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. / Ph. D.)

am Fachbereich Erziehungswissenschaften 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

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:

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

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

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 action 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.

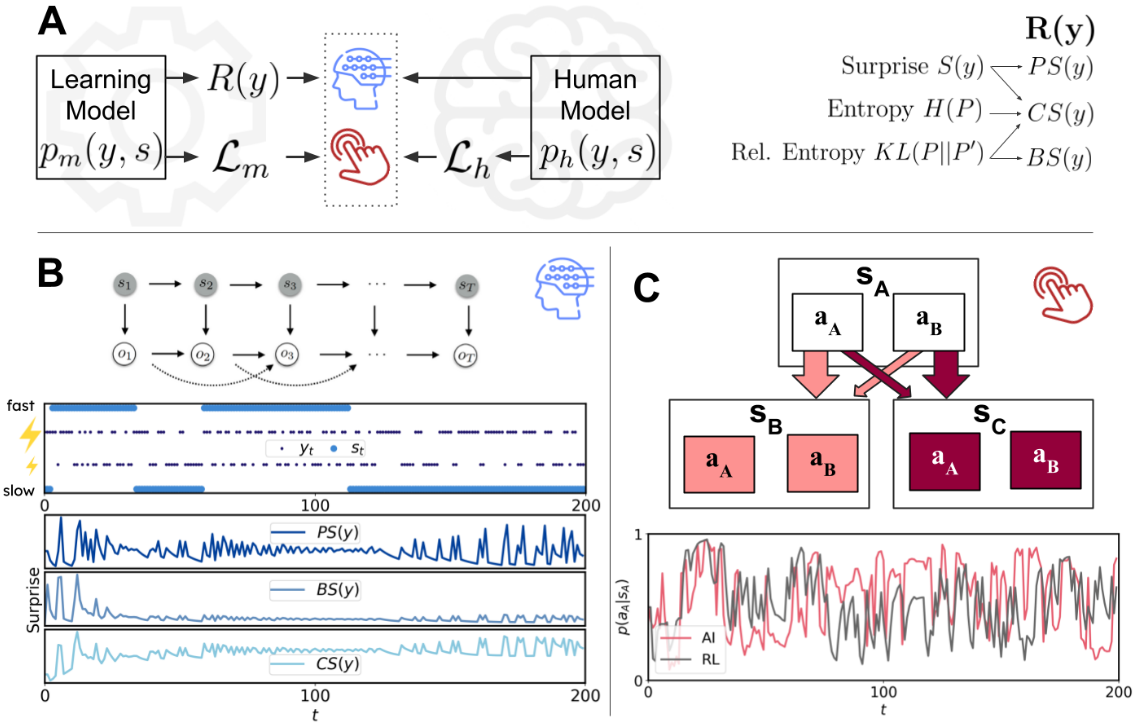

Figure 1: **An investigation of perceptual learning and action using information theoretic tools.** A) The left-hand side includes a schematic overview of the experimental approach using EEG signals and action selection. The human participant is thought to maintain a task model $p_h(y, s)$ which is updated 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 particularly 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 combine 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 (∼125ms) and later (∼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 precision 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).

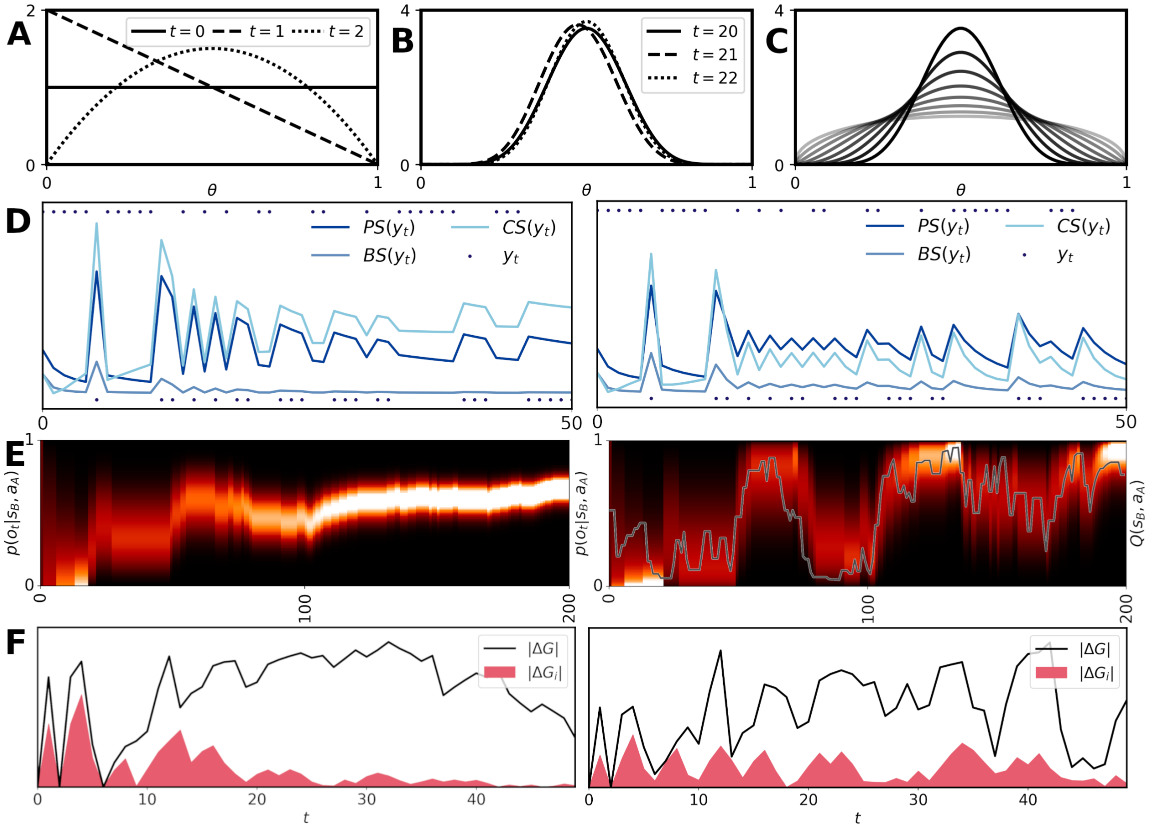

Figure 2: **Model dynamics.** A) The evolution of a uniform Beta-prior as a binary observation $y \sim \{0, 1\}$ is made for $y_{t=0} = 0$ and $y_{t=1} = 1$ sequentially. B) The same sequential observations are made given an informed Beta distribution concentrated at $\theta = 0.5$. C) A forgetting kinetic 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, corresponding 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 breakthroughs 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 surprise, 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

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

Original work: Study 3

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

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: ………………………………………………… 20/12/2022

I confirm the declaration made by Sam Gijsen under III.:

Name: Miro Grundei Signature: M. Grundei…………………………………………

Name: Robert T. Lange Signature: R. Lange…………………………………………

Name: Dirk Ostwald Signature: ……………………………………………………………

Name: Felix Blankenburg Signature: ……………………………………………………………

Name: Pia Schröder Signature: ……………Pia Schröder………………………………………

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