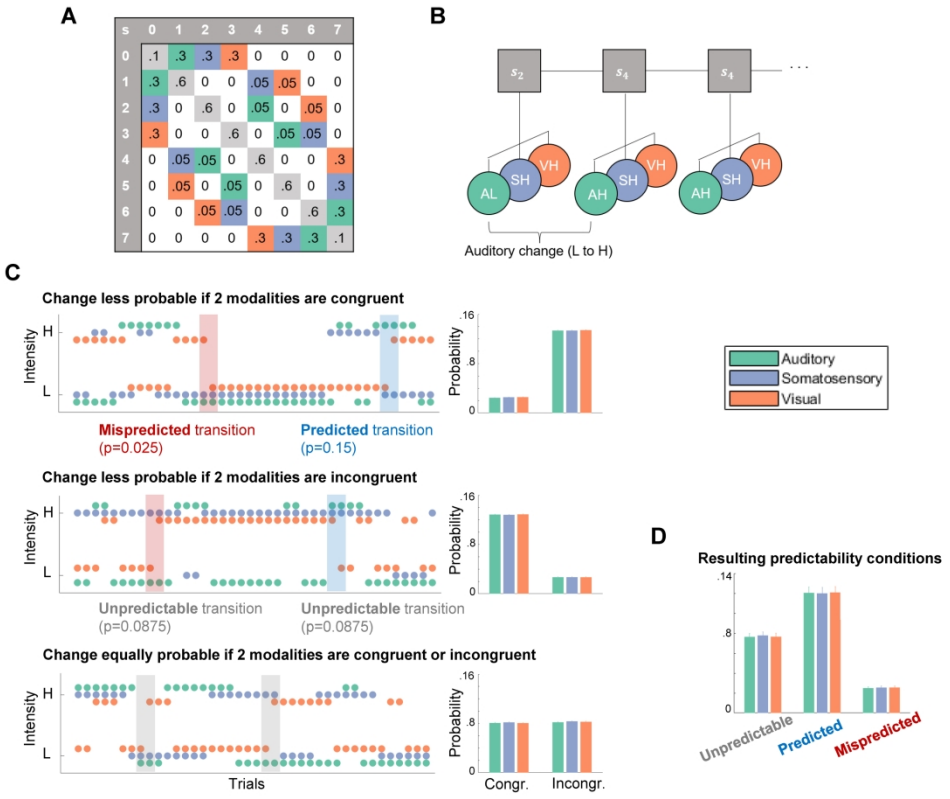
Yucel, G., McCarthy, G., & Belger, A. (2007). fMRI reveals that involuntary visual deviance processing is resource limited. *Neuroimage*, *34*(3), 1245-1252.

Zhao, C., Valentini, E., & Hu, L. (2015). Functional features of crossmodal mismatch responses. *Experimental Brain Research*, *233*(2), 617-629.



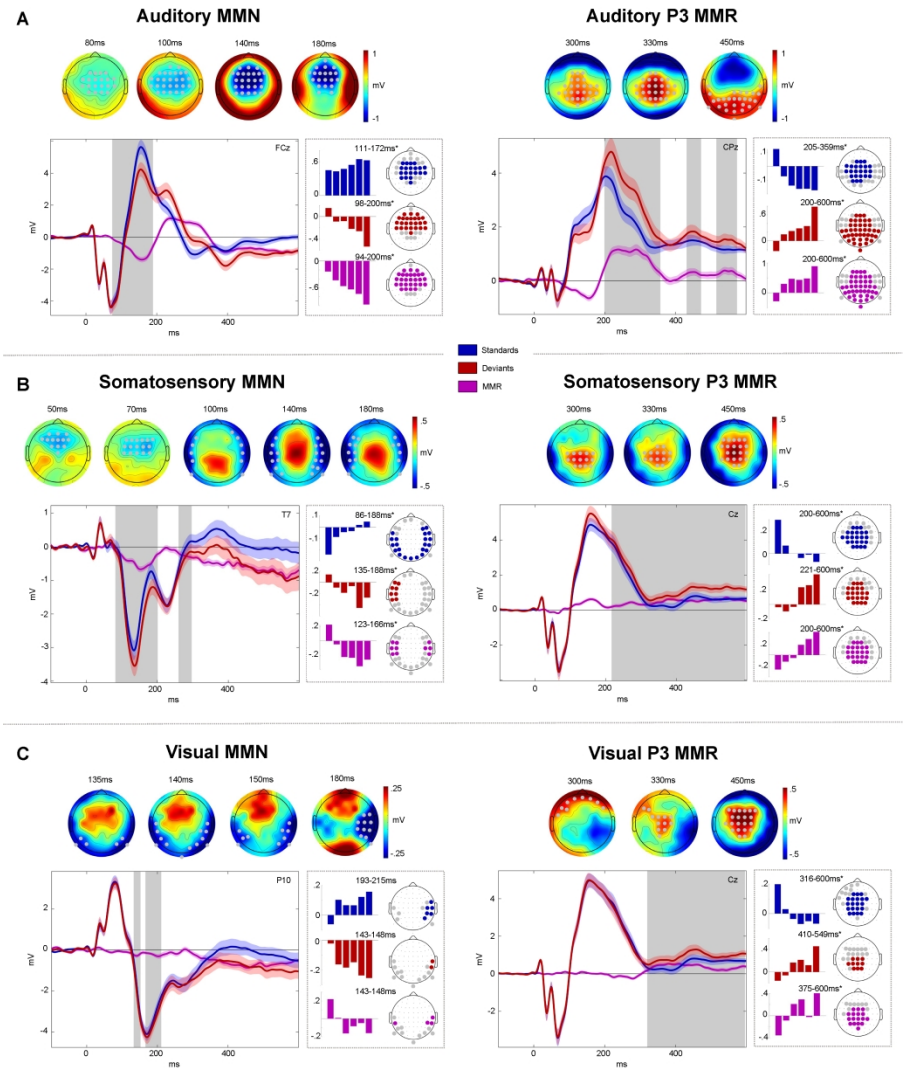
Experimental paradigm. Participants were seated in front of a screen and received sequences of simultaneously presented bilateral auditory beep stimuli (green), somatosensory electrical pulse stimuli (purple) and visual flash stimuli (orange) each at either low or high intensity. On consecutive trials, 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 Probabilistic sequence generation). 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.

215x114mm (300 x 300 DPI)



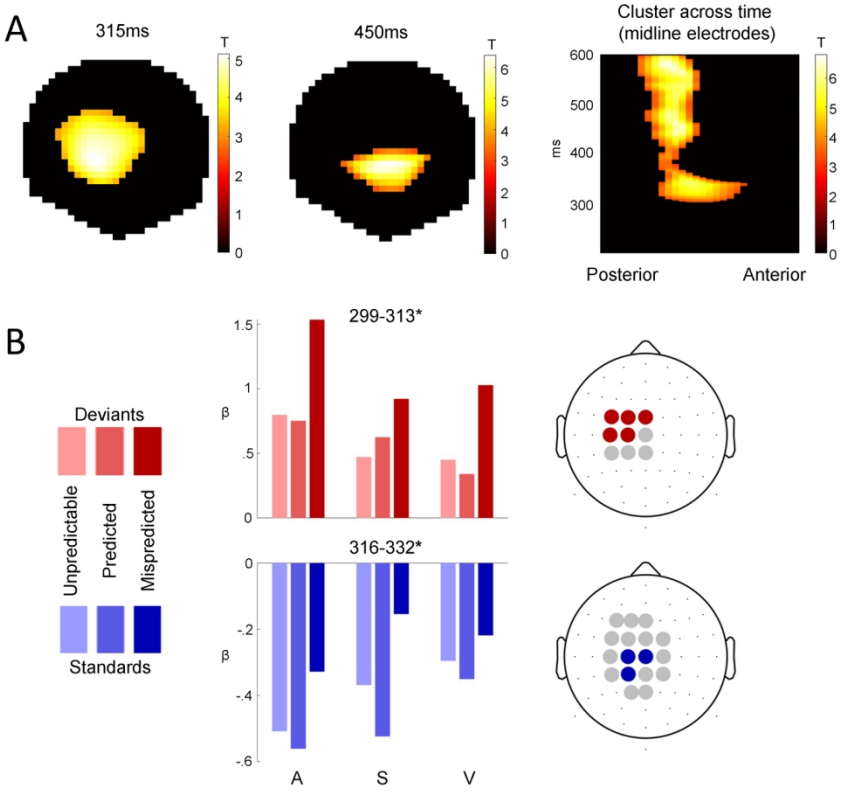
Probabilistic sequence generation. A) Schematic of state transition matrix. Colors depict transitions in the respective modality which were assigned specific transition probabilities: Green=auditory, purple=somatosensory, orange=visual, light-gray=tri-modal repetition, white=multi-modal change (set to zero). 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=0.025) if the other two modalities are congruent (and likely if they are incongruent; p=0.15). Middle: Transition probabilities determine that for each modality a change is likely (p=0.15) if the other two modalities are congruent (and unlikely if they are incongruent; p=0.025). Bottom: Transition probabilities determine that for each modality a change is equally likely (p=0.0875) if the other two modalities are congruent or incongruent. D) Averaged empirical change probabilities for predictability conditions.

266x215mm (300 x 300 DPI)



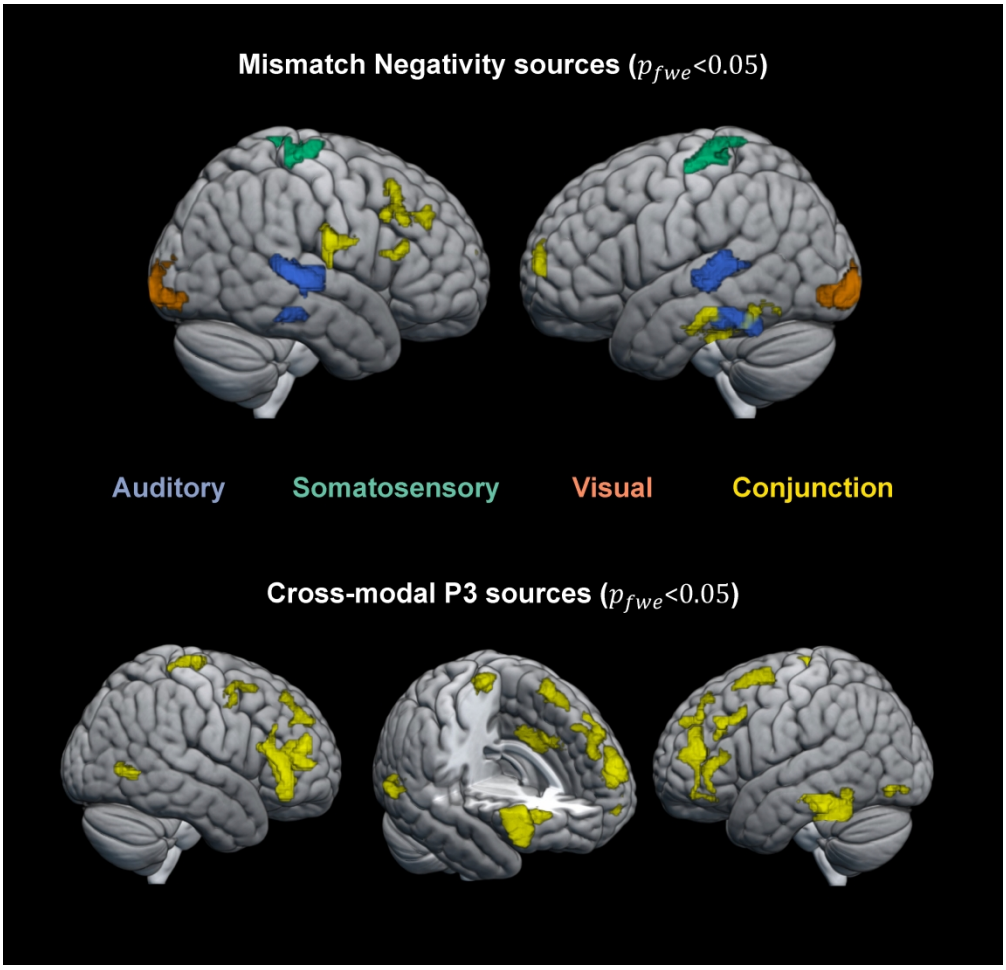
Mismatch responses. Panels A-C show MMRs of auditory (A), somatosensory (B) and visual (C) modalities. Within panels: Left: MMN. Right: P3 MMR. Gray dots (top) and gray boxes (bottom) indicate significant MMR electrodes and time points with p_fwe<0.05. Top row: MMR scalp topographies (deviants-standards). 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<0.05).

304x406mm (300 x 300 DPI)



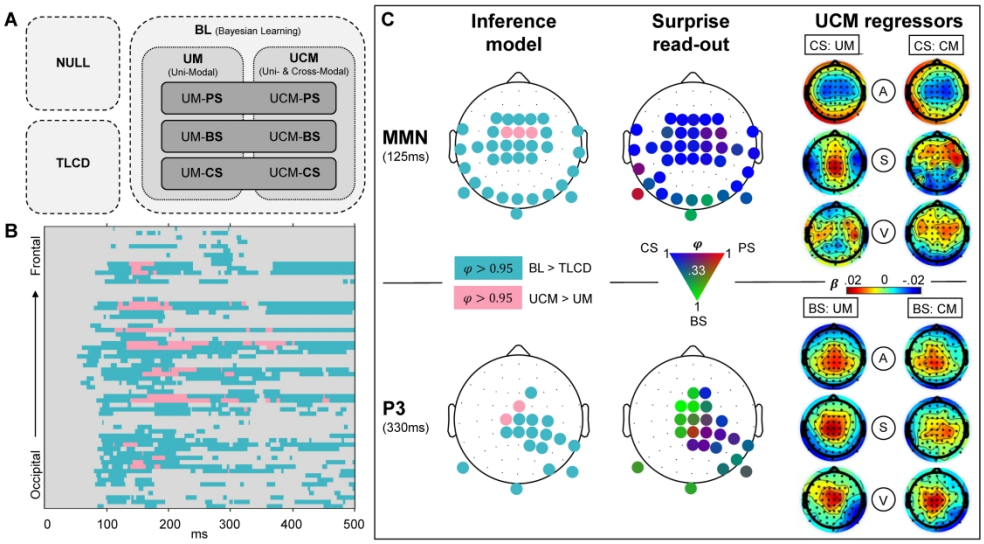
Cross-modal P3 effects. A) T-Maps of the conjunction of deviant>standard contrasts across the auditory, somatosensory and visual modalities. B) Beta estimates averaged across electrodes within significant clusters with peak p_fwe<0.05, resulting from two-way ANOVAs testing for differences between unpredictable, predicted and mispredicted deviants (red) and standards (blue).

152x127mm (300 x 300 DPI)



Source localisation. Top row: significant sources (p_fwe<0.05) for the auditory (purple), somatosensory (green) and visual (orange) MMNs as well as their conjunction (yellow). Bottom row: significant sources (p_fwe<0.05) for the conjunction (yellow) of the P3 MMR in the auditory, somatosensory and visual modalities.

635x609mm (150 x 150 DPI)



Modelling 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): Uni-modal 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 post-stimulus time points. Color depicts exceedance probability (EP) φ>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 φ>0.95). Middle column: Results of comparison 3. EPs between 0.33 and 1 of the three surprise functions are represented by a continuous 3-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=uni-modal) for CS read-out models (top) and BS read-out models (bottom).

330x190mm (300 x 300 DPI)

Original work: Study 3

Gijsen, S., Grundei, M. & Blankenburg, F. Active inference and the two-step task. Scientific Reports 12, 17682 (2022).

scientific reports

Check for updates

OPEN

Active inference and the two-step task

Sam Gijsen[1,2]✉, Miro Grundei[1,2] & Felix Blankenburg[1,2]

Sequential decision problems distill important challenges frequently faced by humans. Through repeated interactions with an uncertain world, unknown statistics need to be learned while balancing exploration and exploitation. Reinforcement learning is a prominent method for modeling such behaviour, with a prevalent application being the two-step task. However, recent studies indicate that the standard reinforcement learning model sometimes describes features of human task behaviour inaccurately and incompletely. We investigated whether active inference, a framework proposing a trade-off to the exploration-exploitation dilemma, could better describe human behaviour. Therefore, we re-analysed four publicly available datasets of the two-step task, performed Bayesian model selection, and compared behavioural model predictions. Two datasets, which revealed more model-based inference and behaviour indicative of directed exploration, were better described by active inference, while the models scored similarly for the remaining datasets. Learning using probability distributions appears to contribute to the improved model fits. Further, approximately half of all participants showed sensitivity to information gain as formulated under active inference, although behavioural exploration effects were not fully captured. These results contribute to the empirical validation of active inference as a model of human behaviour and the study of alternative models for the influential two-step task.

Sequential decision problems capture an important essence of the challenges regularly encountered by humans. Although the environmental structure and statistics often cannot be observed directly, they may be inferred through repeated interactions. Especially in case of a dynamically changing environment, a host of strategies have been proposed to underlie human decision making. Reinforcement learning is eminently cast as pursuing the long-term maximization of scalar reward[1]. This approach can either be *model-free* or *model-based*. A model-free approach states that action-selection proceeds based on the extent to which an action has been reinforced in the past. However, such a strategy ignores available knowledge about the environmental structure and is difficult to reconcile with goal-directed actions, consequently failing to capture important aspects of human behaviour[2]. In contrast, a model-based strategy is able to exploit structural knowledge in pursuit of goals and is therefore capable of predicting action outcomes.

The two-step task provides a prominent example of the application of reinforcement learning to the study of human decision-making[3]. The task requires the sequential traversal of two stages via binary action selection to accumulate rewards or avoid punishment. Specifically, the task was designed to disambiguate between model-free and model-based strategies. Model-based inference uses the probabilistic transition between the stages to steer itself towards lucrative states, while a model-free approach foregoes such transition-based planning and instead only relies on observed stimulus-action mapping. The two-step task has been highly influential and has seen widespread adoption, including the study of pathology such as obsessive-compulsive disorder[4] and gambling disorder[5]. This research has generally used the hybrid reinforcement learning model as introduced by Daw et al.[3], which combines independent model-free and model-based strategies. However, it has been shown that not all aspects of human behaviour on the two-step task are captured by this commonly-used reinforcement learning model[6]. Furthermore, it has been argued that the hybrid model may mischaracterise model-free behaviour as model-based[7], or vice versa[6,8].

Sequential decision problems additionally invoke the exploration-exploitation trade-off. Do we choose an option that is well-understood and known to be rewarding? Or should we risk foregoing this immediate reward so as to learn more about alternatives and in doing so potentially find an even more rewarding option? This conflict has been a prominent area of research in psychology[9], neuroscience[10,11], and computer science[1,12]. Exploration behaviour can result from stochasticity in action selection, randomizing choice rather than deterministically choosing the most rewarding action (random exploration). Additionally, action-values may not be purely reward-based but can receive an additional information bonus, biasing actions toward uncertain options (directed

[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. ✉email: sam.gijsen@fu-berlin.de

exploration). Directed exploration therefore describes an intentional process to minimize information discrepancies between options, which has so far only been observed in a subset of studies on human behaviour[11,13–15]. It has been a particularly powerful descriptor of behaviour in the domain of visual sensing, in which efficient exploration is likely the primary goal[16–18]. However, as directed exploration is not part of the hybrid reinforcement learning model, its potential role in the two-step task has not received much attention. Given that the two-step task features outcome-probabilities that drift over time, a model-based approach may benefit from an information gathering mechanism to promote exploration and discover rewarding actions. This begs the question whether further variance in behavioural data may be explained by modeling directed exploration dynamics.

Active inference has been proposed in neuroscience as a framework for describing the exploration-exploitation trade-off[19]. Derived from the free energy principle[20], active inference leverages the concept of a generative model that is iteratively optimized and allows for planning and decision making[21]. By taking a probabilistic inference approach, it is closely related to the Bayesian-brain hypothesis[22,23]. Action selection is cast as a minimization of *expected free energy*, which combines terms for both the realization of an agent's preferences and exploration. Under active inference, exploration results from seeking information gain, for which the generative model is used to infer the degree to which actions and their resulting observations may change belief distributions. Such a choice bias promotes directed exploration until the agent gains confidence in its understanding of its environment. At that point, further observations have a diminished impact on beliefs and the realization of preferences becomes a more dominant determinant of behaviour.

Active inference has a body of theoretical work[24,25] and has been studied in-silico[26–28]. However, empirical validation based on human-behaviour has only recently started to emerge. For example, active inference was used to characterise atypical choice behaviour in individuals with substance use disorder[29,30] and various other psychopathology[31,32]. Furthermore, model parameters fitted to human behaviour have been shown to correlate across time[33] and to potentially hold predictive power of future symptomatology[30]. Congruent with active inference predictions, human choice behaviour has also been shown to not merely be a function of reward or utility, but also entropy maximization[34,35]. Nevertheless, it is unclear whether a central feature of the framework, namely its proposed resolution to the exploration-exploitation dilemma, is able to capture human behaviour better than existing models in a variety of task settings. This is an important aspect of the framework, given that without the information-gain incentive, especially on simpler tasks, active inference can reduce to generate highly similar behaviour as a purely reward-maximizing reinforcement learning agent.

In the current study, we leverage the widely studied two-step task to investigate the suitability of active inference as a description of human choice behaviour. To this end, we use four publicly available datasets. We compare the behavioural predictions of active inference to those of the hybrid reinforcement learning model and perform Bayesian model selection analyses. By doing so, we contribute to the emerging, empirical study of active inference as well as investigate an alternative model for behaviour on the two-step task.

Methods

Participants and behavioural task. We studied human behavioural data in the two-step task originally designed by Daw et al.[3]. In this paradigm, participants first choose from two available initial-stage actions, each of which is uniquely associated with a likely ($p = 0.7$) and unlikely ($p = 0.3$) state transition to one of two final-stage states. There, another decision needs to be made between two final-stage actions, which yields a binary outcome. The outcome probability for each second-stage action follows an independent Gaussian random walk (Fig. 1).

The two-step task was initially designed to disambiguate between model-free and model-based strategies. A pure model-free strategy solely relies on learning from observed stimulus-action mapping. In contrast, model-based reasoning may use the known latent structure of transitions between initial-stage actions and final-stage states. The two-step task allows for the distinction between these two strategies by means of model comparison as well as analyses of averaged responses. The latter relies on the insight that model-free inference will lead to a greater probability to repeat an initial-stage action if it lead to the preferred outcome on the previous trial, independent of transition type. Model-based inference, in contrast, exploits the knowledge of state transitions and tends to repeat this initial-stage action only following a common transition. In case of a rare transition, the agent becomes more likely to switch to the other initial-stage action so as to increase its probability to access the promising, final-stage action.

The data was obtained from four publicly available datasets. These datasets were selected due to their similar task structure. First, 197 subjects participated in the online 'Daw two-step task' by Kool et al.[36]. Two further datasets were made available by da Silva and Hare[6]. These "Magic Carpet" (n = 24) and "Spaceship" (n = 21) experiments focused on providing intuitive and thorough instructions regarding all aspects of the task. Finally, the "Shock" dataset by Lockwood et al.[37] consists of 36 participants and differs from the aforementioned three experiments by using future electric shocks (or their absence) as the binary outcome rather than a monetary reward (or its absence). In addition, this study features two conditions in which participants are told the electric shocks will either be delivered to themselves ('self' condition) or another, anonymous participant ('other' condition). Finally, it used different parameters for the Gaussian random walk ($\mu = 0, \sigma = 0.2$, reflecting boundaries at [0, 1]) than the other datasets ($\mu = 0, \sigma = 0.025$, reflecting boundaries at [0.25, 0.75]). All participants were explicitly told about the task structure and received training on the task prior to data-collection. Please refer to the original manuscripts for full experimental descriptions.

Logistic regression analyses. Miller et al.[38] introduced and verified a logistic regression analysis of initial-stage choice behaviour based on behaviour, transitions, and outcomes on multiple preceding trials in the two-step task. The method thereby allows for insight into how the local history of these factors influences deci-

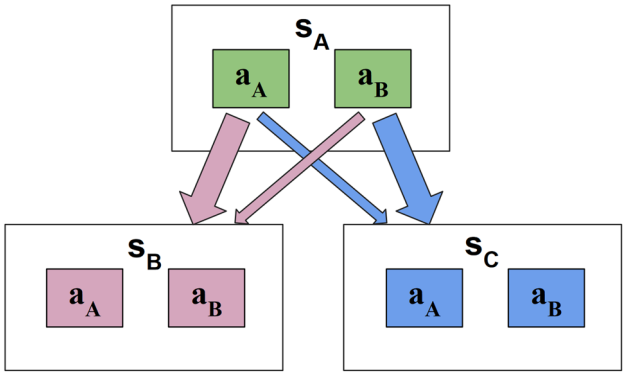


Figure 1. A graphical abstraction of the two-step task. Each trial always starts in the same (here, green) initial state s_A where participants choose between two green options. Each option is associated with a common and rare transition to one of two final-stage states. The green option a_A here has a 0.7 probability to move the participant to the pink s_B final state (common transition), and a 0.3 probability to transition to the blue s_C final stage (rare transition). The transition probabilities for the green option a_B are the opposite of those of a_A. Both final-stage states have two further options to choose from (pink a_A vs. a_B and blue a_A vs. a_B), which are associated with a binary outcome probability, each drifting independently over time The exact stimuli vary between studies.

sion making without relying on a computational model. Here we adopt this analysis to describe behaviour across datasets and to validate model predictions of behaviour. The analysis was shown to alleviate issues of the original regression analyses introduced by Daw et al.[3,38], which only considered the previous trial, by providing more accurate descriptions of behaviour and being less sensitive to learning rates. da Silva and Hare[6] used a variation of this method to describe aspects of behaviour that are not visible in case only the previous trial is considered and showed the shortcoming of the hybrid reinforcement learning model in capturing this behaviour. Specifically, da Silva and Hare[6] model the final-stage outcome in trial t as $o_t = +1$ for positive outcomes (monetary rewards or absence of shocks) and $o_t = -1$ otherwise (shocks or absence of monetary rewards), while transitions are coded as $\tau_t = +1$ for common transitions and $\tau_t = -1$ for rare transitions. The dependent variable y is the initial-stage action (i.e., which of the two stimuli was chosen) on trial t', coded as -1 and $+1$. Then, for each trial t of the $T = 4$ preceding trials, x_t denotes the initial-stage choice made on that trial. This yields the following regression model:

$$\log\left(\frac{p(y_t = +1)}{p(y_t = -1)}\right) = \sum_{t=1}^{T} \beta_0^{t'-t} x_{t'-t} + \beta_o^{t'-t} o_{t'-t} x_{t'-t} + \beta_\tau^{t'-t} \tau_{t'-t} x_{t'-t} + \beta_{o\times\tau}^{t'-t} o_{t'-t} \tau_{t'-t} x_{t'-t} \quad (1)$$

where the β-coefficients indicate the influence of variables in trial $t' - t$ on the initial-stage choice in trial t': $\beta_0^{t'-t}$ quantifies the influence of the initial-stage choice, $\beta_o^{t'-t}$ quantifies the effect of observation, $\beta_\tau^{t'-t}$ quantifies the effect of transition, and $\beta_{o\times\tau}^{t'-t}$ quantifies the interaction between observation and interaction. The model was fit for each participant separately using Scikit-learn[39]. To apply the analysis to the computational models and mitigate the randomness of choice selection, each participant's maximum likelihood parameters were used to simulate model behaviour 20 times. The resulting regression coefficients were averaged across iterations.

Computational modeling. In order to investigate the role of directed exploration on the two-step task, two families of models were implemented. A description of transition probability learning is followed by the common reinforcement learning approach and by a probabilistic framework implementing active inference with information-gain incentives. Parameters of both models are fit to participant behaviour and relative model performance is analysed via Bayesian model comparison. The initial-stage had only one state s_A (green in Fig. 1), while the final-stage had two possible states: s_B and s_C (pink and blue in Fig. 1). Each state had two actions, denoted a_A and a_B. On trial t, the initial-stage state is denoted by $s_{1,t}$ and the final-stage state $s_{2,t}$, and similarly actions by $a_{1,t}$ in the initial-stage and $a_{2,t}$ in the final-stage. Outcomes o_t refer to the binary, final-stage observations, consisting of the absence or presence of a monetary reward or a future electric shock depending on the dataset. As these only occur in the final-stage, they are always zero in the initial-stage.

Transition learning. The transition structure $p(s_{2,t}|s_{1,t}, a_{1,t})$ specifies how the two available initial-stage actions may transition the agent from $s_{1,t}$ to $s_{2,t}$. As each dataset included a familiarization and training phase prior to the start of the experiment, participants were aware that the transition probabilities of initial-stage actions were mirrored and could either be $p(s_B|s_A, a_A) = p(s_C|s_A, a_B) = 0.7$ or $p(s_B|s_A, a_A) = p(s_C|s_A, a_B) = 0.3$ (with $p(s_B|s_A, a_A) = 1 - p(s_C|s_A, a_A)$ and $p(s_B|s_A, a_B) = 1 - p(s_C|s_A, a_B)$). Consequently, which initial-stage action commonly led to which final-stage state was unknown at the start of the task and thus had to be inferred. Here, we model this transition learning similar to previous studies, by having agents count transitions and on each trial choose the most likely transition structure based on the observed frequencies. The three options included

the two possible true structures described above as well as flat ($p = 0.5$) transition probabilities[3,6,36]. This simple strategy settles on the correct solution after only a few trials and allows for the use of identical transition learning between the active inference and model-based reinforcement learning models. More sophisticated methods would have a limited contribution as the correct solution is found quickly and does not change across the experiment. As the identity of the states could be directly observed in all studies, the agents also have access to this information without the need for inference.

Hybrid reinforcement learning. The hybrid reinforcement learning model combines the action evaluations of model-free and model-based algorithms. The model was introduced by Daw et al.[3], who used it to quantify the relative contributions of these two strategies to behaviour. Both strategies map state-action pairs to their expected discounted future return, captured by $Q(s, a)$. The model-free strategy corresponds to the State-action-reward-state-action (SARSA(λ)) algorithm[40], which updates the value of state-action pairs s, a at stage $p = \{1, 2\}$ as follows:

$$Q_{MF}(s,a) = Q_{MF}(s,a) + \alpha_p \delta_{p,t} \lambda_{p,t}(s,a) \tag{2}$$

$$\delta_{p,t} = o_t + Q_{MF}\left(s_{p+1,t}, a_{p+1,t}\right) - Q_{MF}(s_{p,t}, a_{p,t}) \tag{3}$$

where δ is the outcome prediction error, α (α_1 or α_2 depending on the stage) is the learning rate, and λ is the eligibility parameter, modulating the effect of the final-stage prediction error on the values of initial-stage actions. Due to there not being an outcome on the initial-stage, the prediction error for this stage depends on the value of the selected final-stage action. The final-stage prediction error depends on the outcome o_t and thus the $Q-$ values for both stages are updated at the final stage, using the following errors:

$$\delta_{1,t} = Q_{MF}(s_{2,t}, a_{2,t}) - Q_{MF}(s_{1,t}, a_{1,t}) \tag{4}$$

$$\delta_{2,t} = o_t - Q_{MF}(s_{2,t}, a_{2,t}) \tag{5}$$

The model-based algorithm differs from the model-free approach by using knowledge about the transitions between initial-stage actions and final-stage states. Final-stage actions are evaluated directly from prediction errors as in model-free learning, however, the value of each initial-stage action a_j depends on its probabilistic mapping to final-stage states (and thereby to final-stage actions).

$$Q_{MB}(s_A, a_j) = p(s_B|s_A, a_j) \max_{a_2 \in \mathcal{A}_B} Q_{MF}(s_B, a_2) + p(s_C|s_A, a_j) \max_{a_2 \in \mathcal{A}_C} Q_{MF}(s_C, a_2) \tag{6}$$

where $\mathcal{A}_B$ and $\mathcal{A}_C$ are the sets of available actions in the respective final-stage states (s_B and s_C respectively).

The Q-values of the model-free and model-based algorithms are combined according to the weighting para meter w:

$$Q_{net}(s_A, a_j) = wQ_{MB}(s_A, a_j) + (1 - w)Q_{MF}(s_A, a_j) \tag{7}$$

Note that there is no need to weigh model-free and model-based estimates for final-stage choices as the algorithms do not differ there. Furthermore, the hybrid model includes pure model-free and model-based inference as special cases for $w = 0$ and $w = 1$, respectively. Next, an action is selected using a softmax operator:

$$p(a_{p,t} = a|s_{p,t}) = \frac{\exp(\beta_p Q_{net}(s_{p,t}, a) + \rho \times rep(a))}{\sum_{a'} \exp(\beta_p Q_{net}(s_{p,t}, a') + \rho \times rep(a'))} \tag{8}$$

where ρ is a commonly included parameter modeling initial-stage response stickiness and $rep(a')$ is 1 if a is the initial-stage action that was chosen in the last trial and 0 otherwise[36], and β is the inverse temperature parameter that controls the randomness of the action selection. Separate β_1 and β_2 parameters are fitted for each stage to allow for different levels of choice randomness.

Active inference. Active inference agents rely on a generative model of the task. The estimated transition probabilities (denoted by θ_1) and outcome probabilities (θ_2; please see below) are together denoted as θ, with the generative model taking the following form:

$$p(o_t, s_{2,t}|s_{1,t}, \theta) = p(o_t|s_{2,t}, \theta)p(s_{2,t}|s_{1,t}, \theta)p(\theta) \tag{9}$$

In the current study we focus on information-gain incentives of final-stage outcome probabilities θ_2 and omit state inference incentives from active inference due to the static transition probabilities θ_1 and observable state identity. Note that if it is assumed for participants following training to be aware that the transition structure is one of two mirrored options ($p = [0.3\ 0.7]$ or $p = [0.7\ 0.3]$), then both initial-stage actions provide equal amounts of information about transition probabilities. As a result, action-selection will only be sensitive to information discrepancies about outcome probabilities.

Under active inference, agents resolve the exploration-exploitation dilemma by basing action selection on a single expression. Actions are more probable to be selected if they minimise expected surprise about future observations[21], that is, if they minimise expected free energy. This quantity can be expressed in variety of ways, with an intuitive decomposition featuring extrinsic and intrinsic value terms[21]:

$$G_t(a) = -\underbrace{E_{p(o_t;\pi_t(\theta_2)|a_t=a)}\left[\ln p(o_t|C)\right]}_{\text{Extrinsic Value}} - \underbrace{E_{p(o_t;\pi_t(\theta_2)|a_t=a)}[D_{KL}(\pi_t(\theta_2)|o_t, a_t = a\|\pi_t(\theta_2))]}_{\text{Intrinsic Value}} \tag{10}$$

where $p(o_t|C)$ denote the prior preferences over outcomes and D_{KL} is the Kullback-Leibler divergence, here between beliefs $\pi_t(\theta_2)$ about final-stage outcome probabilities θ_2 before (prior) and after (posterior) hypothetically observing an outcome resulting from action selection. $p(o_t; \pi_t(\theta_2)|a)$ can be understood as the distribution obtained from $p(o_t|\theta_2, a)\pi_t(\theta_2)$ once θ_2 has been marginalised out, noting that $\pi_t(\theta_2|a) = \pi_t(\theta_2)$ by construction. For a given action, the extrinsic value term is a measure of how likely prior preferences are to be attained, while the intrinsic value term quantifies the expected information gain. The formulation above suffices for final-stage action selection, however, for the initial-stage the action-dependent state transitions need to be accounted for. As state inference was not included, this leads to the simplified setting wherein both prior preference realization and information-gain is limited to the final-stage actions. For a multi-step policy, we formulate the computation of the expected free energy for an initial-stage action a_j. $G(a_j)$ depends on the probabilistic mapping of the action to final-stage states s_2 and, by extension, the final-stage actions that are available in these states. Specifically, the estimated action-specific transition probabilities are multiplied by the expected free energy of the associated final-stage actions:

$$G(a_j) = p(s_B|s_A, a_j, \theta_1) \sum_{a_2 \in \mathcal{A}_B} G(a_2) + p(s_C|s_A, a_j, \theta_1) \sum_{a_2 \in \mathcal{A}_C} G(a_2) \tag{11}$$

where $\mathcal{A}_B$ and $\mathcal{A}_C$ are the sets of available actions in the corresponding final-stage states (s_B and s_C respectively).

The prior preferences capture the relative attractiveness of the different outcomes, with desired outcomes being assigned higher probabilities. Here we constrain the prior preferences regarding action outcomes to a Bernoulli distribution implying $o_t = 1$ is preferred over $o_t = 0$ following Marković et al.[27]:

$$P(o_t|C) = \frac{1}{Z(\lambda)} e^{o_t \lambda} e^{-(1-o_t)\lambda}. \tag{12}$$

The λ-parameter specifies the precision of the prior preferences. For $\lambda = 0$ (i.e., zero precision), the outcomes are valued equally and the agent will thus only maximize intrinsic value, corresponding to pure information gathering about the outcome probabilities encoded by $\pi_t(\theta_2)$. As λ increases, the agent will value information-gain less and focus more on realizing prior preferences, thereby becoming increasingly risk-seeking. This precision parameter thus balances exploratory and exploitative behaviour.

Previous studies have modeled a tendency for participants to repeat initial-stage actions independent of the outcome[36]. Such behaviour may be modelled under active inference as a habit, which we here assume to be static across the experiment for simplicity, again constrained to a Bernoulli distribution:

$$E_a(a_j) = \frac{1}{Z(\kappa)} e^{\delta_{a_{t-1},a_j}\kappa} e^{-(1-\delta_{a_{t-1},a_j})\kappa} \tag{13}$$

with precision parameter κ and $E_a(a)$ always set to zero for final-stage actions. Action selection at stage $p = \{1, 2\}$ may then proceed by applying a softmax operation (σ) to the expected free energies, together with the habitual bias:

$$p(a_{p,t}) = \sigma\left[-\gamma_p G(a_{p,t}) + E_a\right] \tag{14}$$

with γ_p functioning as an inverse temperature parameter, controlling the stochasticity of action selection. As in the hybrid model, two separate γ parameters are fit to participant data, allowing for different levels of randomness in initial- and final-stage action selection.

Finally, previous studies on multi-armed bandit tasks have found evidence for a bias in prior outcome probabilities[41]. To capture any such biases and their effects on action selection, the mean ($E[\pi_0(\theta_2)] = \frac{\alpha_0}{\alpha_0+\beta_0}$) of the prior Beta distribution is included as a free parameter in the model fitting procedure (please see below).

Next, we describe the learning rule of observations for the active inference models. This necessarily deviates from the hybrid reinforcement learning model as the aforementioned computations require probability distributions, rather than point estimates. Note that this only concerns the learning of final-stage outcome probabilities, as transition learning is shared across all models as detailed above. Liakoni et al.[42] introduced surprise-based learning algorithms for changepoint paradigms that feature occasional resampling of the environmental statistics. We will briefly describe one such algorithm and subsequently modify it to better suit the current environment with drifting parameters.

For brevity, references to s and a have their subscripts dropped as they will generally refer to the final-stage. Wherever the initial-stage state or actions are concerned, these will be explicitly denoted by s_1 and a_1. This process of observation emission corresponds to sampling from a Bernoulli distribution parameterized by an expectation $\theta_{2,a}$ for each final-stage action, encoding the probability of observing $o_t = 0$ or $o_t = 1$. A Bayesian agent requires a prior distribution over the estimated θ_2 probabilities, for which conjugate Beta priors are appropriate:

$$p(\theta_2) = \prod_{s=1}^{2} \prod_{a=1}^{2} \mathcal{B}e(\alpha_{s,a}, \beta_{s,a}) \tag{15}$$

In such a setting, Bayesian inference corresponds to the following simple update rules for the parameters of the Beta distributions:

$$\alpha_{s,a} = \alpha_{s,a} + \delta_{a_t,a} o_t \tag{16}$$

$$\beta_{s,a} = \beta_{s,a} + \delta_{a_t,a}(1 - o_t) \tag{17}$$

with δ being the Kronecker delta, which is 1 if both variables are equal, and 0 otherwise. However, such updating will quickly lead to inflexible beliefs, making it unsuited for dynamic environments. Liakoni et al.[42] propose an updating scheme based on the Bayes-factor surprise $S_{BF} \geq 0$, a ratio between the subjective probability of an observation under the current beliefs and prior beliefs. Together with a prior belief of the volatility of the environment, $\nu \in [0, 1]$, this surprise quantity enables a simple learning rule that moves the current belief distribution over final-stage outcome probabilities $\pi_t(\theta_2)$ closer to the uninformed prior beliefs $\pi_0(\theta_2)$ (here: $\mathcal{B}e(\alpha_0, \beta_0)$) based on the degree to which current observations are more likely under these prior parameters. This is achieved by scaling concentration parameters with a surprise-modulated adaptation rate $\chi \in [0, 1]$. The rate for the current trial χ_t may be computed as follows:

$$\chi_t = \chi(S_{BF}, m) \tag{18}$$

$$\chi(S, m) = \frac{mS}{1 + mS} \tag{19}$$

$$\begin{aligned} S_{BF} &= \frac{p(o_t; \pi_0(\theta_2))}{p(o_t; \pi_t(\theta_2))} \\ m &= \frac{\nu}{1 - \nu} \end{aligned} \tag{20}$$

$$p(o_t; \pi_t(\theta_2)) = \begin{cases} \frac{\alpha_{s,a,t}}{\alpha_{s,a,t} + \beta_{s,a,t}} \text{ if } o_t = 0 \\ \frac{\beta_{s,a,t}}{\alpha_{s,a,t} + \beta_{s,a,t}} \text{ if } o_t = 1 \end{cases} \tag{21}$$

where $m \geq 0$ depends on volatility ν and modulates the effect of surprise on learning and $p(o_t; \pi(\theta_2))$ refers to the subjective probability of observing o_t under the belief $\pi(\theta_2)$, which is easily computed using the Beta-parameters α and β. Liakoni et al.[42] continue to derive the following update rules for the Beta-distributions:

$$\alpha_{s,a,t} = (1 - \chi_t)\alpha_{s,a,t-1} + \chi_t \alpha_0 + \delta_{a_t,a} o_t \tag{22}$$

$$\beta_{s,a,t} = (1 - \chi_t)\beta_{s,a,t-1} + \chi_t \beta_0 + \delta_{a_t,a}(1 - o_t) \tag{23}$$

Effectively, surprising events shrink the concentration parameters of $\pi_t(\theta_2)$ towards those of $\pi_0(\theta_2)$, causing previous observations to be forgotten and thereby increasing the effect current observations have on the belief distribution, enabling flexible learning. In the current drift paradigm we are not interested in weighing between a changepoint or continuation, but rather the degree to which the generative probability has diffused and is thereby incompatible with currently held beliefs. This discrepancy between beliefs and the world can be quantified by predictive surprise ($PS(o_t) \geq 0$), which has a significant relevance in behavioural and imaging neuroscience[43–46] as well as active inference[19]. The influence of surprise on beliefs may then be mediated by a prior volatility parameter $\nu_{PS} \in [0, 1]$:

$$\chi_t = \chi(PS, m) \tag{24}$$

$$\chi(S, m) = \frac{mS}{1 + mS} \tag{25}$$

$$PS = -\ln p(o_t; \pi_t(\theta_2)) \tag{26}$$

$$m = \frac{\nu_{PS}}{1 - \nu_{PS}} \tag{27}$$

The update rules for the parameters corresponding to sampled actions then remain similar:

$$\alpha_{s,a,t} = (1 - \chi_t)\alpha_{s,a,t-1} + \delta_{a_t,a} o_t l \tag{28}$$

$$\beta_{s,a,t} = (1 - \chi_t)\beta_{s,a,t-1} + \delta_{a_t,a}(1 - o_t) l \tag{29}$$

where l indicates a learning rate that may be fitted per participant and thus does not have to equate to 1. As probabilistic learning models are understudied in the context of the current paradigm, we also consider the possibility that beliefs decay independent of observations, akin to static forgetting. This may apply to beliefs of sampled actions, instead of or in addition to surprise-based learning, and unsampled actions, where it signals a gradual loss of confidence in beliefs about unexplored options, independently of $PS(o_t)$:

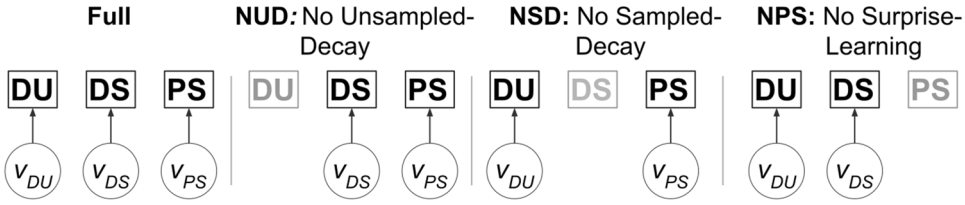


Figure 2. Graphic summary of learning models. Four variants of the learning model are compared, differentiated by their forgetting kinetics with associated prior volatility parameters ν. The NUD and NSD models omit decay of concentration parameters of beliefs for unsampled and sampled actions respectively (corresponding to $\nu_{UD} = 0$ and $\nu_{SD} = 0$), while the NPS model excludes surprise-based learning ($\nu_{PS} = 0$). The more complex Full model incorporates all three learning dynamics.

$$\alpha_{s,a,t} = (1 - \nu)\alpha_{s,a,t-1} + \nu\alpha_0 \tag{30}$$

$$\beta_{s,a,t} = (1 - \nu)\beta_{s,a,t-1} + \nu\beta_0 \tag{31}$$

where ν indicates a prior volatility parameter, with separate parameters $\nu = \{\nu_{SD}, \nu_{UD}\}$ for sampled and unsampled actions respectively. Here, beliefs are assumed to decay back to the prior distributions over time rather than shrinking to zero, as very small concentration parameters may lead to computational instability.

We formulate variations of the learning model that allow for surprise-based learning, static forgetting implemented as decay of concentration parameters, or both (Fig. 2). The rates of these mechanisms are governed by free volatility parameters. The resulting model-based hypotheses may be compared via model-comparison. Many further model variants may be hypothesized, including the possibility of 'shared' parameters between forgetting and surprise-learning. We limit the analysis to a smaller amount of models for two reasons. First, the binary reward structure of the task was deemed likely to provide insufficiently detailed behavioural data to dissociate large numbers of highly similar models. Second, the current work aims to understand the explanatory power of active inference in the two-step task, rather than an exhaustive study of underlying learning dynamics. A comparison between potential key features of learning (decay and surprise-based learning) was performed to approximate an adequate level of flexibility to enable inference on the action-selection level. Hereby we intend to ameliorate the impact of potential complex interactions between learning dynamics and action-selection on model fitting.

Model fitting and model comparison. In order to fit the free parameters of the computational models to the participant choice data we used a constrained minimization algorithm ('L-BFGS-B' as implemented in Scipy[47]). To mediate the problem of local optima, the optimization was ran 25 times for each participant with different (uniformly) randomized initializations for all parameters. The iteration that yielded the highest log likelihood was used for the model comparison procedures.

Fixed-effects model comparisons were performed using the Akaike's Information Critria (AIC) and Bayesian Information Criteria (BIC).

$$BIC := k\ln(n) - 2\ln(\hat{L}) \tag{32}$$

$$AIC := 2k - 2\ln(\hat{L}) \tag{33}$$

with k being the number of free parameters in the model, n the amount of trials, and $\hat{L}$ denoting the maximized value of the subject- and model-specific log likelihood function. These approximations of log model evidence were then subjected to random-effects Bayesian model selection as implemented in SPM (*spm_BMS.m*; Wellcome Trust Centre for Neuroimaging, Institute for Neurology, University College London, London, UK). This algorithm yields model-specific exceedance probabilities, representing the probability that the model has a higher frequency than the other included models on the group-level. The method additionally provides protected exceedance probabilities, which are more conservative by accounting for the possibility that apparent differences in model frequencies arise due to chance[48]. Models that share certain characteristics may be grouped into 'families', and the subsequent comparison of families rather than individual models allows for inference on the contribution of these characteristics to the model fit[49]. This analysis is implemented in SPM's *spm_compare_families.m*, for which we set priors to be equal across families rather than models ("F-unity" priors). As protected exceedance probabilities are not available for family-level inference, we present exceedance probabilities for these analyses.

To investigate the strengths and shortcomings of models in their ability to describe participant data, the models were used to simulate behaviour on the two-step task. The subject-specific parameters that maximized the log likelihood, as described above, were used to simulate action selection on the two-step task for 20 independent runs. Identical logistic regression analyses were then applied to this synthetic choice data, of which the resulting beta-coefficients allow for a comparison to the participant-derived coefficients. For λ-parameter reliability analyses, these synthethic datasets were subsequently subjected to the parameter-fitting procedure. The recovered λ-parameters were then compared with the parameter values fitted to the subject's actual choice data.

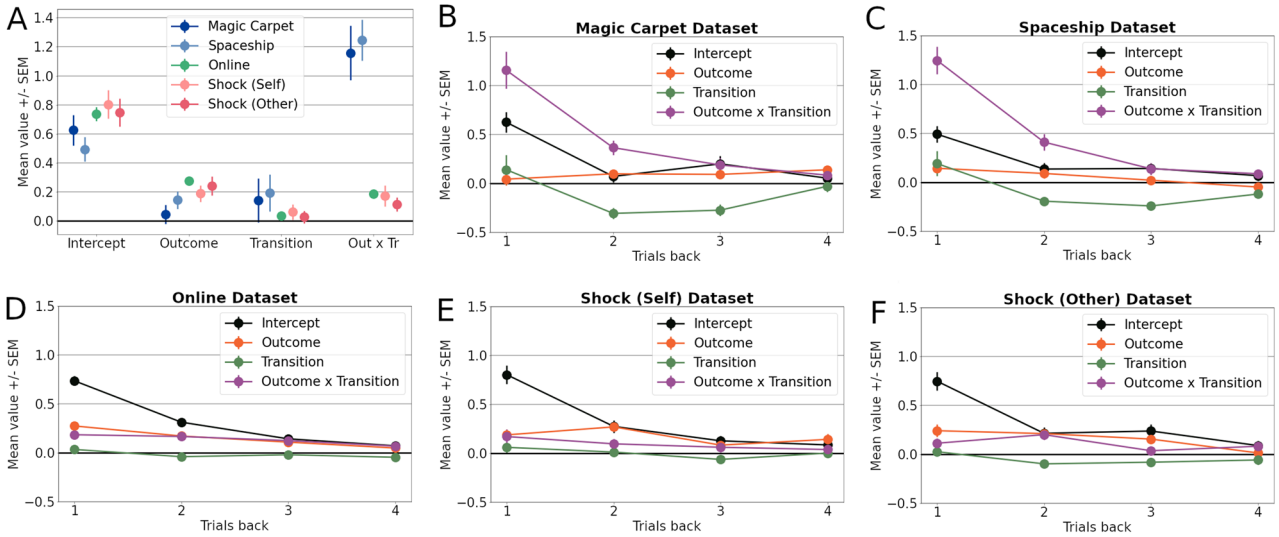


Figure 3. Logistic regression analyses. Regression analysis of first-stage actions based on the previous four trials. (**A**) Regression coefficients for the intercept, main effects of outcome and transition, and outcome-transition interaction only on the preceding trial across the different datasets. (**B**–**F**) Each subplot shows the coefficients for one dataset on all four preceding trials.

Results

Logistic regression analyses. By inspecting the regression coefficients resulting from the logistic regression analyses, considerable differences in participant behaviour were observed across the datasets (Fig. 3). Most notably, behaviour in the Magic Carpet and Spaceship paradigms showed a greater outcome-transition interaction, a proxy of model-based inference. They additionally showed larger main effects of transition: small positive coefficients for the previous trial, with larger negative coefficients for behaviour two and three trials back. This indicates that common transitions lead participants to actively switch to the other initial-stage action independent of trial outcomes, potentially indicating information-seeking behaviour[6]. These effects were largely absent in the remaining datasets. Main effects of outcome were relatively small in the Magic Carpet and Spaceship tasks, while the intercept terms were comparable across all included datasets.

Comparing hybrid reinforcement learning and active inference. The hybrid reinforcement learning (Hybrid-RL) model and the active inference model were compared in their ability to explain participant choice data via Bayesian model selection. To do so independently of the variations of the probabilistic learning model displayed in Fig. 2, these were clustered into a family and compared against a family containing only the Hybrid-RL model. The active inference family was found to describe single-trial participant behaviour better for the Magic Carpet and Spaceship datasets (both exceedance probabilities $\phi > 0.99$), with consistent results between AIC (expected posterior probabilities of $\langle r_{AI} \rangle = [0.86, 0.89]$ respectively) and BIC ($\langle r_{AI} \rangle = [0.82, 0.87]$ respectively) (Fig. 4A,C). For the other datasets, the metrics did not agree on the best performing family. Using AIC, the Online and Shock datasets slightly favoured Hybrid-RL ($\langle r_{RL} \rangle = [0.57, 0.50, 0.59]$), while the active inference family scored better using BIC ($\langle r_{AI} \rangle = [0.68, 0.66, 0.65]$).

The different learning models were subsequently compared against one another (Fig. 4B,D). In multiple cases, there was ambiguity about the best scoring model. For the Magic Carpet and Spaceship datasets, only AIC provided strong evidence for a model: the "Full"-variant with both belief decay and surprise-based learning (protected exceedance probabilities $\tilde{\phi} = [0.98, 0.98]$ respectively). Using BIC, model scores were distributed across the Full, NSD, and NPS model variants. For the Online and Shock datasets, the simpler 'NSD'-model which omits decay of concentration parameters for the sampled action tended to score highest. However, this was only clearly the case for the Online dataset, with $\tilde{\phi} > 0.99$ using both AIC and BIC. Expected posterior probabilities were quite widely distributed over the Full, NSD, and NPS variants. This was also found for the Shock dataset, indicating that the best fitting learning model differed across subjects. As the Online dataset contained more subjects, exceedance probabilities were still high due to greater confidence a majority of subjects used the NSD variant. The NUD variant, omitting decay of concentration parameters of unsampled actions, scored poorly across all datasets and metrics. This indicates the importance for a decay of concentration parameters of beliefs about unsampled actions, likely to appropriately capture the behavioural flexibility observed in participants. Overall, a dissociation between learning models was incomplete.

Simulation analysis. To gain insight into the extent to which the models capture participant behaviour, we simulated model behaviour on the two-step task. The maximum likelihood parameter estimates were used and the resulting synthetic datasets were submitted to the logistic regression analyses (Fig. 5A). The Full model was used for the Magic Carpet and Spaceship datasets, while the NSD variant was chosen for the Online and Shock

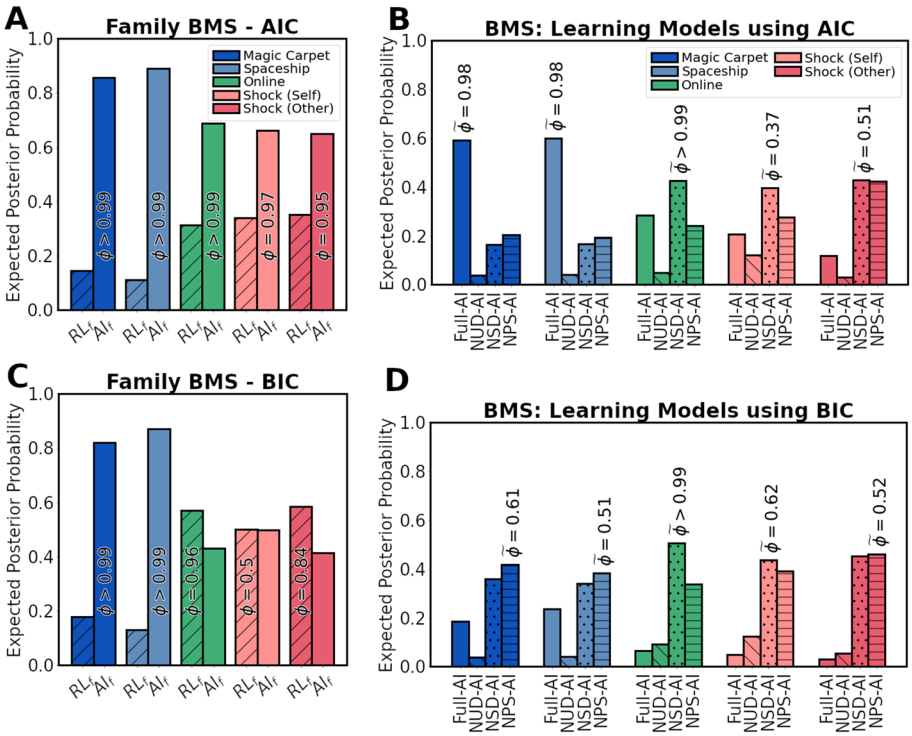


Figure 4. Bayesian model comparison. (**A**,**C**) Expected posterior probabilities resulting from the Bayesian Model Comparison between the RL_f family containing the hybrid-reinforcement learning model and the AI_f family, containing the active inference models with varying learning dynamics (see Fig. 2) using AIC (**A**) and BIC (**C**). ϕ indicates the exceedance probability in favour of the best-scoring family. (**B**,**D**) Expected posterior probabilities for the different learning models contained within the AI_f family. $\tilde{\phi}$ indicates the protected exceedance probability of the best-scoring model.

datasets. On the Magic Carpet and Spaceship datasets, the active inference model appears to better capture the strong interaction between outcome and transition type, while it is underestimated by the Hybrid-RL model. The active inference model displays minor main effects of transition, although not to the extent that they are present in the participant data, while being absent in the Hybrid-RL model. For the Online and Shock datasets, the smaller interaction terms and close-to-zero interaction terms are reproduced by both models. The main-effects of outcome are reproduced by the Hybrid-RL model, but not by active inference.

Correlation analyses. To follow up on the differences between the model families, we inspected potential contributions to differences in model fit. Specifically, as regression analyses suggest that behavioural datasets which are better explained by active inference show both greater model-based inference as well as potential directed exploration, we investigated which of these correlated with relative model fit. We additionally compute correlations between the λ-parameter and the main-effects of transition to understand whether this averaged-based metric indeed tends to be greater for subjects sensitive to information-gain as identified by a low λ-parameter.

The correlation analyses between single-subject relative model fits and the interaction term between outcome and transition type (as a proxy for model-based inference) were positive and larger for the Magic Carpet (Pearson's $r = 0.71$, $p < 0.001$) and the Spaceship ($r = 0.63$, $p = 0.002$) datasets (Table 1). For the Online and Shock (Other, Self) datasets, correlation coefficients were lower ($r = [0.19, 0.18, 0.08]$, $p = [0.009, 0.29, 0.63]$ respectively). As the λ-parameter was not normally distributed, we adopt two common approaches. First, we offset the data such that the lowest value was 1, allowing for a log-transformation of the data and computing Pearson's r (r_p). Second, we used the data without transformation to compute the non-parametric, rank-based correlation measure Spearman's r (r_s). Repeating the aforementioned analyses for the λ-parameter and relative model fit, a similar contrast between datasets appeared. Modest correlations were observed in the Magic Carpet ($r_p = -0.35, p = 0.09, r_s = -0.14, p = 0.51$) and Spaceship ($r_p = -0.45, p = 0.04, r_s = -0.40, p = 0.07$) datasets. Meanwhile, the correlations were smaller for the Online and Shock datasets ($r_p = [0.10, 0.18, 0.08]$, $p = [0.16, 0.29, 0.66]$ and $r_s = [0.09, 0.11, -0.10]$, $p = [0.26, 0.53, 0.56]$, respectively). The correlations between the λ-parameter and the transition-coefficients were again somewhat larger for the Magic Carpet ($r_p = 0.29, p = 0.16$, $r_s = 0.26, p = 0.23$) and Spaceship ($r_p = 0.55, p = 0.01, r_s = 0.34, p = 0.13$) tasks, compared to the Online and Shock datasets ($r_p = [0.14, -0.08, -0.07]$, $p = [0.06, 0.63, 0.70]$ and $r_s = [0.09, 0.07, -0.17]$, $p = [0.23, 0.70, 0.33]$, respectively).

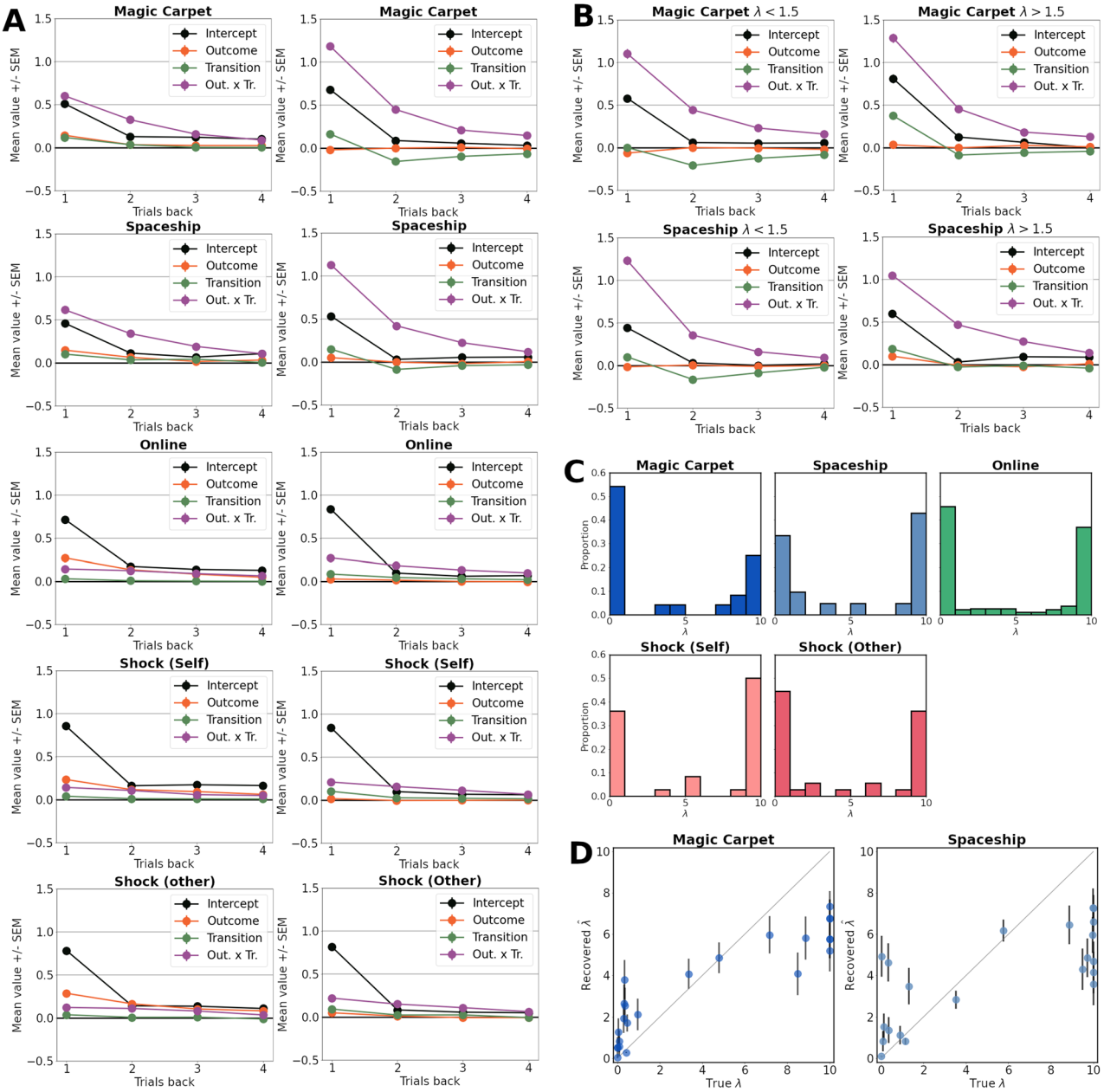


Figure 5. Simulation analyses. (**A**) The Hybrid-RL (Left) and active inference (Right) models were used to simulate behaviour on the two-step task using the maximum-likelihood parameters for each participant. The resulting data were subjected to the same logistic regression analyses as the participant data of Fig. 3. (**B**) Data was simulated for participants with low (Left) and high (Right) values of λ separately. The coefficients resulting from the logistic regression are plotted per group. (**C**) The recovered values of the λ-parameter per dataset, displaying highly bimodal distributions. (**D**) A comparison of true but unknown λ-parameter values and the recovered values. Each datapoint depicts one participant.

Evidence for directed exploration. Next, the role of the λ-parameter was further investigated. The distribution of parameter values resulting from the maximum-likelihood procedure is displayed in Fig. 5C. In every dataset, a bimodal distribution was found, with a majority of participants having a λ-parameter of smaller than one or close to the upper bound of ten (Fig. 5C). To analyse the effect of the λ-parameter on the main-effect of transition in the regression analysis directly, participants were stratified into two groups: a low λ group ($\lambda < 1.5$) and a high λ group ($\lambda > 1.5$). The associated model parameters were subsequently used to simulate group-specific behaviour on the two-step task. The low λ group displayed moderately stronger transition effects than the high λ group (Fig. 5B). Nevertheless, even this low λ group did not exhibit transition effects to the extent that these were observed in behavioural participant data.

In an exploratory analysis, we compared whether task performance in terms of average obtained reward differed between the low and high λ groups. A Welch's t-test was used due to unequal

Paradigm	corr($\beta_{o\times\tau}^{t'-1}$, $\Delta\hat{L}$), (p)	corr(λ, $\Delta\hat{L}$), (p)	corr(λ, $\beta_{\tau}^{t'-2,3}$), (p)
Magic Carpet	**0.71** (< 0.001)	− 0.35 (0.09)/− 0.14 (0.51)	0.29 (0.16)/0.26 (0.23)
Spaceship	**0.63** (0.002)	− **0.45** (0.04)/− 0.40 (0.07)	**0.55** (0.01)/0.34 (0.13)
Online	**0.19** (0.009)	0.10 (0.16)/0.09 (0.26)	0.14 (0.06)/0.09 (0.23)
Shock (Self)	0.18 (0.29)	0.18 (0.29)/0.11 (0.53)	− 0.08 (0.63)/0.07 (0.70)
Shock (Other)	0.08 (0.63)	0.08 (0.66)/− 0.10 (0.56)	− 0.07 (0.70)/− 0.17 (0.33)

Table 1. Correlation analyses. Correlation coefficients and uncorrected, associated p values. Due to λ not being normally distributed, both Pearson's r using log-transformed data (left) and Spearman's r (right) using untransformed data are presented for its analyses. $\beta_{o\times\tau}^{t'-1}$: regression beta-coefficient for the interaction between outcome and transition type for the preceding trial, $\Delta\hat{L}$: difference in values of maximized log likelihood function for the Hybrid-RL and best fitting active inference model, $\beta_{\tau}^{t'-2,3}$: beta-coefficient encoding the main effect of transition type at the second and third previous trials. Significant values are in [bold].

variances and sample sizes for some datasets. The high λ group obtained slightly more reward on most datasets ([+ 11.9%, + 4.0%, + 1.3%, − 0.5%, + 3.3%] for the Magic Carpet, Spaceship, Online, Shock (Self) and Shock (Other) datasets respectively), with uncorrected $p < 0.05$ for the Magic Carpet dataset ($t(p) = [-2.28(0.03), -0.49(0.47), -1.04(0.29), 0.12(0.91), -0.94(0.35)]$).

Finally, we verified that the λ-parameter could be accurately recovered given the constellation of model parameters as identified in the participant sample. Synthetic choice data was repeatedly generated using the set of parameters of every subject, which subsequently underwent parameter fitting. The results of this recovery analysis for the λ-parameter is displayed in Fig. 5D. Correlations between true and recovered parameter values were high for both the Magic Carpet ($r_p = 0.85, p < 0.0001, r_s = 0.90, p < 0.0001$) and Spaceship tasks ($r_p = 0.74, p = 0.0001, r_s = 0.60, p = 0.004$). Parameter values were often recovered accurately for low values of λ, although especially in the Spaceship dataset this was not successful for all participants. Many of the high λ -parameter values were underestimated, which likely stems from increases in λ having diminished effects. As λ increases past values of approximately 3, the expected free energy is already heavily biased towards preference maximization. For example, due to the linear scaling an increase in λ from 0.1 to 0.5 is comparable in size to an increase from 2 to 10. The large upper bound on λ of 10 was to ensure a pure preference-realization strategy could be captured by the model-fitting procedure and not due to the idea that variations in large λ values would meaningfully change behaviour. Overall, for most participants, the λ-parameter could be recovered successfully, although with considerably noise and inaccuracy for multiple subjects.

Discussion

In this paper, we collected and analysed datasets of human participants performing the two-step task and compared the explanatory power of the active inference framework to that of the standard hybrid reinforcement learning (Hybrid-RL) model. Logistic regression analyses revealed marked differences in initial-stage choice behaviour of participants between datasets, with two datasets exhibiting markedly greater model-based inference. Bayesian model comparison indicated that only these two datasets were significantly better described by active inference. The degree to which participants used model-based inference was found to strongly correlate with the relative performance of the models, while the exploration-exploitation parameter λ showed only a small to moderate relationship. Model simulations indicated that the probabilistic learning underlying active inference likely contributed to the better model fits and that exploration behaviour of subjects was better captured by active inference, albeit only partially. This was the case even when analysing only those subjects who were classified as most strongly pursuing directed exploration.

The Magic Carpet and Spaceship datasets featured comparatively small effects of outcome yet stronger effects of both transition type and the interaction of outcome and transition type. This replicates the results reported by da Silva and Hare[6], suggesting a greater degree of model-based inference in these datasets compared to the Online and Shock tasks. The Shock study[37] provided thorough instructions to participants, as did the Online study[36]. However, neither provided detailed explanations for all aspects of the task, such as transition type. da Silva and Hare[6] argued that providing intuitive reasons for task dynamics aided participants in adopting a more accurate model of the task. This is plausible as task knowledge is known to be able to influence the task model[50], with direct comparisons having been performed for the two-step task specifically[51]. Although the current work does not aim to explain why behaviour differed between tasks, we recover a similar behavioural discrepancy.

By partitioning model space, we were able to compare the Hybrid-RL model to active inference independent of the specific learning model. The discrepancy between datasets also persisted on this level, as we showed strong evidence in favour of active inference for the Magic Carpet and Spaceship datasets, yet there was no consensus between AIC and BIC for the Shock and Online datasets. As BIC penalizes models for their complexity more than AIC does, a lack of consensus implies that the additional parameters in the active inference models did not sufficiently increase the fit of the model. As such, additional caution is warranted when inspecting the comparison of the individual active inference models for the Shock and Online datasets.

Overall, we were unable to convincingly dissociate between learning models, indicated by the lack of consensus between metrics. Nevertheless, on the Magic Carpet and Spaceship tasks, the more complex, full model scored well using AIC, combining both decay and surprise-based learning for chosen actions. A trend was observed where a simpler model without decay of concentration parameters for sampled actions scored best for

the Online and Shock datasets. This model only uses surprise-based learning to update beliefs about outcomes of chosen actions, although a model with only decay and without surprise-based learning scored similarly. These results suggest that learning dynamics are the result of only decay or surprise-based learning, but not both. The high protected exceedance probabilities but low posterior probabilities for the Online dataset likely result from the large amount of participants. This combination suggests a very high probability that a model without decay is most prevalent in the sample, yet behaviour of many participants is best described by other models. A consistent finding, however, is the very low scoring of the model without decay of beliefs about unsampled actions. This may highlight an important distinction between the learning rule based on probability distributions and the Hybrid-RL model. Using probabilistic learning, sampling an action increases the concentration parameters of a belief distribution, often equated to increases in confidence[52]. In absence of a decay or forgetting mechanism, these higher concentration parameters impede flexible behaviour when an action is revisited at a later time and appears rewarding. Thus, the poor scores of this model imply the use of such a decay mechanism by participants in the Magic Carpet and Spaceship tasks. Due to the lower scoring of active inference in the remaining datasets, it could alternatively indicate that rather than using belief distributions, participant behaviour on the Online and Shock datasets is best described as learning point estimates (such as Q-values in reinforcement learning), which intrinsically allow for continued flexible behaviour. Nevertheless, forgetting mechanisms are commonplace in cognitive and neuroscience[46,53,54] and have been previously shown to improve model fits on the two-step task[55,56].

To gain insight into what drives the relative model scores, we simulated behaviour using the Hybrid-RL and best fitting active inference models. The strong positive interactions between outcome and transition type observed with the Magic Carpet and Spaceship tasks were reproduced by the active inference model, but not the Hybrid-RL model. This interaction has generally been interpreted to show model-based inference[3], but using the current analyses by Miller et al.[38] shows the progression of this term over trials and thereby reveals further information about the learning dynamics. Besides the strength of the interaction, its relative and diminished influence on preceding trials shows the sensitivity of beliefs to the history of observed outcomes. As such, it is mainly interpreted as a reflection of underlying learning dynamics rather than a function of action selection. The ability to capture this interaction term accurately thus further shows promise for the role of probability distributions to model learning on the two-step task, displaying a better description of behaviour than the Q-value learning strategy of the Hybrid-RL model. This finding fits with the general idea underlying the Bayesian brain hypothesis[23]. However, this advantage does not extend to the Online and Shock datasets, as the weaker interaction-terms are reproduced by both models.

The main effects of outcome have been interpreted as a proxy of model-free inference[3]. These effects, which are especially prevalent on the Online and Shock datasets, are not captured by the active inference model as the underlying learning rule we used is purely model-based. This discrepancy between human behaviour and active inference simulations likely contributes to the lower active inference scores on the model comparison analyses for these two datasets. Although active inference often receives a (generative) model-based treatment in the literature[21,57,58] and is therefore the focus of the current work, the framework is not incompatible with multi-system theories (such as the co-existence of separate model-free and model-based inference under the Hybrid-RL model). For example, active inference as implemented here could be used in conjunction with algorithms more similar to the model-free part of the Hybrid-RL model, which may be implemented as a form of habitual behaviour[19]. This could form the basis of interesting future extensions and may lead to a model better able to account for the distinct behaviour observed across datasets. Nevertheless, such an approach may only yield marginal improvements for the datasets investigated here. First, behaviour in the Online and Shock datasets does not reveal strong main effects of transition type and are thus unlikely to benefit from modelling directed exploration via expected free energy. Second, main effects of outcome are minor on the Magic Carpet and Spaceship tasks, which are thus unlikely to feature significant model-free behavioural components. Taken together, the simulation analyses thus far appear to confirm the model comparison analyses which indicated only the Magic Carpet and Spaceship datasets to be better described by active inference.

A critical component of active inference is its formulation of the exploration-exploitation trade-off. The studies by da Silva and Hare[6] included post-hoc reports of subjects, which included descriptions of intentionally visiting a specific final-stage state multiple times, before actively aiming for the other state. These reports are congruent with the observed transition effects and the authors proposed this may indicate directed exploration behaviour. The main effects of transition type observed on the Magic Carpet and Spaceship datasets are considerably weaker in the active inference simulations. For the active inference model to provide a behavioural description that is incompatible with a pure reward driven strategy, sufficient weight needs to be assigned to the information gathering term. This was found for approximately half of all subjects, suggesting that sensitivity to information gain was common. Although necessary, it is not a sufficient condition to produce significant effects of transition-type. For example, small learning rates, as well as small decay rates, diminish the effect of new observations on beliefs and thereby also decrease the information gain term of the expected free energy. Thus, noticeable transition-type effects also require observations to be able to significantly impact beliefs under current model assumptions. By performing the simulation analyses for subjects assigned a low or high λ-parameter separately, we show that this parameter does contribute to the main-effects of transition when using participant parameters, although the stratification of the datasets leads to small sample sizes. The correlation analyses support this, as participants with greater transition effects tend to be assigned lower λ values. In addition, these lower parameter values appear associated with relatively better model fits for active inference. However, the correlations are small-to-moderate in strength and even the simulations for the low λ-parameter group still do not fully capture the transition effects seen in the behavioural data. It is important to note that the two-step task was not specifically designed to disambiguate directed exploration strategies. Participants in the low λ group did not obtain more reward, probably due to the little influence agents have on the obtained rewards as shown by Kool et al.[36]. Ultimately, as per the exploration-exploitation trade-off, exploration is only worthwhile if sacrificed

short-term reward can be recouped by utilizing obtained information via different action selection strategies. In sum, active inference as implemented here likely does not provide a full description of observed exploration behaviour, although surrogate tasks should be considered.

Alternative specifications for an information-bonus have been considered by previous research. Compared to computing the expected Kullback-Leibler divergence, simpler accounts include an 'all-or-nothing' bonus, in which the least-explored option receives a fixed information bonus[59]. Others have modelled directed exploration by scaling the information-bonus linearly by using a count of the number of times an option has been selected[15]. More sophisticated strategies are time-horizon dependent[60], a further possible extension to the current active inference implementation already discussed in the literature[61]. Another possibility altogether is the existence of distinct exploration and exploitation states, rather than a weighted trade-off that is computed on a single-trial basis[11].

An important issue for studying exploration is the coupling of reward and information. In many paradigms, including the two-step task, participants receive only outcome information about the selected action. To the extent that participants act to obtain rewards, more rewarding actions naturally tend to be sampled more often, thus correlating how much information participants have about an option and how rewarding it is[59]. Researchers have attempted to manipulate the value of either strategy using ideas such as forced sampling[15] and periodic introductions of novel options[62]. Additionally, Horvath et al.[63] modeled a bandit-task using an approach similar to active inference, but withheld reward-information on a subset of trials. This allowed for a demonstration of human sensitivity to an information-bonus computed as the expected Kullback-Leibler divergence. As the two-step task was designed for the disambiguation of model-free and model-based inference, the resulting lack of decoupling between reward and information may have interfered with identifying a directed exploration strategy due to its correlation with a reward-maximization strategy. The level of noise in the parameter recovery supports this interpretation.

Some considerations of the current work deserve mention. First, due to interactions between learning and action selection processes, it is possible active inference would explain exploration behaviour better when combined with a different learning algorithm than ours. This may be the case directly with respect to the employed learning rule, for example by using a hierarchical model that estimates environmental volatility. It could, however, also be a function of the (misconstrued) task model, on which directed exploration relies to predict information gain. Secondly, the current study implements active inference by specifying expected free energy as the loss function for action selection. Over time, the scope of active inference in the literature has been extended to also include learning and inference schemes, often based on message-passing implementations[64]. Although we focused on final-stage outcomes to stay close to the two-step task literature, the exploration of more expansive applications of the framework may be considered in the future, including potential information-gain incentives for inference about the transition structure. Future research might also extend its scope beyond directed exploration incentives to explore dynamic habitual control via learned, rather than static, habits, which might allow for an active inference-based analogue to model-free inference. Moreover, the Hybrid-RL model learns point estimates and active inference uses probability distributions, preventing direct comparisons of action selection strategies on this task. This complicates the interpretation of relative model fits as these may result from differences in either learning or action selection. Nevertheless, we aimed to mediate this issue by using additional logistic regression analyses instead of relying only on model comparisons. Overall, the aim to investigate the applicability of active inference in describing data not specifically recorded for this purpose using diverse datasets did complicate drawing certain conclusions. As such, the heterogeneity of the datasets is difficult to fully account in part due to considerable differences in sample sizes. The analyses presented here should thus be regarded as a proof-of-principle to steer future work testing the empirical validation of active inference.

To conclude, we replicated results by da Silva and Hare[6] and extended on behavioural discrepancies between datasets of the two-step task. Participants in the Magic Carpet and Spaceship datasets not only appeared to perform more model-based inference, but also showed more volatile learning dynamics and greater transition effects, which indicate the use of directed exploration strategies. These two datasets were better described by an active inference model than a Hybrid-RL model, while the models scored similarly for the Online and Shock datasets. For the Magic Carpet and Spaceship datasets, the use of a learning model based on probability distributions appeared to contribute to the better model fits and captured behaviour better than the Hybrid-RL model. For such a model, a decay mechanism for beliefs about unsampled options was found to be important. Model parameters indicated that approximately half of all subjects were sensitive to information gain of actions. However, active inference was only partly able to capture the observed transition effects, and thus likely did not provide a full account of exploration behaviour.

Data availability

The datasets analysed during the current study were previously made available by Kool et al.[36] (github.com/wkool/tradeoffs), Lockwood et al.[37] (osf.io/3stp9/files), and da Silva and Hare[6] (github.com/carolfs/muddled_models).

Code availability

The modeling code used in this work is available at github.com/SamGijsen/AI2step.

Received: 6 May 2022; Accepted: 30 September 2022
Published online: 21 October 2022

References

1. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, 2018).

2. Decker, J. H., Otto, A. R., Daw, N. D. & Hartley, C. A. From creatures of habit to goal-directed learners: Tracking the developmental emergence of model-based reinforcement learning. *Psychol. Sci.* **27**, 848–858 (2016).
3. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
4. Voon, V. *et al.* Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. *Transl. Psychiatry* **5**, e670–e670 (2015).
5. Wyckmans, F. *et al.* Reduced model-based decision-making in gambling disorder. *Sci. Rep.* **9**, 1–10 (2019).
6. da Silva, C. F. & Hare, T. A. Humans primarily use model-based inference in the two-stage task. *Nat. Hum. Behav.* **4**, 1053–1066 (2020).
7. Akam, T., Costa, R. & Dayan, P. Simple plans or sophisticated habits? State, transition and learning interactions in the two-step task. *PLoS Comput. Biol.* **11**, e1004648 (2015).
8. Feher da Silva, C. & Hare, T. A. A note on the analysis of two-stage task results: How changes in task structure affect what model-free and model-based strategies predict about the effects of reward and transition on the stay probability. *PloS One* **13**, e0195328 (2018).
9. Cohen, J. D., McClure, S. M. & Yu, A. J. Should i stay or should i go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 933–942 (2007).
10. Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
11. Wilson, R. C., Bonawitz, E., Costa, V. D. & Ebitz, R. B. Balancing exploration and exploitation with information and randomization. *Curr. Opin. Behav. Sci.* **38**, 49–56 (2021).
12. Osband, I., Blundell, C., Pritzel, A. & Van Roy, B. Deep exploration via bootstrapped dqn. *Adv. Neural Inf. Process. Syst.* **29**, 4026–4034 (2016).
13. Meder, B. & Nelson, J. D. Information search with situation-specific reward functions. *Judgm. Decis. Mak.* **7**, 119–148 (2012).
14. Tsividis, P., Gershman, S., Tenenbaum, J. & Schulz, L. Information selection in noisy environments with large action spaces. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* vol. 36 (2014).
15. Dezza, C., Yu, A. J., Cleeremans, A. & Alexander, W. Learning the value of information and reward over time when solving exploration-exploitation problems. *Sci. Rep.* **7**, 1–13 (2017).
16. Itti, L. & Baldi, P. Bayesian surprise attracts human attention. *Vis. Res.* **49**, 1295–1306 (2009).
17. Yang, S.C.-H., Lengyel, M. & Wolpert, D. M. Active sensing in the categorization of visual patterns. *Elife* **5**, e12215 (2016).
18. Mirza, M. B., Adams, R. A., Mathys, C. & Friston, K. J. Human visual exploration reduces uncertainty about the sensed world. *PloS One* **13**, e0190429 (2018).
19. Friston, K. *et al.* Active inference and learning. *Neurosci. Biobehav. Rev.* **68**, 862–879 (2016).
20. Friston, K. The free-energy principle: A unified brain theory?. *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
21. Sajid, N., Ball, P. J., Parr, T. & Friston, K. J. Active inference: Demystified and compared. *Neural Comput.* **33**, 674–712 (2021).
22. Gershman, S. J. What does the free energy principle tell us about the brain? arXiv preprint arXiv:1901.07945 (2019).
23. Knill, D. C. & Pouget, A. The bayesian brain: The role of uncertainty in neural coding and computation. *TRENDS Neurosci.* **27**, 712–719 (2004).
24. Friston, K., Kilner, J. & Harrison, L. A free energy principle for the brain. *J. Physiol. Paris* **100**, 70–87 (2006).
25. Friston, K. A free energy principle for biological systems. *Entropy* **14**, 2100–2121 (2012).
26. FitzGerald, T. H., Schwartenbeck, P., Moutoussis, M., Dolan, R. J. & Friston, K. Active inference, evidence accumulation, and the urn task. *Neural Comput.* **27**, 306–328 (2015).
27. Marković, D., Stojić, H., Schwöbel, S. & Kiebel, S. J. An empirical evaluation of active inference in multi-armed bandits. *Neural Netw.* **144**, 229–246 (2021).
28. Friston, K. *et al.* Active inference and epistemic value. *Cogn. Neurosci.* **6**, 187–214 (2015).
29. Smith, R. *et al.* Imprecise action selection in substance use disorder: Evidence for active learning impairments when solving the explore-exploit dilemma. *Drug Alcohol Depend.* **215**, 108208 (2020).
30. Smith, R. *et al.* Slower learning rates from negative outcomes in substance use disorder over a 1-year period and their potential predictive utility. *Comput. Psychiatry* **6**, 117–141 (2022).
31. Smith, R. *et al.* A bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *PLoS Comput. Biol.* **16**, e1008484 (2020).
32. Smith, R. *et al.* Greater decision uncertainty characterizes a transdiagnostic patient sample during approach-avoidance conflict: A computational modelling approach. *J. Psychiatry Neurosci.* **46**, E74–E87 (2021).
33. Smith, R. *et al.* Long-term stability of computational parameters during approach-avoidance conflict in a transdiagnostic psychiatric patient sample. *Sci. Rep.* **11**, 1–13 (2021).
34. Schwartenbeck, P. *et al.* Evidence for surprise minimization over value maximization in choice behavior. *Sci. Rep.* **5**, 1–14 (2015).
35. Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R. & Friston, K. The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cereb. Cortex* **25**, 3434–3445 (2015).
36. Kool, W., Cushman, F. A. & Gershman, S. J. When does model-based control pay off?. *PLoS Comput. Biol.* **12**, e1005090 (2016).
37. Lockwood, P. L., Klein-Flügge, M. C., Abdurahman, A. & Crockett, M. J. Model-free decision making is prioritized when learning to avoid harming others. *Proc. Natl. Acad. Sci.* **117**, 27719–27730 (2020).
38. Miller, K. J., Brody, C. D. & Botvinick, M. M. Identifying model-based and model-free patterns in behavior on multi-step tasks. bioRxiv preprint at https://www.biorxiv.org/content/10.1101/096339v2 (2016).
39. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
40. Rummery, G. A. & Niranjan, M. *On-line Q-learning Using Connectionist Systems* Vol. 37 (Citeseer, Princeton, 1994).
41. Guo, D. & Yu, A. J. Why so gloomy? A bayesian explanation of human pessimism bias in the multi-armed bandit task. *Adv. Neural Inf. Process. Syst.* **31**, 5176–5185 (2018).
42. Liakoni, V., Modirshanechi, A., Gerstner, W. & Brea, J. Learning in volatile environments with the bayes factor surprise. *Neural Comput.* **33**, 269–340 (2021).
43. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Techn. J.* **27**, 379–423 (1948).
44. Kopp, B. *et al.* P300 amplitude variations, prior probabilities, and likelihoods: A bayesian erp study. *Cogn. Affect. Behav. Neurosci.* **16**, 911–928 (2016).
45. O'Reilly, J. X. *et al.* Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc. Natl. Acad. Sci.* **110**, E3660–E3669 (2013).
46. Gijsen, S., Grundei, M., Lange, R. T., Ostwald, D. & Blankenburg, F. Neural surprise in somatosensory bayesian learning. *PLoS Comput. Biol.* **17**, e1008068 (2021).
47. Virtanen, P. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272. https://doi.org/10.1038/s41592-019-0686-2 (2020).
48. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies-revisited. *Neuroimage* **84**, 971–985 (2014).
49. Penny, W. D. *et al.* Comparing families of dynamic causal models. *PLoS Comput. Biol.* **6**, e1000709 (2010).

50. Green, C., Benson, C., Kersten, D. & Schrater, P. Alterations in choice behavior by manipulations of world model. *Proc. Natl. Acad. Sci.* **107**, 16401–16406 (2010).
51. Castro-Rodrigues, P. *et al.* Explicit knowledge of task structure is a primary determinant of human model-based action. *Nat. Hum. Behav.* **6**, 1126–1141 (2022).
52. Meyniel, F. Brain dynamics for confidence-weighted learning. *PLoS Comput. Biol.* **16**, e1007935 (2020).
53. Ostwald, D. *et al.* Evidence for neural encoding of bayesian surprise in human somatosensation. *NeuroImage* **62**, 177–188 (2012).
54. Meyniel, F., Maheu, M. & Dehaene, S. Human inferences about sequences: A minimal transition probability model. *PLoS Comput. Biol.* **12**, e1005260 (2016).
55. Toyama, A., Katahira, K. & Ohira, H. A simple computational algorithm of model-based choice preference. *Cogn. Affect. Behav. Neurosci.* **17**, 764–783 (2017).
56. Toyama, A., Katahira, K. & Ohira, H. Biases in estimating the balance between model-free and model-based learning systems due to model misspecification. *J. Math. Psychol.* **91**, 88–102 (2019).
57. Mirza, M. B., Adams, R. A., Friston, K. & Parr, T. Introducing a Bayesian model of selective attention based on active inference. *Sci. Rep.* **9**, 1–22 (2019).
58. Parr, T. & Friston, K. J. The anatomy of inference: Generative models and brain structure. *Front. Comput. Neurosci.* **12**, 90 (2018).
59. Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans use directed and random exploration to solve the explore-exploit dilemma. *J. Exp. Psychol. Gen.* **143**, 2074 (2014).
60. Zhang, S. & Yu, A. J. Forgetful bayes and myopic planning: Human learning and decision-making in a bandit setting. *Adv. Neural Inf. Process. Syst.* **26**, 2607–2615 (2013).
61. Smith, R., Friston, K. J. & Whyte, C. J. A step-by-step tutorial on active inference and its application to empirical data. *J. Math. Psychol.* **107**, 102632 (2022).
62. Costa, V. D. & Averbeck, B. B. Primate orbitofrontal cortex codes information relevant for managing explore-exploit tradeoffs. *J. Neurosci.* **40**, 2553–2561 (2020).
63. Horvath, L. *et al.* Human belief state-based exploration and exploitation in an information-selective symmetric reversal bandit task. *Comput. Brain Behav.* **4**, 442–462 (2021).
64. Parr, T., Markovic, D., Kiebel, S. J. & Friston, K. J. Neuronal message passing using mean-field, bethe, and marginal approximations. *Sci. Rep.* **9**, 1–18 (2019).

Acknowledgements

We acknowledge and thank Kool et al., Lockwood et al., and da Silva and Hare for making their data publicly available.

Author contributions

S.G. and F.B. were responsible for conceptualization and the methodology. S.G. performed the formal analyses and wrote a first version of the manuscript. All authors revised and edited the work.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.G.

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:	Gijsen, Sam Christian Jan
Institute:	Department of Education and Psychology, Freie Universität Berlin
Doctoral study subject:	Psychology
Title:	The brain as a generative model: information-theoretic surprise in learning and action

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. (Submitted) EEG mismatch responses in a multi-modal roving stimulus paradigm provide evidence for probabilistic inference across audition, somatosensation and vision. Human Brain Mapping
3. Gijsen, S., Grundei, M. & Blankenburg, F. Active inference and the two-step task. Scientific Reports 12, 17682 (2022).

* shared authorship

III. Explanation of own share of these works:

Regarding II. 1.: Study conceptualisation and design (substantial), methodology (substantial), programming of task (in part), data collection (substantial), data analysis (substantial), discussion of results (substantial), writing/revising the manuscript (substantial).

Regarding II. 2.: Methodology (in part), discussion of results (in part), writing/revising the manuscript (in part).

Regarding II. 3.: Study conceptualisation and design (vast majority), methodology (vast majority), data analysis (all), discussion of results (substantial), writing/revising the manuscript (vast majority).

IV. Names, addresses, and e-mail addresses for the relevant co-authors:

Regarding II. 1.: Miro Grundei (1,2), m.grundei@fu-berlin.de
Robert T. Lange (3), robert.t.lange@tu-berlin.de
Dirk Ostwald (4), dirk.ostwald@ovgu.de
Felix Blankenburg (1,2), felix.blankenburg@fu-berlin.de

Regarding II. 2.: Miro Grundei (1,2), see above
Pia Schröder (1), pia.schroeder@fu-berlin.de
Felix Blankenburg (1,2), see above

Regarding II. 3.: Miro Grundei (1,2), see above
Felix Blankenburg (1,2), 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, Einstein Center for Neurosciences, Berlin, Germany
(4) Institute of Psychology, Otto von Guericke Universität Magdeburg, 39106 Magdeburg, Germany

Date, doctoral candidate signature: ……

I confirm the declaration made by Sam Gijsen under III.:

Name: Miro Grundei Signature: ……

Name: Robert T. Lange Signature: ……

Name: Dirk Ostwald Signature: ……

Name: Felix Blankenburg Signature: ……

Name: Pia Schröder 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, 20.12.2022

Sam Gijsen