Surprise: a unified theory and experimental predictions

Alireza Modirshanechi, Johanni Brea, Wulfram Gerstner

EPFL, School of Computer and Communication Sciences and School of Life Sciences, Lausanne, Switzerland

November 1, 2021

Abstract

Surprising events trigger measurable brain activity and influence human behavior by affecting learning, memory, and decision-making. Currently there is, however, no consensus on the definition of surprise. Here we identify 16 mathematical definitions of surprise in a unifying framework, show how these definitions relate to each other, and prove under what conditions they are indistinguishable. We classify these surprise measures into four main categories: (i) change-point detection surprise, (ii) information gain surprise, (iii) prediction surprise, and (iv) confidence-correction surprise. We design experimental paradigms where different categories make different predictions: we show that surprise-modulation of the speed of learning leads to sensible adaptive behavior only for change-point detection surprise whereas surprise-seeking leads to sensible exploration strategies only for information gain surprise. However, since neither change-point detection surprise nor information gain surprise perfectly reflect the definition of ‘surprise’ in natural language, a combination of prediction surprise and confidence-correction surprise is needed to capture intuitive aspects of surprise perception. We formalize this combination in a new definition of surprise with testable experimental predictions. We conclude that there cannot be a single surprise measure with all functions and properties previously attributed to surprise. Consequently, we postulate that multiple neural mechanisms exist to detect and signal different aspects of surprise.

Author note

AM is grateful to Vasiliki Liakoni, Martin Barry, and Valentin Schmutz for many useful discussions in the course of the last few years, and to Andrew Barto for insightful discussions through and after EPFL Neuro Symposium 2021 on “Surprise, Curiosity and Reward: from Neuroscience to AI”. We thank K. Robbins and collaborators for their publicly available experimental data (Robbins et al., 2018). All code needed to reproduce the results reported here will be made publicly available after publication acceptance. This research was supported by Swiss National Science Foundation (no. 200020_184615). Correspondence concerning this article should be addressed to Alireza Modirshanechi, School of Computer and Communication Sciences and School of Life Sciences, EPFL, Lausanne, Switzerland. E-mail: alireza.modirshanechi@epfl.ch.
Introduction

Imagine you open the curtains one morning and find the street in front of your apartment covered by fresh snow. If you have expected a warm and sunny morning according to the weather forecast, you feel ‘surprised’ as you see the white streets; as a consequence of surprise, the activity of many neurons in your brain changes (Antony et al., 2021; Kolossa et al., 2015; Mars et al., 2008; Modirshanechi et al., 2019; Ostwald et al., 2012; Squires et al., 1976) and your pupils dilate (Antony et al., 2021; Nassar et al., 2012; Preuschoff et al., 2011). Surprise affects how we predict and perceive our future and how we remember our past (Antony et al., 2021; Dubey & Griffiths, 2020; Gershman et al., 2017; Gottlieb & Oudeyer, 2018; Oudeyer & Kaplan, 2009; Sinclair & Barense, 2018; Soltani & Izquierdo, 2019; Yu & Dayan, 2005). For example, some studies suggest that you would rely less on the weather forecast for your future plans after the snowy morning (Behrens et al., 2007; Findling et al., 2021; Gerstner et al., 2018; Nassar et al., 2010; Soltani & Izquierdo, 2019; Xu et al., 2021; Yu & Dayan, 2005). Other studies predict that you would remember more vividly the face of the random stranger who walked past the street in that very moment you felt surprised (Rouhani & Niv, 2021; Rouhani et al., 2018), and some predict that this moment of surprise might have even modified your memory of another snowy morning in the past (Exton-McGuinness et al., 2015; Gershman et al., 2017; Gershman et al., 2014; Sinclair & Barense, 2018). To understand and explain the computational role (Marr, 1982) of surprise in different brain functions, one first needs to ask ‘what does it really mean to be surprised?’ and formalize how surprise is perceived by our brain. For instance, when you see the white street, do you feel ‘surprised’ because what you expected turned out to be wrong (Faraji et al., 2018; Gläscher et al., 2010; Shannon, 1948; Tribus, 1961) or because you need to change your trust in the weather forecast (Baldi, 2002; Liakoni et al., 2021; Schmidhuber, 2010)?

Computational models of perception, learning, memory, and decision-making often assume that humans implicitly perceive their sensory observations as probabilistic outcomes of a generative model with hidden variables (Dubey & Griffiths, 2019; Findling et al., 2021; Fiser et al., 2010; Friston, 2010; Gershman et al., 2017; Liakoni et al., 2021; Schmidhuber, 2010; Soltani & Izquierdo, 2019; Yu & Dayan, 2005). In the example above, the observation is whether it snows or not and the hidden variables characterise how the probability of snowing depends on old observations and relevant context information (such as the current season, yesterday’s weather, and the weather forecast). Different brain functions are then modeled as aspects of statistical inference (Barber, 2012) and probabilistic control (Sutton & Barto, 2018) in such generative models (Behrens et al., 2007; Daw et al., 2011; Dubey & Griffiths, 2019; Findling et al., 2021; Friston, 2010; Friston et al., 2017; Gershman et al., 2017; Gläscher et al., 2010; Horvath et al., 2021; Liakoni et al., 2021; Meyniel et al., 2016; Nassar et al., 2012; Rouhani et al., 2020; Ryali et al., 2018; Schmidhuber, 2010; Soltani & Izquierdo, 2019; Yu & Dayan, 2005). In these probabilistic settings, surprise of an observation depends on the relation between the observation and our expectation of what to observe.

In the past decades, different definitions and formal measures of surprise have been proposed and studied (Baldi, 2002; Barto et al., 2013; Faraji et al., 2018; Friston, 2010; Gläscher et al., 2010; Itti & Baldi, 2006, 2009; Kolossa et al., 2015; Liakoni et al., 2021; Palm, 2012; Schmidhuber, 2010; Shannon, 1948; Tribus, 1961). These surprise measures have been successful both in explaining the role of surprise in different brain functions (Findling et al., 2021; Friston, 2010; Gershman et al., 2017; Liakoni et al., 2021; Schmidhuber, 2010; Soltani & Izquierdo, 2019; Yu & Dayan, 2005) and in identifying signatures of surprise in behavioral and physiological measurements (Gijsen et al., 2021; Gläscher et al., 2010; Kolossa et al., 2015; Lieder et al., 2013; Maheu et al., 2019; Mars et al., 2008; Meyniel, 2020; Meyniel et al., 2016; Modirshanechi et al., 2019; Mousavi et al., 2020; Ostwald et al., 2012; Preuschoff et al., 2011; Rubin et al., 2016). However, there are still many open questions (Baldi, 2002; Barto et al., 2013; Liakoni et al., 2021; Palm, 2012; Schmidhuber, 2010) including, but not limited to: (i) Since the notion of surprise has been linked to a multitude of computational roles and physiological phenomena, should we still assume that the word ‘surprise’ (and quantitative measures thereof) refers to a single common phenomenon or rather to a
multitude of different phenomena? (ii) If ‘surprise’ refers to a multitude of different phenomena, can we assign different measures of surprise to different computational roles and different phenomena? In other words, are different surprise measures qualitatively distinguishable in experiments? (iii) Can we identify mathematical relations between different surprise measures? In particular, is one measure a special case of another one, completely distinct, or do they have some common ground? (iv) Does the word surprise as used in common language match one of the known mathematical definitions?

In this work, we analyze and discuss several previously proposed surprise measures (Baldi, 2002; Burda et al., 2019; Faraji et al., 2018; Friston, 2010; Gläscher et al., 2010; Itti & Baldi, 2006, 2009; Kolossa et al., 2015; Liakoni et al., 2021; Pathak et al., 2017; Schmidhuber, 2010; Shannon, 1948) in a unifying framework. First, we give definitions for each of these measures, show their similarities and differences, and prove and discuss their mathematical properties. We then propose experimental paradigms where different surprise measures entail different predictions for perception, learning, and decision-making. Our theoretical analysis provides a comprehensive understanding of current paradigms (Kuhn, 1962) for the study of surprise, shows that a single notion of surprise cannot explain the multitude of computational roles and physiological phenomena to which surprise has been linked, and shows how different measures of surprise relate to how we use the word ‘surprise’ in common language. Furthermore, we suggest a new definition of surprise which can capture some intuitive aspects of surprise perception that cannot be explained by current definitions. Our analyses enable experimental scientists to dissociate the contributions of different surprise measures to behavioral and physiological measurements.

Subjective world-model: A unifying generative model

Our primary goal is to study the theoretical properties of different formal measures of surprise and their predictions for perception, learning, and decision-making in a common mathematical framework. To do so, we consider a generative model of a volatile environment (Fig. 1A) that captures a few key features of daily life and unifies many existing model environments in neuroscience and psychology (Behrens et al., 2007; Daw et al., 2011; Findling et al., 2021; Gijsen et al., 2021; Gläscher et al., 2010; Glaze et al., 2015; Heilbron & Meyniel, 2019; Horvath et al., 2021; Huys et al., 2015; Liakoni et al., 2021; Mars et al., 2008; Meyniel et al., 2016; Nassar et al., 2012; Nassar et al., 2010; Ostwald et al., 2012; Wilson et al., 2013; Xu et al., 2021). The generative model describes the subjective interpretation of the environment from the point of view of an agent (e.g., a human participant or an animal). Importantly, we assume that the agent takes the possibility into account that the environment may undergo abrupt changes at unknown points in time (similar to Glaze et al., 2015; Heilbron and Meyniel, 2019; Liakoni et al., 2021; Nassar et al., 2010; Xu et al., 2021). Note, however, that we do not assume that the environment has the same dynamics as those assumed by the agent.

General definition

At each time $t$, the agent’s model of the environment is characterized by a tuple of 4 random variables $(X_t, Y_t, \Theta_t, C_t)$ (Fig. 1A). $X_t$ and $Y_t$ are observable, whereas $\Theta_t$ and $C_t$ are unobservable (hidden). We refer to $X_t$ as the cue and to $Y_t$ as the observation at time $t$. Examples of an observation are an image on a computer screen (da Silva & Hare, 2020; Daw et al., 2011; Gläscher et al., 2010; Kolossa et al., 2015; Mars et al., 2008; Xu et al., 2021), an auditory tone (Imada et al., 1993; Lieder et al., 2013; Meyniel et al., 2016; Walz et al., 2015), and an electrical or mechanical stimulation (Esmaeili et al., 2021; Gijsen et al., 2021; Ostwald et al., 2012). The cue variable $X_t$ can be interpreted as a predictor of the next observation, since it describes all observable information that is relevant for predicting the observation $Y_t$. Examples of a cue variable are the previous observation $Y_{t-1}$ (Maheu et al., 2019; Meyniel et al., 2016; Modirshanechi et al., 2019), the last action of a participant (which we will denote by $A_{t-1}$) (Behrens et al., 2007; Findling et al., 2021; Horvath et al., 2021), and a conditioned stimulus in Pavlovian conditioning tasks (Gershman...
At time \( t \), given the cue variable \( X_t = x \), the agent assumes that the observation \( Y_t = y \) comes from a time-invariant conditional distribution \( P_{Y|X}(y|x; \theta) \) parameterized by the hidden variable \( \Theta_t = \theta \). We do not put any constraints on the sets to which \( x, y, \) and \( \theta \) belong; in a non-parametric setup, \( \theta \) can even have infinite dimension (Gershman & Blei, 2012; Ghahramani, 2013). We refer to \( \Theta_t \) as the environment parameter at time \( t \). The sequence of variables \( \Theta_{1:t} = [\Theta_1, ..., \Theta_t] \) describe the temporal dynamics of the observations \( Y_{1:t} \) given the cue variables \( X_{1:t} \) in the agent’s model of the environment. Similar to well-known models of volatile environments (Adams & MacKay, 2007; Behrens et al., 2007; Fearnhead & Liu, 2007; Findling et al., 2021; Glaze et al., 2015; Heilbron & Meyniel, 2019; Liakoni et al., 2021; Meyniel et al., 2016; Nassar et al., 2012; Nassar et al., 2010; Wilson et al., 2013; Xu et al., 2021; Yu & Dayan, 2005), the agent assumes that the environment undergoes abrupt changes at random points in time. An abrupt change at time \( t \) is specified by the event \( C_t = 1 \) and happens with a probability \( p_c \in [0, 1] \); otherwise \( C_t = 0 \). If the environment abruptly changes at time \( t \) (i.e., \( C_t = 1 \)), then the agent assumes that the environment parameter \( \Theta_t \) is sampled from a prior distribution \( \pi(0) \) independently of \( \Theta_{t-1} \); if there is no change (\( C_t = 0 \)), then \( \Theta_t \) remains the same as \( \Theta_{t-1} \). We refer to \( p_c \) as the change-point probability.

We use \( \mathbb{P} \) to refer to probability mass functions (pmf) for discrete random variables and to probability density functions (pdf) for continuous random variables. In general, we show random variables by capital letters and their values by small letters. However, for any pair of arbitrary random variables \( W \) and \( V \) and their values \( w \) and \( v \), whenever there is no risk of ambiguity, we either drop the capital- or the small-letter notation and, for example, write \( \mathbb{P}(W = w|V = v) \) as \( \mathbb{P}(w|v) \). When there is a risk of ambiguity, we keep the capital notation for the random variables, e.g., we write \( \mathbb{P}(W = v, V = v) \) as \( \mathbb{P}(W = v, v) \). Given this convention, the agent’s model of the environment described above is formalized in Definition 1.

**Definition 1.** (Subjective world-model) An agent’s model of the environment is defined for \( t > 0 \) as a joint probability distribution over \( Y_{1:t}, X_{1:t}, \Theta_{1:t}, \) and \( C_{1:t} \) as

\[
\mathbb{P}(y_{1:t}, x_{1:t}, \theta_{1:t}, c_{1:t}) = \mathbb{P}(c_1) \mathbb{P}(\theta_1) \mathbb{P}(x_1) \mathbb{P}(y_1|x_1, \theta_1) \prod_{\tau=2}^{t} \mathbb{P}(c_\tau) \mathbb{P}(\theta_\tau|\theta_{\tau-1}, c_\tau) \mathbb{P}(x_\tau|x_{\tau-1}, y_{\tau-1}) \mathbb{P}(y_\tau|\theta_\tau, x_\tau),
\]

where \( c_1 \) is by definition equal to 1 (i.e., \( \mathbb{P}(c_1) = \delta_{(c_1,1)} \)), \( \mathbb{P}(\theta_1) = \pi(0)(\theta_1) \) for an arbitrary distribution \( \pi(0) \), and

\[
\mathbb{P}(c_\tau) = \text{Bernoulli}(c_\tau; p_c) \\
\mathbb{P}(\theta_\tau|\theta_{\tau-1}, c_\tau) = \pi(0)(\theta_\tau) \delta_{(c_\tau,1)} + \delta_{(\theta_\tau,\theta_{\tau-1})} \delta_{(c_\tau,0)} \\
\mathbb{P}(y_\tau|x_\tau, \theta_\tau) = P_{Y|X}(y_\tau|x_\tau; \theta_\tau),
\]

where \( \delta \) is the delta distribution. We do not make any assumption about \( \mathbb{P}(x_1) \) and \( \mathbb{P}(x_\tau|x_{\tau-1}, y_{\tau-1}) \).

See Table 1 for a summary of the notation. Fig. 1B-E shows how the generative model of Definition 1 relates to other existing model environments in neuroscience and psychology – see Appendix A: Special cases and links to related works for details.

**Additional notation, belief, and marginal probability**

We define \( \mathbb{P}^{(t)} \) as \( \mathbb{P} \) conditioned on the sequences of observations \( y_{1:t} \) and cue variables \( x_{1:t} \). For example, for an arbitrary random variable \( W \) with value \( w \), we write \( \mathbb{P}(w|y_{1:t}, x_{1:t}) = \mathbb{P}^{(t)}(w) \). We occasionally use an alternative notation and write \( \mathbb{P}^{(t)}(W = w) \) as \( \mathbb{P}^{(t)}_W(w) \) where the random variable is specified in the subscript.
Figure 1: **Subjective model of the environment.** A. The Bayesian network (Barber, 2012) corresponding to the most general case of our generative model in Eq. 1 and Eq. 2. The arrows show conditional dependence, the grey nodes show the hidden variables ($C_{1:t+1}$ and $\Theta_{1:t+1}$), the blue nodes show the observations ($Y_{1:t+1}$), and the green nodes show the cue variables ($X_{1:t+1}$). A variety of tasks can be written in the form of a reduced version of our generative model. Specifically:

B. Standard generative model for studying passive learning in volatile environments (Adams & MacKay, 2007; Fearnhead & Liu, 2007; Liakoni et al., 2021; Nassar et al., 2012; Nassar et al., 2010; Wilson et al., 2013),

C. Generative model corresponding to variants of bandit and reversal bandit tasks (Behrens et al., 2007; Findling et al., 2021; Horvath et al., 2021), where the cue variable $X_t = A_t$ is a participant’s action,

D. Generative model for modeling human inferences about binary sequences (Gijsen et al., 2021; Maheu et al., 2019; Meyniel et al., 2016; Modirshanechi et al., 2019; Mousavi et al., 2020), and

E. classic Markov Decision Processes (MDPs) (Daw et al., 2011; Gläscher et al., 2010; Huys et al., 2015; Lehmann et al., 2019; Schultz et al., 1997; Sutton & Barto, 2018), where the cue variable $X_t = (A_{t-1}, Y_{t-1})$ consists of previous action and observation. See Appendix A: Special cases and links to related works for details.

Following this notation, we define an agent’s belief about the parameter $\Theta_t$ at time $t$ as the posterior distribution

$$\pi^{(t)}(\theta) = \mathbb{P}^{(t)}(\Theta_t = \theta).$$

We can equivalently write $\pi^{(t)} = \mathbb{P}_{\Theta_t}^{(t)}$. The belief plays a crucial role in the perception of surprise (see next sections), and we assume that an agent constantly updates its belief as it makes new observations; see Barber, 2012 and Liakoni et al., 2021 for examples of inference algorithms in generative models similar to ours.

Another important quantity is the marginal probability of observing $y$ given the cue $x$ and a belief $\pi^{(t)}$:

$$P(y|x; \pi^{(t)}) = \mathbb{E}_{\pi^{(t)}} \left[ P_{Y|X}(y|x; \Theta) \right] = \int P_{Y|X}(y|x; \theta) \pi^{(t)}(\theta) d\theta,$$

where the integration is replaced by summation whenever $\theta$ is discrete.

**Theories of surprise: A technical review**

Conditioned on the previous observations $y_{1:t}$ and cue variables $x_{1:t+1}$, how surprising is the next observation $y_{t+1}$? In this section, we address this question by examining previously proposed measures of surprise (Baldi, 2002; Burda et al., 2019; Faraji et al., 2018; Friston, 2010; Gläscher et al., 2010; Itti & Baldi, 2006, 2009; Kolossa et al., 2015; Liakoni et al., 2021; Pathak et al., 2017; Schmidhuber, 2010; Shannon, 1948). We define all surprise measures in the same mathematical framework and discuss their differences and similarities.
| Notation | Meaning |
|----------|---------|
| $X_t$ | Cue at time $t$ |
| $Y_t$ | Observation at time $t$ |
| $\Theta_t$ | Environment parameter at time $t$ |
| $C_t$ | Change-point indicator at time $t$ |
| $p_c$ | Change-point probability, i.e., the probability of $C_t = 1$ |
| $P_{Y|X}(y|x; \theta)$ | Time invariant distribution of observation $y$ given cue $x$, parameterized by $\theta$ |
| $\mathbb{P}$ | The distribution corresponding to the subjective model of the environment; see Definition 1 |
| $\mathbb{P}^{(t)}$ | $\mathbb{P}$ conditioned on observations and cues until time $t$, i.e., $x_{1:t}$ and $y_{1:t}$ |
| $\mathbb{P}_{W}^{(t)}$ | An alternative notation for the distribution of random variable $W$ given $x_{1:t}$ and $y_{1:t}$ |
| $\pi^{(0)}$ | Prior distribution over the environment parameter |
| | Equivalently, the distribution of $\Theta_t$ given $C_t = 1$ |
| $\pi^{(t)}$ | The belief about parameter $\Theta_t$ at time $t$, i.e., $\mathbb{P}^{(t)}(\Theta_t = \theta)$ |
| $P(y|x; \pi^{(t)})$ | The marginal probability of observation $y$ given cue $x$ and belief $\pi^{(t)}$; see Eq. 4 |

**Table 1:** Summary of notation introduced in Subjective world-model: A unifying generative model

Surprise measures are commonly used in experiments to study whether a behavioral or physiological variable $Z$ (e.g., the amplitude of the EEG P300 component (Kolossa et al., 2015; Meyniel et al., 2016)) is sensitive to or representative of surprise. Given two measures of surprise $S$ and $S'$, a typical experimental question is which one of them (if any) more accurately explains the variations of the variable $Z$ (Gijssen et al., 2021; Kolossa et al., 2015; Mousavi et al., 2020; Ostwald et al., 2012; Visalli et al., 2021); see Fig. 2A1. However, if there exists a one-to-one (monotonic) mapping between $S$ and $S'$ (e.g., as in Fig. 2A2), then the two surprise measures have the same explanatory power with respect to $Z$ – because any function of $S$ can be written in terms of $S'$ and vice-versa\(^1\). We, therefore, say $S$ and $S'$ are indistinguishable if there exists a strictly increasing function $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $S = f(S')$ for all choices of belief $\pi^{(t)}$, cue $x_t$, and observation $y_t$ (c.f. Fig. 2A2). Loosely speaking, we consider two surprise measures indistinguishable if whenever $S$ increases $S'$ increases as well. One of our goals in this section is to determine under what conditions different surprise measures are indistinguishable (Fig. 2B).

Based on how they depend on an agent’s belief $\pi^{(t)}$, we divide existing surprise measures into three categories: (i) probabilistic mismatch, (ii) observation-mismatch, and (iii) belief-mismatch surprise measures (Fig. 3). Probabilistic mismatch surprise measures depend on the belief $\pi^{(t)}$ through the marginal prob-

---

\(^1\)For example, assume that $S = f(S')$ for an invertible function $f$. If an estimator of the variable $Z$ is found using the measure $S$ as $\tilde{Z} = g(S)$, then we can rewrite the same estimator in terms of $S'$ as $\tilde{Z} = \tilde{g}(S') = g(f^{-1}(S))$. Because $g(S)$ and $\tilde{g}(S')$ essentially have the same explanatory power given any function $g$ and any measure of performance, the two surprise measures $S$ and $S'$ are equally informative about the variable $Z$ in this regard. However, this is not necessarily true if one restricts the estimators to a particular class of functions – e.g., if the estimators are constrained to be linear with respect to surprise measures while $f$ is nonlinear. Such limitations can be avoided by using non-parametric statistical methods like Spearman or Kendall correlations (Corder & Foreman, 2014).
ability \( P(y_{t+1}|x_{t+1}; \pi^{(t)}) \); an example is the Shannon surprise (Shannon, 1948). Observation-mismatch surprise measures depend on the belief \( \pi^{(t)} \) through some estimate of the next observation \( \hat{y}_{t+1} \) according to the belief \( \pi^{(t)} \); an example is the absolute difference between \( y_{t+1} \) and \( \hat{y}_{t+1} \) (Rouhani & Niv, 2021). To compute the belief-mismatch surprise measures, however, we need to have the whole distribution \( \pi^{(t)} \); an example is the Bayesian surprise (Baldi, 2002; Schmidhuber, 2010).

Probabilistic mismatch surprise 1: Bayes Factor surprise

An abrupt change in the parameters of the environment influences the sequence of observations. Therefore, a sensible way to define the surprise of an observation is that ‘surprise’ measures the probability of an abrupt change in the eye of the agent, given the present observation. To detect an abrupt change, it is not enough to measure how unexpected the observation is according to the current belief of the agent. Rather, the agent should measure how much more expected the new observation is under the prior belief than under the current belief. The Bayes Factor surprise was introduced by Liakoni et al., 2021 to quantify this concept of surprise.

Here, we apply their definition to our generative model. Similar to Xu et al., 2021, we define the Bayes Factor surprise of observing \( y_{t+1} \) given the cue \( x_{t+1} \) as the ratio of the marginal probability of observing \( y_{t+1} \) given \( x_{t+1} \) and \( C_{t+1} = 1 \) (i.e., assuming a change) to the marginal probability of observing \( y_{t+1} \) given \( x_{t+1} \) and \( C_{t+1} = 0 \) (i.e. assuming no change); formally, we write

\[
S_{\text{BF}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \frac{P^{(t)}(y_{t+1}|x_{t+1}, C_{t+1} = 1) \pi^{(0)}}{P^{(t)}(y_{t+1}|x_{t+1}, C_{t+1} = 0) \pi^{(t)}}. \tag{5}
\]

The name arises because \( S_{\text{BF}}(y_{t+1}|x_{t+1}; \pi^{(t)}) \) is the Bayes Factor (Bayarri & Berger, 1997; Efron & Hastie, 2016; Kass & Raftery, 1995) used in statistics to test whether a change has occurred at time \( t \). For a given \( P(y_{t+1}|x_{t+1}; \pi^{(0)}) \), the Bayes Factor surprise is a decreasing function of \( P(y_{t+1}|x_{t+1}; \pi^{(t)}) \): Hence, more probable events are perceived as less surprising. However, the key feature of \( S_{\text{BF}}(y_{t+1}|x_{t+1}; \pi^{(t)}) \) is that it measures not only how unexpected (unlikely) the observation \( y_{t+1} \) is according to the current belief \( \pi^{(t)} \)
but also how expected it would be if the agent had reset its belief to the prior belief. More precisely, for a given $P(y_{t+1}|x_{t+1}; \pi^{(t)})$, the Bayes Factor surprise is an increasing function of $P(y_{t+1}|x_{t+1}; \pi^{(0)})$.

Such a comparison is necessary to evaluate whether a reset of the belief (or an increase in the update rate of the belief) can be beneficial in order to have a more accurate estimate of the environment’s parameters (c.f. Soltani and Izquierdo, 2019). This intuition is formulated in a precise way by Liakoni et al., 2021 in their Proposition 1, where they show that, in the generative model of Fig. 1B, the exact Bayesian inference for the update of $\pi^{(t)}$ to $\pi^{(t+1)}$ upon observing $y_{t+1}$ leads to a learning rule modulated by the Bayes Factor surprise. Proposition 1 below states that this result is also true for our more general generative model (Definition 1 and Fig. 1A).

**Proposition 1.** (Extension of Proposition 1 of Liakoni et al., 2021) For the generative model of Definition 1, the Bayes Factor surprise can be used to write the updated (according to exact Bayesian inference) belief $\pi^{(t+1)}$, after observing $y_{t+1}$ with the cue $x_{t+1}$, as

$$\pi^{(t+1)}(\theta) = (1 - \gamma_{t+1})\pi^{(t+1)}_{\text{integration}}(\theta) + \gamma_{t+1}\pi^{(t+1)}_{\text{reset}}(\theta),$$

where $\gamma_{t+1}$ is an adaptation rate modulated by the Bayes Factor surprise

$$\gamma_{t+1} = \frac{mS_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)})}{1 + mS_{BF}(y_{t+1}|x_{t+1}; \pi^{(0)})} \quad \text{with} \quad m = \frac{p_c}{1 - p_c},$$

and

$$\pi^{(t+1)}_{\text{integration}}(\theta) = \frac{P_{Y|X}(y_{t+1}|x_{t+1}; \theta)\pi^{(t)}(\theta)}{P(y_{t+1}|x_{t+1}; \pi^{(t)})}, \quad \text{and} \quad \pi^{(t+1)}_{\text{reset}}(\theta) = \frac{P_{Y|X}(y_{t+1}|x_{t+1}; \theta)\pi^{(0)}(\theta)}{P(y_{t+1}|x_{t+1}; \pi^{(0)})}.$$
Figure 3: Categorization of surprise measures based on the form of their dependence upon the belief. Surprise depends on expectations. Therefore, all surprise measures depend on the belief \( \pi(t) \). However, the specific form of the dependence changes between one measure and another. ‘Observation-mismatch’ surprise measures use the marginal distribution \( P(y_{t+1}|x_{t+1}; \pi(t)) \) to calculate an estimate \( \hat{y}_{t+1} \) of the next observation, which is then compared with the real observation \( y_{t+1} \) by an error function such as \( ||\hat{y}_{t+1} - y_{t+1}||_1 \), where \( ||.||_1 \) stands for \( \ell_1 \)-norm. ‘Probabilistic mismatch’ surprise measures use the marginal probability \( P(y_{t+1}|x_{t+1}; \pi(t)) \) directly, without extracting a specific estimate. ‘Belief-mismatch’ surprise measures use the belief \( \pi(t) \) directly, without extracting the marginal distribution \( P(y_{t+1}|x_{t+1}; \pi(t)) \).

Abbreviations: Ab: Absolute error surprise, Sq: Squared error surprise, BF: Bayes Factor surprise, Sh: Shannon surprise, SPE: State Prediction Error, Ba: Bayesian surprise, Po: Postdictive surprise, CC: Confidence Corrected surprise, and \( F^* \): Minimized Free Energy. See section Theories of surprise: A technical review for details.

Therefore, the Bayes Factor surprise \( S_{BF} \) controls the trade-off between the integration of the new observation into the old belief (via \( \pi_{\text{integration}}^{(t+1)} \)) and resetting the old belief to the prior belief (via \( \pi_{\text{reset}}^{(t+1)} \)) – c.f. Liakoni et al., 2021.

Probabilistic mismatch surprise 2: Shannon surprise

No matter if there has been an abrupt change (\( C_{t+1} = 1 \)) or not (\( C_{t+1} = 0 \)), an unlikely event may be perceived as surprising. Therefore, another way to measure the surprise of an observation is to quantify how unlikely the observation is in the eye of the agent. Shannon surprise (Shannon, 1948), also known as surprisal (Tribus, 1961), is a way to formalize this concept of surprise. It comes from the field of information theory (Shannon, 1948) and statistical physics (Tribus, 1961) and is widely used in neuroscience (Gijsen et al., 2021; Kolossa et al., 2015; Konovalov & Krajbich, 2018; Kopp & Lange, 2013; Maheu et al., 2019; Mars et al., 2008; Meyniel et al., 2016; Modirshanechi et al., 2019; Mousavi et al., 2020; Visalli et al., 2021).

Formally, in the generative model of Definition 1, one can define the Shannon surprise of observing \( y_{t+1} \) given the cue \( x_{t+1} \) as

\[
S_{Sh1}(y_{t+1}|x_{t+1}; \pi(t)) = -\log P(y_{t+1}|x_{t+1}) \\
= -\log \left( p_c P(y_{t+1}|x_{t+1}; \pi(0)) + (1 - p_c) P(y_{t+1}|x_{t+1}; \pi(t)) \right).
\]  

The Shannon surprise \( S_{Sh1} \) measures how unexpected or unlikely \( y_{t+1} \) is considering the possibility that
there might have been an abrupt change in the environment. As a result, for a fixed $P(y_{t+1}|x_{t+1}; \pi^{(t)})$, the Shannon surprise is a decreasing function of $P(y_{t+1}|x_{t+1}; \pi^{(0)})$ (c.f. Eq. 9), whereas the Bayes Factor surprise is an increasing function of $P(y_{t+1}|x_{t+1}; \pi^{(0)})$ (c.f. Eq. 5). This essential difference between the Shannon and the Bayes Factor surprise has been exploited by Liakoni et al., 2021 in order to propose experiments where these two measures of surprise make different predictions.

Experimental evidence (Nassar et al., 2012; Nassar et al., 2010; Wilson et al., 2013) indicates that in volatile environments like the one in Fig. 1B, human participants do not actively consider the possibility that there may be an abrupt change while predicting the next observation $y_{t+1}$ even though they update their belief after observing $y_{t+1}$ by considering the possibility that there might have been a change before the current observation at time $t+1$. To arrive at a Shannon surprise measure consistent with this observation, we suggest a 2nd definition as

$$S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = - \log \mathbb{P}^{(t)}(y_{t+1}|x_{t+1}, C_{t+1} = 0) = - \log P(y_{t+1}|x_{t+1}; \pi^{(t)}). \quad (10)$$

In other words, $S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)})$ neglects the potential presence of change-points, and, therefore, it is independent of both $p_c$ and $P(y_{t+1}|x_{t+1}; \pi^{(0)})$. For a non-volatile environment that does not allow for abrupt changes ($p_c = 0$), the two definitions of Shannon surprise are identical: $S_{\text{Sh}1} = S_{\text{Sh}2}$ (Fig. 2B).

Proposition 2 shows that the Bayes Factor surprise $S_{\text{BF}}$ is related to $S_{\text{Sh}1}$ and $S_{\text{Sh}2}$:

**Proposition 2. (Relation between the Shannon surprise and the Bayes Factor surprise)** In the generative model of Definition 1, the Bayes Factor surprise $S_{\text{BF}}(y_{t+1}|x_{t+1}; \pi^{(t)})$ can be written as

$$S_{\text{BF}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \frac{(1 - p_c)e^{\Delta S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)})}}{1 - p_c e^{\Delta S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)})}} = e^{\Delta S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)})}, \quad (11)$$

where

$$\Delta S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) - S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(0)}),$$

$$\Delta S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)}) - S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(0)}). \quad (12)$$

Proposition 2 states that the Bayes Factor $S_{\text{BF}}(y_{t+1}|x_{t+1}; \pi^{(t)})$ has a behavior similar to the difference in Shannon surprise (i.e., $\Delta S_{\text{Sh}1}$ or $\Delta S_{\text{Sh}2}$) as opposed to Shannon surprise itself (i.e., $S_{\text{Sh}1}$ or $S_{\text{Sh}2}$). The difference in Shannon surprise (i.e., $\Delta S_{\text{Sh}1}$ or $\Delta S_{\text{Sh}2}$) compares the Shannon surprise under the current belief with that under the prior belief. Two direct consequences of this proposition are summarized in Corollaries 1 and 2.

Corollary 1 states that the modulation of learning as presented in Proposition 1 can also be written in the form of the difference in Shannon surprise (i.e., $\Delta S_{\text{Sh}1}$ or $\Delta S_{\text{Sh}2}$).

**Corollary 1.** The adaptation rate $\gamma_{t+1}$ in Proposition 1 can be written as

$$\gamma_{t+1} = p_c \exp \left( \Delta S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) \right) \quad \text{or} \quad \gamma_{t+1} = \text{Sigmoid} \left( \tilde{m} \Delta S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)}) \right), \quad (13)$$

with $\tilde{m} = \log \frac{p_c}{1-p_c} = \log m$ (c.f. Proposition 1) and $\text{Sigmoid}(u) = \frac{1}{1+e^{-u}}$

Corollary 2 indicates that, under a flat prior, the Bayes Factor surprise and the two definitions of the Shannon surprise are indistinguishable from each other (Fig. 2B):

**Corollary 2.** (Flat prior prediction) In the generative model of Definition 1, if the probability of observing $y_{t+1}$ with the cue $x_{t+1}$ is flat under the prior belief $\pi^{(0)}$ (i.e., if $P(y_{t+1}|x_{t+1}; \pi^{(0)})$ is uniform), then there are one-to-one mappings between $S_{\text{BF}}(y_{t+1}|x_{t+1}; \pi^{(t)})$, $S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)})$, and $S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)})$.

A consequence of Corollary 2 is that experimental paradigms with flat priors of the agent cannot be used to distinguish $S_{\text{BF}}$ from $S_{\text{Sh}1}$ or $S_{\text{Sh}2}$ (c.f. Fig. 2).
Probabilistic mismatch surprise 3: State prediction error

The State Prediction Error (SPE) was introduced by Gläscher et al., 2010 in the context of model-based reinforcement learning in Markov Decision Processes (MDPs – c.f. Fig. 1E) (Daw et al., 2011; Daw et al., 2005; Sutton & Barto, 2018). Similar to the Shannon surprise, the SPE considers less probable events as the more surprising ones.

Whenever observations \( y_{t+1} \) come from a discrete distribution so that we have \( P_{Y|X}(y_{t+1}|x_{t+1}; \theta) \in [0, 1] \) for all \( \theta, x_{t+1}, \) and \( y_{t+1} \), we can generalize the definition of Gläscher et al., 2010 to the setting of our generative model. Analogously to our two definitions of Shannon surprise (c.f. Eq. 9 and Eq. 10), we give also two definitions for SPE:

\[
S_{SPE1}(y_{t+1}|x_{t+1}; \pi^{(i)}) = 1 - P(y_{t+1}|x_{t+1})
= 1 - \left( p_c P(y_{t+1}|x_{t+1}; \pi^{(0)}) + (1 - p_c) P(y_{t+1}|x_{t+1}; \pi^{(i)}) \right),
\]

(14)

and

\[
S_{SPE2}(y_{t+1}|x_{t+1}; \pi^{(i)}) = 1 - P(y_{t+1}|x_{t+1}; C_{t+1} = 0) = 1 - P(y_{t+1}|x_{t+1}; \pi^{(i)}).
\]

(15)

In non-volatile environments \( (p_c = 0) \), the two definitions of SPE are identical (Fig. 2B). In particular, in an MDP without abrupt changes \( (p_c = 0; \text{ Fig. 1E}) \), both definitions are equal to \( 1 - \mathbb{P}(s_t, a_t \rightarrow s_{t+1}) \), where \( \mathbb{P}(s_{t}, a_{t} \rightarrow s_{t+1}) \) is an agent’s estimate (at time \( t \)) of the probability of the transition to state \( s_{t+1} \) after taking action \( a_t \) in state \( s_t \); c.f. Gläscher et al., 2010.

Proposition 3 states that both definitions \( S_{SPE1} \) and \( S_{SPE2} \) can always be written as invertible functions of Shannon surprise (Fig. 2B):

**Proposition 3.** (Relation between the Shannon surprise and the SPE) In the generative model of Definition 1, for \( i \in \{1, 2\} \), the state prediction error \( S_{SPEi}(y_{t+1}|x_{t+1}; \pi^{(i)}) \), can be written as

\[
S_{SPEi}(y_{t+1}|x_{t+1}; \pi^{(i)}) = 1 - \exp\left( -S_{Sh}(y_{t+1}|x_{t+1}; \pi^{(i)}) \right).
\]

(16)

Therefore, the SPE and the Shannon surprise are in principle indistinguishable (Fig. 2).

Observation-mismatch surprise: Absolute and squared errors

Assume an agent predicts \( \hat{y}_{t+1} \) for the next observation \( y_{t+1} \). Then, a measure of surprise can be defined as the prediction error or the mismatch between the prediction \( \hat{y}_{t+1} \) and the actual observation \( y_{t+1} \) (Heydari & Holroyd, 2016; Nassar et al., 2010; Prat-Carrabin et al., 2021; Rouhani et al., 2020; Talmi et al., 2013) (Fig. 3). For the sake of completeness, we discuss four possible definitions for observation-mismatch surprise measures in this section.

Before turning to an ‘observation-mismatch’, we first need to define an agent’s prediction for the next observation. Analogously to our two definitions for the Shannon surprise (c.f. Eq. 9 and Eq. 10), we define two different predictions for the next observation \( y_{t+1} \) given the cue \( x_{t+1} \):

\[
E_1[Y_{t+1}] = p_c \mathbb{E}_{P(\cdot|x_{t+1}; \pi^{(0)})}[Y_{t+1}] + (1 - p_c) \mathbb{E}_{P(\cdot|x_{t+1}; \pi^{(i)})}[Y_{t+1}]
\]

(17)

and

\[
E_2[Y_{t+1}] = \mathbb{E}_{P(\cdot|x_{t+1}; \pi^{(i)})}[Y_{t+1}].
\]

(18)

Although \( E_1[Y_{t+1}] \) is a more reasonable prediction for \( y_{t+1} \) given the fact that there is always a possibility of an abrupt change according to our generative model of the environment (Definition 1), Nassar et al., 2010 have shown that, in a Gaussian task, \( E_2[Y_{t+1}] \) explains human participants’ predictions better than \( E_1[Y_{t+1}] \).
We note that the observation $y_{t+1}$ is in general case multi-dimensional (Niv et al., 2015). As two natural ways of measuring mismatch, we define the squared and the absolute error surprise, for $i \in \{1, 2\}$ as

$$S_{Ab,i}(y_{t+1}|x_{t+1}; \pi(t)) = ||y_{t+1} - E_i[y_{t+1}]||_1,$$

$$S_{Sq,i}(y_{t+1}|x_{t+1}; \pi(t)) = ||y_{t+1} - E_i[y_{t+1}]||_2^2,$$

where $||.||_1$ and $||.||_2$ stand for the $\ell_1$- and $\ell_2$-norms, respectively, and $E_1$ and $E_2$ are defined in Eq. 17 and Eq. 18, respectively. Similar definitions have been used in neuroscience (Heydari & Holroyd, 2016; Nassar et al., 2010; Prat-Carrabin et al., 2021; Talmi et al., 2013) and machine learning (Burda et al., 2015; Pathak et al., 2017). For two special cases, we show in Propositions 4 and 5 below that the absolute and the squared error surprise can be written as functions of the SPE and the Shannon surprise (Fig. 2B).

**Proposition 4.** (Relation between the observation-mismatch surprise measures and the SPE for categorical distributions) In the generative model of Definition 1, if $Y_{t+1}$ is represented as one-hot coded vectors, i.e., vectors with one element equal to 1 and the others equal to 0, then we have, for $i \in \{1, 2\}$,

$$S_{Ab,i}(y_{t+1}|x_{t+1}; \pi(t)) = 2S_{SPE,i}(y_{t+1}|x_{t+1}; \pi(t)),$$

and

$$S_{Sq,i}(y_{t+1}|x_{t+1}; \pi(t)) = 2S_{SPE,i}(y_{t+1}|x_{t+1}; \pi(t)) + \text{Conf.}\left[P(.|x_{t+1}; \pi(t))\right],$$

where $\text{Conf.}\left[P(.|x_{t+1}; \pi(t))\right]$ can be seen as a measure of confidence in the prediction (see Appendix B: Proofs).

**Proposition 5.** (Relation between the squared error surprise and the Shannon surprise for Gaussian distributions – from Pathak et al., 2017) In the generative model of Definition 1, if the marginal distribution of $Y_{t+1} \in \mathbb{R}^N$ given the cue $x_{t+1}$ and under the belief $\pi(t)$ is a Gaussian distribution with a covariance matrix equal to $\sigma I_{N \times N}$, where $I_{N \times N}$ is the $N \times N$ identity matrix, then $S_{S}\pi(2)(y_{t+1}|x_{t+1}; \pi(t))$ is an invertible function of $S_{Sh2}(y_{t+1}|x_{t+1}; \pi(t))$.

We note that, according to Proposition 3, the SPE is an invertible function of the Shannon surprise. Hence, for categorical distributions with one-hot coding, the SPE, the Shannon surprise, and the absolute error surprise are indistinguishable, and for Gaussian distributions with unit diagonal covariance, the SPE, the Shannon surprise, and the squared error surprise are indistinguishable (Fig. 2).

**Belief-mismatch surprise 1: Bayesian surprise**

Another way to think about surprise is to define surprising events as those that change an agent’s belief about the world (Baldi, 2002; Storck et al., 1995). Bayesian surprise or information gain (Baldi & Itti, 2010; Itti & Baldi, 2006, 2009; Schmidhuber, 2010) is a way to formalize this concept of surprise. Whereas the Bayes Factor surprise measures how likely it is that the environment has changed given the new observation, the Bayesian surprise measures how much the agent’s belief changes given the new observation.

Bayesian surprise (Baldi, 2002; Itti & Baldi, 2006, 2009; Schmidhuber, 2010; Storck et al., 1995) has been originally introduced in non-volatile environments, i.e., where there is no change ($p_c = 0$) and as a result $\Theta_1 = \Theta_2 = \ldots = \Theta_t = \Theta$. In this case, the Bayesian surprise of observing $y_{t+1}$ with cue $x_{t+1}$ is defined as $D_{KL}[P^{(t)}(\Theta) || P^{(t+1)}(\Theta)]$ (Baldi, 2002; Itti & Baldi, 2006, 2009; Schmidhuber, 2010; Storck et al., 1995), where $D_{KL}$ stands for the Kullback-Leibler (KL) divergence (Cover, 1999) (c.f. Table 1 for notation).

Hence, in non-volatile environments, Bayesian surprise measures the pseudo-distance $D_{KL}$ between two distributions, i.e., the belief $\pi(t) = P^{(t)}(\Theta)$ before and the belief $\pi(t+1) = P^{(t+1)}(\Theta)$ after observing $y_{t+1}$. To generalize this definition to volatile environments, we have to choose the equivalent two distributions that
we want to compare. The natural choice for $\mathbb P_{\Theta}^{(t+1)}$ is $\mathbb P_{\Theta}^{(t+1)} = \pi^{(t+1)}$; however, it is unclear whether $\mathbb P_{\Theta}^{(t)}$ should be taken as the momentary belief $\mathbb P_{\Theta}^{(t)} = \pi^{(t)}$ or its one-step forward-propagation $\mathbb P_{\Theta}^{(t)}$ before the next observation $y_{t+1}$ is integrated. If $p_c \neq 0$, the two choices are different:

$$\pi^{(t)} = \mathbb P_{\Theta}^{(t)} \neq \mathbb P_{\Theta}^{(t)} = p_c \pi^{(0)} + (1 - p_c) \pi^{(t)}.$$  

(22)

Therefore, for the case of volatile environments, we give two definitions for the Bayesian surprise:

$$S_{Ba1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL}[p_c \pi^{(0)} + (1 - p_c) \pi^{(t)}||\pi^{(t+1)}],$$  

(23)

and

$$S_{Ba2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL}[\pi^{(t)}||\pi^{(t+1)}].$$  

(24)

The 1st definition is more consistent with the original definition of the Bayesian surprise (Baldi, 2002; Itti & Baldi, 2006, 2009; Schmidhuber, 2010; Storck et al., 1995) applied to our generative model because the belief before the observation should include the knowledge that the environment is volatile. However, the 2nd definition looks more intuitive from the neuroscience perspective (Gijzen et al., 2021; Mousavi et al., 2020). Note that, in Eq. 23 and Eq. 24, the observation $y_{t+1}$ does not appear explicitly on the right hand side; the observation has, however, influenced the update of the belief to its new distribution $\pi^{(t+1)}$. For the case of $p_c = 0$, the two definitions are identical (Fig. 2B).

In Lemma 1 and Remark 1 in Appendix B: Proofs, we show that the Bayesian surprise is correlated with the difference between the Shannon surprise and its expectation (over all possible values of $\Theta_{t+1}$). There are two consequences of this observation. First, Bayesian surprise is distinguishable from Shannon surprise since it cannot be found only as a function of Shannon surprise. Second, we need access to the full belief distribution $\pi^{(t)}$ for computing the expectation (Fig. 3).

In general, surprise measures similar to the Bayesian surprise can be defined also by measuring the change in the belief via distance or pseudo-distance measures different from the KL-divergence (Baldi, 2002; Baldi & Itti, 2010).

**Belief-mismatch surprise 2: Postdictive surprise**

We saw that the Bayesian surprise measures how much the new belief $\pi^{(t+1)}$ has changed after observing $y_{t+1}$. Kolossa et al., 2015 introduced ‘postdictive surprise’ with a similar idea in mind but focused on changes in the marginal distribution $P(y|x_{t+1}; \pi^{(t+1)})$ (c.f. Eq. 4). More precisely, whereas the Bayesian surprise measures the amount of update in the space of distributions over the parameters (i.e., how differently the agent thinks about the parameters), the postdictive surprise measures the amount of update in the space of distributions over the observations (i.e., how differently the agent predicts the next observations).

Analogous to our two definitions for the Bayesian surprise (c.f. Eq. 23 and Eq. 24), there are two definitions for the postdictive surprise in volatile environments:

$$S_{Po1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL}[p_c P(.|x_{t+1}; \pi^{(0)}) + (1 - p_c) P(.|x_{t+1}; \pi^{(t)})||P(.|x_{t+1}; \pi^{(t+1)})],$$  

(25)

and

$$S_{Po2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL}[P(.|x_{t+1}; \pi^{(t)})||P(.|x_{t+1}; \pi^{(t+1)})],$$  

(26)

where the dot refers to a dummy variable $y$ that is integrated out when evaluating $D_{KL}$. Note that for $p_c = 0$, the two definitions are identical (Fig. 2B).

Although the amount of update is computed over the space of observations, $S_{Po1}$ and $S_{Po2}$ cannot be categorized as probabilistic mismatch surprise measures, since the update depends explicitly on the belief $\pi^{(t)}$. The statement becomes more obvious in our Lemma 2 in Appendix B: Proofs.
Belief-mismatch surprise 3: Confidence Corrected surprise

Since surprise arises when an expectation is violated, the violation of an agent’s expectation should be more surprising when the agent is more confident about its expectation. Based on the observation that neither Shannon nor Bayesian surprise explicitly captures the concept of confidence, Faraji et al., 2018 proposed the ‘Confidence Corrected Surprise’ as a new measure of surprise that explicitly takes confidence into account.

To define the Confidence Corrected surprise, we first define $\pi_{\text{flat}}$ as the flat (uniform) distribution over the space of parameters, i.e., over the set to which $\Theta$ belongs. Then, following Faraji et al., 2018, we define the normalized likelihood after observing $y_{t+1}$ (i.e., the posterior given the flat prior) as

$$
\pi_{\text{flat}}(\theta|y_{t+1}, x_{t+1}) = \frac{P_{Y|X}(y_{t+1}|x_{t+1}; \theta)\pi_{\text{flat}}(\theta)}{P(y_{t+1}|x_{t+1}; \pi_{\text{flat}})} = \frac{P_{Y|X}(y_{t+1}|x_{t+1}; \theta)}{\int P_{Y|X}(y_{t+1}|x_{t+1}; \theta)d\theta}.
$$

(27)

If the prior $\pi^{(0)}$ is equal to $\pi_{\text{flat}}$ (i.e., if the prior is uniform), then $\pi_{\text{flat}}(\theta|y_{t+1}, x_{t+1})$ is the same as $\pi_{\text{reset}}^{(t+1)}(\theta)$ defined in Proposition 1. Note that the prior $\pi_{\text{flat}}$ does not necessarily need to be a proper distribution (i.e., does not necessarily need to be normalized) as long as $\int P_{Y|X}(y_{t+1}|x_{t+1}; \theta)d\theta$ is finite and the posterior $\pi_{\text{flat}}(\theta|y_{t+1}, x_{t+1})$ is a proper distribution (Efron & Hastie, 2016). Using this terminology, the original definition for the Confidence Corrected surprise is (Faraji et al., 2018)

$$
S_{\text{CC1}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{\text{KL}}\left[\pi^{(t)}||\pi_{\text{flat}}(.|y_{t+1}, x_{t+1})\right].
$$

(28)

To interpret $S_{\text{CC1}}$, Faraji et al., 2018 defined the commitment (or confidence) $C[\pi]$ corresponding to an arbitrary belief $\pi$ as its negative entropy (Cover, 1999), i.e.,

$$
C[\pi] = \mathbb{E}_\pi \left[ \log \pi(\Theta) \right].
$$

(29)

Then, in a non-volatile environment (i.e., $p_c = 0$), they show that $S_{\text{CC1}}$ can be written as (Faraji et al., 2018)

$$
S_{\text{CC1}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Sh}}(y_{t+1}|x_{t+1}; \pi^{(t)}) + S_{B_1}(y_{t+1}|x_{t+1}; \pi^{(t)}) + C[\pi^{(t)}] - A(y_{t+1}, x_{t+1}),
$$

(30)

where $A(y_{t+1}, x_{t+1}) = S_{\text{Sh}}(y_{t+1}|x_{t+1}; \pi_{\text{flat}}) + C[\pi_{\text{flat}}]$ is independent of the current belief $\pi^{(t)}$; note that because $p_c = 0$, we have $S_{\text{Sh}1} = S_{\text{Sh}2}$ and $S_{B_1} = S_{B_2}$. Therefore, in a non-volatile environment (i.e., $p_c = 0$), $S_{\text{CC1}}$ is correlated with the sum of the Shannon and the Bayesian surprise regularized by the confidence of the agent’s belief. However, such an interpretation is no longer possible in volatile environments ($p_c > 0$), and Eq. 30 must be replaced by Proposition 6 below.

In order to account for the information of the true prior $\pi^{(0)}$ and to avoid the cases where $\pi_{\text{flat}}(.|y_{t+1}, x_{t+1})$ is not a proper distribution, we also give a 2nd definition for the Confidence Corrected surprise as

$$
S_{\text{CC2}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{\text{KL}}\left[\pi^{(t)}||\pi_{\text{reset}}^{(t+1)}\right],
$$

(31)

where $\pi_{\text{reset}}^{(t+1)}(\theta)$ is defined in Proposition 1. Whenever $\pi^{(0)} = \pi_{\text{flat}}$, the two definitions are identical (Fig. 1B). Proposition 6 shows how the Confidence Corrected surprise relates to the Shannon surprise, the Bayesian surprise, and the confidence in the general case.

Proposition 6. (Relation between the Confidence Corrected surprise, Shannon surprise, and Bayesian surprise) For the generative model of Definition 1, the original definition of the Confidence Corrected surprise can be written as

$$
S_{\text{CC1}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) - S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi_{\text{flat}})
$$

$$
+ S_{B_1}(y_{t+1}|x_{t+1}; \pi^{(t)}) - D_{\text{KL}}\left[\pi^{(t)}||c, p_c \pi^{(0)} + (1 - p_c) \pi^{(t)}\right] + C[\pi^{(t)}] - C[\pi_{\text{flat}}],
$$

(32)
Proposition 6 conveys three important messages. First, both definitions of the Confidence Corrected surprise depend on differences in the Shannon surprise as opposed to the Shannon surprise itself (c.f. first line in Eq. 32 and Eq. 33). Second, both definitions depend on the difference between the Bayesian surprise (i.e., the change in the belief given the new observation) and the a priori expected change in the belief (because of the possibility of a change in the environment; c.f. second line in Eq. 32 and Eq. 33). Third, both definitions regularize the contributions of Shannon surprise and Bayesian surprise by the relative confidence of the current belief compared to either the flat or the prior belief (c.f. third line in Eq. 32 and Eq. 33). ‘Relative confidence’ quantifies how different the current belief is with respect to a reference belief; note that \( C[\pi^{(t)}] - C[\pi_{\text{flat}}] = D_{\text{KL}}[\pi^{(t)}||\pi_{\text{flat}}]. \)

Hence, the Confidence Corrected surprise should be distinguishable from both the Shannon and the Bayesian surprise (for \( p_c < 1 \)). An interesting consequence of Proposition 6, however, is that \( S_{\text{CC2}} \) is identical to \( S_{\text{Ba2}} \) when the environment becomes so volatile that its parameter changes at each time step (i.e., in the limit of \( p_c \to 1 \)).

**Corollary 3.** For the generative model of Definition 1, when \( p_c \to 1 \), we have \( S_{\text{CC2}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Ba2}}(y_{t+1}|x_{t+1}; \pi^{(t)}) \).

**Belief-mismatch surprise 4: Minimized free energy**

Although an agent can perform computations over the joint probability distribution in Eq. 1 and Eq. 2, finding the belief \( \pi^{(t+1)}(\theta) \) (i.e., the posterior distribution in Eq. 3) can be computationally intractable (Barber, 2012; Liakoni et al., 2021). Therefore, it has been argued that the brain uses approximate inference (instead of exact Bayesian inference) for finding the belief (Daw & Courville, 2008; Faraji et al., 2018; Findling et al., 2021; Fiser et al., 2010; Friston, 2010; Friston et al., 2017; Liakoni et al., 2021; Mathys et al., 2011). An approximation of the belief \( \pi^{(t+1)}(\theta) \) can for example be found via variational inference (Blei et al., 2017; MacKay, 2003) over a family of distributions \( q(\theta; \phi) \) parameterized by \( \phi \). Such approaches are popular in neuroscience studies of learning and inference in the brain (Friston, 2010; Friston et al., 2017; Gershman, 2019).

Formally, in variational inference, the belief \( \pi^{(t+1)}(\theta) \) is approximated by \( \tilde{\pi}^{(t+1)}(\theta) = q(\theta; \phi^{(t+1)}) \), where \( \phi^{(t+1)} \) is the minimizer of the variational loss or free energy, i.e., \( \phi^{(t+1)} = \arg\min_\phi F^{(t+1)}(\phi) \) (Friston, 2010; Friston et al., 2017; Liakoni et al., 2021; MacKay, 2003; Markovic et al., 2021; Sajid et al., 2021).

To define \( F^{(t+1)}(\phi) \), we introduce a new notation:

\[
\mathbb{P}_{\Theta^{t+1}}(\theta, y_{t+1}|x_{t+1}; \pi) = P_{Y|X}(y_{t+1}|x_{t+1}; \theta)(p_c\pi^{(0)}(\theta) + (1 - p_c)\pi(\theta)),
\]

where \( \pi(\theta) \) is an arbitrary distribution over the parameter space. Using this notation, we can write the joint distribution over the observation and the parameter \( \mathbb{P}^{(t)}(\theta_{t+1}, y_{t+1}|x_{t+1}) = \mathbb{P}_{\Theta^{t+1}}(\theta, y_{t+1}|x_{t+1}; \pi^{(t)}) \) and the updated belief \( \pi^{(t+1)}(\theta) \) as \( \mathbb{P}_{\Theta^{t+1}}(\theta|y_{t+1}, x_{t+1}; \pi^{(t)}) \). The variational loss or free energy can then be defined as (Friston, 2010; Friston et al., 2017; Liakoni et al., 2021; Markovic et al., 2021; Sajid et al., 2021)

\[
F^{(t+1)}(\phi) = \mathbb{E}_{q(\cdot; \phi)} \left[ \log q(\Theta; \phi) - \log \mathbb{P}_{\Theta^{t+1}}(\Theta, y_{t+1}|x_{t+1}; \tilde{\pi}^{(t)}) \right].
\]
For any value of $\phi$, one can show that (Blei et al., 2017; Sajid et al., 2021)

$$F^{(t+1)}(\phi) = D_{KL}\left[q(\cdot; \phi)||\mathbb{P}_{\Theta_t+1}(\cdot|y_{t+1}, x_{t+1}; \hat{\pi}(t))\right] + S_{Sh1}(y_{t+1}|x_{t+1}; \hat{\pi}(t))$$

where the right side of the inequality is independent of $\phi$, and $\mathbb{P}_{\Theta_t+1}(\cdot|y_{t+1}, x_{t+1}; \hat{\pi}(t))$ is the exact Bayesian update of the belief (according to the generative model in Definition 1) given the latest approximation of the belief $\hat{\pi}(t)$ (Liakoni et al., 2021; Markovic et al., 2021).

The minimized free energy $F^* = \min_{\phi} F^{(t+1)}(\phi)$ has been interpreted as a measure of surprise (Friston, 2010; Friston et al., 2017; Schwartenbeck et al., 2013), which, according to Eq. 36, can be seen as an approximation of $S_{Sh1}(y_{t+1}|x_{t+1}; \hat{\pi}(t))$. The parametric family of $q(\cdot; \phi)$ and its relation to the exact belief $\pi^{(t+1)}$ determine how well $F^*$ approximates $S_{Sh1}(y_{t+1}|x_{t+1}; \hat{\pi}(t))$ (Fig. 2B). More precisely, the minimized free energy measures both how unlikely the new observation is (i.e., how large $S_{Sh1}(y_{t+1}|x_{t+1}; \hat{\pi}(t))$ is) and how imprecise the best parametric approximation of the belief $\hat{\pi}^{(t+1)}$ is (i.e., how large $D_{KL}[\hat{\pi}^{(t+1)}|| \mathbb{P}_{\Theta_t+1}(\cdot|y_{t+1}, x_{t+1}; \hat{\pi}(t))]$ is). Therefore, the minimized free energy is in the category of belief-mismatch surprise measures (Fig. 3).

A brief review of experimental results

The surprise definitions discussed above have been used in various experiments with the goal of identifying neural and behavioral signatures of surprise. In this subsection, we briefly review some of the existing experimental results.

The neural signatures of the Shannon surprise (and the state prediction error, c.f. Proposition 3) have been found in numerous EEG (Gijsen et al., 2021; Kolossa et al., 2015; Kopp & Lange, 2013; Mars et al., 2008; Meyniel et al., 2016; Modirshanechi et al., 2019; Mousavi et al., 2020), MEG (Maheu et al., 2019; Meyniel, 2020; Mousavi et al., 2020), and fMRI (Gläser et al., 2010; Konovalov & Krajbich, 2018; Loued-Khenissi & Preuschoff, 2020) studies. In particular, it is well-known that the amplitude of the EEG P300 component correlates with the Shannon surprise (Kolossa et al., 2015; Kopp & Lange, 2013; Mars et al., 2008; Meyniel et al., 2016; Modirshanechi et al., 2019; Mousavi et al., 2020). Modirshanechi et al., 2019 showed that such correlations can be exploited to decode the Shannon surprise of stimuli from EEG signals independently of the stimulus modality. Maheu et al., 2019 showed that the MEG amplitudes at different time windows correlate with different values of the Shannon surprise calculated according to different inference styles. The expected (over observations) Shannon surprise (usually interpreted as uncertainty) has also been shown to have signatures in EEG (Kopp & Lange, 2013), MEG (Meyniel, 2020), and fMRI (Konovalov & Krajbich, 2018; Loued-Khenissi & Preuschoff, 2020).

The signatures of the Bayesian and the Confidence Corrected surprise have been compared to those of the Shannon surprise (Gijsen et al., 2021; Kolossa et al., 2015; Mars et al., 2008; Mousavi et al., 2020; Ostwald et al., 2012; Visalli et al., 2021). Whereas some studies have reported that the EEG and MEG amplitudes were always best explained by only one of these surprise measures (Mars et al., 2008; Mousavi et al., 2020; Ostwald et al., 2012), other studies have reported separate significant contributions of different surprise measures to explaining variations of EEG signals (Gijsen et al., 2021; Kolossa et al., 2015; Visalli et al., 2021). In particular, the amplitude of the EEG P300 component at the frontal and frontocentral electrodes (called P3a component) has been reported to be explained better by the Bayesian surprise than the Shannon surprise, and the amplitude of the EEG P300 component at the parietal electrodes (called P3b component) has been reported to be explained better by the Shannon surprise than the Bayesian surprise (Kolossa et al., 2015; Visalli et al., 2021). In addition to the Bayesian and the Shannon surprise, Kolossa et al., 2015 reported a significant positive correlation between the postdictive surprise and the EEG amplitudes in a later time window (i.e., the EEG slow wave component). Gijsen et al., 2021 focused
on the early EEG components and reported strong evidence for the existence of separate signatures for all Shannon, Bayesian, and Confidence Corrected surprise.

The Bayesian surprise has been shown to attract human attention (Itti & Baldi, 2006, 2009) and drive exploration in reward-driven experiments (Horvath et al., 2021) and has also been used in computational models of human curiosity (Gottlieb & Oudeyer, 2018; Gottlieb et al., 2013; Oudeyer et al., 2016; Schmidhuber, 2010). As mentioned earlier, surprise measures similar to the Bayesian surprise and postdictive surprise can be defined also by measuring the change in the belief and the marginal probability, respectively, via distance or pseudo-distance measures different from the KL-divergence (c.f. Eq. 23, Eq. 24, Eq. 25, and Eq. 26) (Baldi, 2002). Findling et al., 2021 showed that introducing computational noise proportional to such a measure of surprise into the inference mechanism can explain human adaptive behavior in a wide range of experiments. Antony et al., 2021 showed that a similar measure of surprise plays a role in the segmentation of memory and correlates with pupil dilation and the activity of dopamine-related regions of the brain in fMRI. The influence of surprise on memory has also been studied in reward-driven experiments where it has been shown that high unsigned reward prediction errors increase the memorability of task-independent stimuli (Rouhani & Niv, 2021; Rouhani et al., 2018); in such settings, reward can be defined as a component of the observation \( y_t \), and the unsigned reward prediction error can be seen as the absolute error surprise (c.f. Eq. 19).

In volatile environments like the one in Fig. 1B, human behavior has been well explained by exact or approximate Bayesian inference (Behrens et al., 2007; Heilbron & Meyniel, 2019; Nassar et al., 2012; Nassar et al., 2010; Wilson et al., 2013), supporting the idea that the Bayes Factor surprise modulates the speed of learning (c.f. Proposition 1). Nassar et al., 2012 showed that the adaptation rate \( \gamma_{t+1} \) (c.f. Proposition 1) correlates with the changes in pupil diameter in a Gaussian task with abrupt changes. In a recent study, Xu et al., 2021 showed that the Bayes Factor surprise modulates human reinforcement learning in a volatile multi-step decision-making experiment and correlates with the EEG P300 amplitude at frontal electrodes. They showed that such a correlation is independent of the correlation of the reward prediction error and novelty with the EEG P300 amplitude (Xu et al., 2021). Although these observations support the computation and the use of the Bayes Factor surprise in the brain, the Bayes Factor surprise and the Shannon surprise are indistinguishable in the above experiments as they all used a flat distribution for the prior marginal probabilities (c.f. Corollary 2). Liakoni et al., 2021 proposed a modification of the Gaussian task of Nassar et al., 2012 where the Bayes Factor surprise and the Shannon surprise have different behaviors.

Most of these previous studies have focused on one measure of surprise and its role and signatures in behavioral and physiological measurements. The few examples that considered more than one surprise measure (Gijsen et al., 2021; Kolossa et al., 2015; Mars et al., 2008; Mousavi et al., 2020; Ostwald et al., 2012; Visalli et al., 2021) (we note that 3 out of 6 were published after 2020) have focused on model-selection methods to compare different models and did not look for fundamentally different predictions of these measures. However, as we will see in the next section, well-designed experiments allow us to dissociate different surprise measures based on qualitatively different predictions.

**Summary and intermediate discussion (i)**

In a unified framework, we discussed 9 previously proposed measures of surprise: (1) the Bayes Factor surprise; (2) the Shannon surprise; (3) the State Prediction Error; (4) the Absolute and (5) the Squared error surprise; (6) the Bayesian surprise; (7) the Postdictive surprise; (8) the Confidence Corrected surprise; and (9) the Minimized Free Energy. We considered different ways to define some of these measures in volatile environments and, overall, analyzed 16 different definitions of surprise.

Based on how different definitions depend on the belief \( \pi^{(t)} \), we divided them into three groups of probabilistic mismatch (\( S_{BF}, S_{Sh1}, S_{Sh2}, S_{SPE1}, \) and \( S_{SPE2} \)), observation-mismatch (\( S_{Ab1}, S_{Ab2}, S_{Sq1}, \) and...
Surprise measures (Fig. 3). We then showed how these measures relate to each other theoretically and, more importantly, under which conditions they are invertible functions of each other (i.e., they are indistinguishable – Fig. 2A); these links are summarized in Fig. 2B.

Importantly, our theoretical analysis provides foundations for biological mechanism of surprise computation and surprise-modulated learning (Berlemont & Nadal, 2021; Frémaux & Gerstner, 2016; Gerstner et al., 2018; Iigaya, 2016; Soltani & Izquierdo, 2019). For example, it has been argued that the computation of observation-mismatch surprise measures is biologically more plausible than more abstract measures such as Shannon surprise (Iigaya, 2016). Our results identify conditions under which observation-mismatch surprise measures behave identically as those probabilistic mismatch surprise measures that are optimal for adaptive learning (c.f. Fig. 2B, Proposition 1, and Corollary 1).

Measures of surprise in neuroscience have been previously divided into two categories (Faraji et al., 2018; Gijsen et al., 2021; Hurley et al., 2011): ‘puzzlement’ and ‘enlightenment’ surprise. Puzzlement surprise measures how puzzling a new observation is for an agent, whereas enlightenment surprise measures how much the new observation has enlightened the agent and changed its belief – a concept closely linked but not identical to the ‘Aha! moment’ (Dubey et al., 2021; Kounios & Beeman, 2009). The Bayesian and the Postdictive surprise can be categorized as enlightenment surprise since both quantify information gain (Fig. 4). Based on our theoretical analyses, however, we suggest to further divide measures of puzzlement surprise into 3 sub-categories (Fig. 4):

1. ‘Prediction surprise’: Surprise measures that quantify how unpredicted, unexpected, or unlikely the new observation is, including the Shannon surprise, State Prediction Error, the Minimized Free Energy, and all observation-mismatch surprise measures (Fig. 4). According to these measures, the agent’s puzzle is to find the most accurate predictions of the next observations.

2. ‘Change-point detection surprise’: Surprise measures designed to modulate the learning rate and to identify environmental changes, including the Bayes Factor surprise and the difference in Shannon surprise (c.f. Corollary 1; Fig. 4). According to these measures, the agent’s puzzle is to detect environmental changes.

3. ‘Confidence correction surprise’: Surprise measures that explicitly account for the agent’s confidence, including the Confidence Corrected surprise. The idea is that higher confidence (or higher commitment to a belief) leads to more puzzlement, where the puzzle is either to detect environmental changes or to find the most accurate prediction. We will introduce a new measure in this family in section Regularized Shannon surprise: A new direction that explicitly captures confidence and defines the agent’s puzzle as finding the most accurate predictions (Fig. 4).

Following this classification, we focus in the following section on one representative example from each of these sub-categories, and, whenever there are two definitions of one surprise measure, we take the second one: we include, in our analysis, the Shannon surprise $S_{Sh} = S_{Sh2}$, the Confidence Corrected surprise $S_{CC} = S_{CC2}$, the Bayesian surprise $S_{Ba} = S_{Ba2}$, and the Bayes Factor surprise $S_{BF}$ (colored in black in Fig. 4).

**Different definitions make different predictions: Case studies**

In the previous section, we clustered definitions of surprise into 4 main categories based on important conceptual differences (Fig. 4). In this section, we design experimental paradigms where these different definitions make different predictions. Our goal is (i) to better understand the behavior of each measure, (ii) to examine whether they are consistent with common sense, and (iii) to aid future experimental studies to dissociate the contribution of different measures of surprise to explaining biological and behavioral...
In our first case-study, we address the following question: ‘how does an agent’s confidence in its belief and predictions influence its perception of surprise?’ . To do so, we study two thought experiments and propose two actual experiments where different surprise measures make different predictions for the dependence of surprise upon confidence.

**General setting: Task formulation.** All our experiments can be formulated in the form of a categorical task: Each observation $y_t$ can be seen as a number between 1 and $N$ sampled from a categorical distribution, e.g., the observations are generated by rolling a die, flipping a coin, or playing a random musical note. We assume that there is no cue variable $x_t$ and no change in the environment ($p_c = 0$). In human experiments, these assumptions should be communicated to participants through instructions at the beginning of the experiment. Because there is no change and there are exactly $N$ possibilities for $y_t$, $\Theta = \Theta_1 = ... = \Theta_t$ is a vector of $N$ probabilities $[p_1, ..., p_N]$ corresponding to different values of $y_t$. Since there are no cues, the conditional distribution $P_{Y|X}$ is specified as

$$P_{Y|X}(y = i|x, \theta = [p_1, ..., p_N]) = P_Y(y = i|\theta = [p_1, ..., p_N]) = p_i. \quad (37)$$

An example of a categorical task is the oddball paradigm (Squires et al., 1976). In a typical oddball task, participants are exposed to a sequence of binary stimuli (i.e., $N = 2$), e.g., they listen to an auditory
sequence composed of two different musical notes, where one stimulus is less frequent than the other, e.g., \( p_1 = 0.9 \) and \( p_2 = 0.1 \). The frequent stimulus is often called ‘standard’, and the infrequent stimulus is often called ‘deviant’. Oddball tasks are among the most widely used paradigms for the study of the neural and behavioral signatures of surprise (Huettel et al., 2002; Maheu et al., 2019; Meyniel et al., 2016; Modirshanechi et al., 2019; Rubin et al., 2016; Squires et al., 1976). We will also consider generalized oddball task with \( N = 3 \) stimuli.

As a natural choice for a categorical task, we assume that the prior and the current beliefs are both Dirichlet distributions

\[
\pi^{(0)}(\theta) = \text{Dir}(\theta; 1_{N \times 1}) = \pi_{\text{flat}}(\theta) \quad \text{and} \quad \pi^{(t)}(\theta) = \text{Dir}(\theta; \alpha^{(t)}),
\]

where \( \alpha^{(t)} = [\alpha_1^{(t)}, \ldots, \alpha_N^{(t)}] \) are the parameters of the Dirichlet distribution for the belief at time \( t \), and the initial belief \( \pi^{(0)} \) is the uniform distribution, i.e., a Dirichlet distribution with parameters \( \alpha^{(0)} = 1_{N \times 1} \).

At any time-point \( t \geq 0 \), we can write

\[
\alpha^{(t)} = \alpha^{(t)}_{\text{sum}} \cdot \hat{\theta}^{(t)} = \alpha^{(t)}_{\text{sum}} \cdot [\hat{p}_1^{(t)}, \ldots, \hat{p}_N^{(t)}],
\]

where we define \( \alpha^{(t)}_{\text{sum}} = \sum_{i=1}^{N} \alpha_i^{(t)} \) and \( \hat{p}_i^{(t)} = \alpha_i^{(t)}/\alpha^{(t)}_{\text{sum}} \). The vector \( \hat{\theta}^{(t)} = [\hat{p}_1^{(t)}, \ldots, \hat{p}_N^{(t)}] = \mathbb{E}_{\pi^{(t)}}[\Theta] \) is an estimate of \( \theta \) given the belief \( \pi^{(t)} \). We discuss the parameter \( \alpha^{(t)}_{\text{sum}} \) below when we define confidence. Using this notation, we can write the marginal probability as \( P(y = i; \pi^{(t)}) = \hat{p}_i^{(t)} \). For example, with the belief \( \pi^{(t)} \) at time \( t \), \( \hat{p}_2^{(t)} \) is the estimated occurrence probability of the second category (e.g., the deviant note in an auditory oddball task). For the initial belief \( \pi^{(0)} \), we have \( \alpha^{(0)}_{\text{sum}} = N \) and \( P(y = i; \pi^{(0)}) = \hat{p}_i^{(0)} = 1/N \) for all \( i \) between 1 and \( N \).

Because \( P(y; \pi^{(0)}) \) is a flat distribution, \( S_{\text{BF}} \) and \( S_{\text{Sh}} \) are invertible functions of each other and hence indistinguishable (Fig. 2). Therefore, in this section, we do not explicitly include \( S_{\text{BF}} \) in our comparisons as all qualitative results for \( S_{\text{Sh}} \) hold true also for \( S_{\text{BF}} \); in particular, if \( S_{\text{Sh}} \) increases or decreases, \( S_{\text{BF}} \) does as well. Moreover, in the setting described above, we have \( S_{\text{Sh1}} = S_{\text{Sh2}}, S_{\text{Ba1}} = S_{\text{Ba2}}, \) and \( S_{\text{CC1}} = S_{\text{CC2}} \) due to the assumptions \( p_c = 0 \) and \( \pi^{(0)} = \pi_{\text{flat}} \) (Fig. 2B).

**General setting: Confidence definition.** To study the effect of confidence on the perception of surprise, we first need to agree how to measure confidence given a belief \( \pi^{(t)} \). In Eq. 29, we have defined the confidence given a belief \( \pi^{(t)} \) by the negative entropy \( C[\pi^{(t)}] \) (Faraaji et al., 2018). However, when \( \pi^{(t)} \) is a Dirichlet distribution, there is no analytic expression for \( C[\pi^{(t)}] \). For reasons of practicality, we therefore define here confidence in a categorical task as the inverse variance

\[
\text{CatConf}(t) = \frac{1}{\mathbb{E}_{\pi^{(t)}}[||\Theta - \hat{\theta}^{(t)}||_2^2]} = \frac{\alpha^{(t)}_{\text{sum}} + 1}{1 - ||\hat{\theta}^{(t)}||_2^2},
\]

where CatConf stands for Categorical Confidence, \( \hat{\theta}^{(t)} = \mathbb{E}_{\pi^{(t)}}[\Theta] \) is the current estimate of the parameter, \( ||.||_2 \) stands for the \( \ell_2 \)-norm, and \( \mathbb{E}_{\pi^{(t)}}[||\Theta - \hat{\theta}^{(t)}||_2^2] = \int ||\theta - \hat{\theta}^{(t)}||_2^2 \pi^{(t)}(\theta) d\theta \) is the variance of \( \Theta \) given the belief \( \pi^{(t)} \). According to Eq. 40, for a fixed estimate \( \hat{\theta}^{(t)} \), confidence is an increasing function of \( \alpha^{(t)}_{\text{sum}} \) (c.f. Fig. 5A), and, for a fixed \( \alpha^{(t)}_{\text{sum}} \), confidence is an increasing function of \( ||\hat{\theta}^{(t)}||_2^2 \) (c.f. Fig. 5B).

\footnote{If the initial belief \( \pi^{(0)} \) is a Dirichlet distribution, then \( \pi^{(t)} \) is also a Dirichlet distribution at any time \( t \) for a wide range of learning rules (Faraaji et al., 2018; Liakoni et al., 2021; Maheu et al., 2019; Markovic et al., 2021; Meyniel et al., 2016; Modirshanechi et al., 2019; Ryali et al., 2018; Yu & Cohen, 2009) – since the Dirichlet distribution is the conjugate prior of the categorical distribution (Efron & Hastie, 2016).}

\footnote{We note that the qualitative behavior of confidence in Fig. 5 and Fig. 6 is the same for both definitions \( C[\pi^{(t)}] \) and CatConf(\( t \)).}
The parameter $\alpha^{(t)}_{\text{sum}}$ has been interpreted as the number of samples the belief $\pi^{(t)}$ is worth (Efron & Hastie, 2016). We note that $||\hat{\theta}^{(t)}||_2^2 = 1/N + \sum_{i=1}^N (\hat{\theta}_i^{(t)} - 1/N)^2$ is a measure of the difference between the estimate $\hat{\theta}^{(t)}$ and the uniform distribution over $N$ categories – i.e., $||\hat{\theta}^{(t)}||_2^2$ takes its maximum value (corresponding to maximum confidence) when the estimate $\hat{\theta}^{(t)}$ has a probability of 1 for one category and zero for the rest, and it takes its minimum value (corresponding to minimum confidence) when $\hat{\theta}^{(t)}$ is distributed uniformly over all categories.

As we show below, there is no general rule for the dependence of different surprise measures on confidence as defined in Eq. 40. Rather, depending on different values of $\alpha^{(t)}_{\text{sum}}$ and $\hat{\theta}^{(t)}$, confidence affects surprise measures differently.

**Thought experiment: (i) Coin flipping.** Imagine we flip a coin. How surprising is it to see a head? How does our surprise change depending on how long we have owned the coin? Or depending on how much we believe in its fairness?

As our first thought experiment, we consider a coin flipping experiment where the observation $Y_{t+1}$ is either a head ($Y_{t+1} = 1$) or a tail ($Y_{t+1} = 2$) – i.e., $N = 2$. Our goal is to study the effect of confidence as defined in Eq. 40 on surprise of observing a head ($Y_{t+1} = 1$) depending on how often we have flipped the coin (i.e., how large $\alpha^{(t)}_{\text{sum}}$ is) and depending on our estimate of its bias (i.e., how large $|\hat{p}_1^{(t)} - 0.5|$ is).

We first fix $\hat{p}_1^{(t)}$ and study the behavior of different surprise measures as a function of the parameter $\alpha^{(t)}_{\text{sum}}$.

In other words, we show that we have a fixed estimate of the coin’s bias (expressed by $\hat{p}_1^{(t)}$), but our confidence about our estimation (measured by how often we have flipped it, $\alpha^{(t)}_{\text{sum}}$) varies (grey, Fig. 5A1). We observe that the Shannon surprise $S_{\text{Sh}}$ of $Y_{t+1} = 1$ is independent of $\alpha^{(t)}_{\text{sum}}$ (blue, Fig. 5A1), as it is expected from a probabilistic mismatch surprise measure (Fig. 3). The Bayesian surprise $S_{\text{Ba}}$, however, decreases with increasing $\alpha^{(t)}_{\text{sum}}$ (red, Fig. 5A1). The reason is that the more confident we are about our belief, the less a single new observation changes our belief. In Appendix C: Methods for case-studies, we show that these results are not limited to the coin flipping experiment ($N = 2$) but are true for any categorical task with arbitrary $N$. The Confidence Corrected surprise $S_{\text{CC}}$, on the other hand, does not change monotonically with respect to $\alpha^{(t)}_{\text{sum}}$ (green, Fig. 5A1): It is a decreasing function of confidence for small values of $\alpha^{(t)}_{\text{sum}}$ but an increasing function of confidence for large values of $\alpha^{(t)}_{\text{sum}}$.

Next we fix the parameter $\alpha^{(t)}_{\text{sum}}$ (i.e., how often we have flipped the coin) and study the behavior of different surprise measures as functions of our estimate of the coin’s bias $|\hat{p}_1^{(t)} - 0.5|$. For $N = 2$, the term $||\hat{\theta}^{(t)}||_2^2$ in Eq. 40 is equal to $0.5 + 2(\hat{p}_1^{(t)} - 0.5)^2$, thus confidence is an increasing function of the estimated bias $|\hat{p}_1^{(t)} - 0.5|$ (grey, Fig. 5B1). The effect of confidence on surprise of observing $Y_{t+1} = 1$ depends on whether the agent’s prediction has been wrong ($\hat{p}_1^{(t)} < 0.5$, i.e., $Y_{t+1} = 1$ was observed even though it was thought to be less likely than $Y_{t+1} = 2$) or correct ($\hat{p}_1^{(t)} > 0.5$); c.f. Fig. 5B1. For a wrong prediction, surprise increases with increasing confidence, i.e., with decreasing $\hat{p}_1^{(t)}$ when $\hat{p}_1^{(t)} < 0.5$, for all 3 surprise measures (Fig. 5B1, left). This means that all surprise measures predict that a wrong prediction made with higher confidence leads to higher surprise than a wrong prediction made with little confidence. For a correct prediction, higher confidence decreases Shannon $S_{\text{Sh}}$ and Bayesian $S_{\text{Ba}}$ surprise (blue and red, respectively, Fig. 5B1, right). In Appendix C: Methods for case-studies, we show that these results are not limited to the coin flipping experiment and are true for any categorical task. Interestingly, the Confidence Corrected surprise $S_{\text{CC}}$ does not change monotonically with respect to the confidence for a correct prediction (green, Fig. 5B1, right). For $\hat{p}_1^{(t)}$ slightly greater than 0.5, it decreases with an increase in confidence, but then, for larger values of $\hat{p}_1^{(t)}$, the slope turns and $S_{\text{CC}}$ increases with an increase in confidence. In other words, $S_{\text{CC}}$ predicts that the more we expect to observe a head ($Y_{t+1} = 1$), the more surprised we are upon observing it. This prediction looks counter-intuitive.
Figure 5: Coin flipping. **A.** Effect of $\alpha_{\text{sum}}^{(t)}$ on surprise when $\hat{p}_{1}^{(t)} = \hat{p}_{2}^{(t)} = 0.5$ is fixed. **A1.** The Shannon (Sh, blue), the Bayesian (Ba, red), and the Confidence Corrected (CC, green) surprise for $Y_{t+1} = 1$ as functions of $\alpha_{\text{sum}}^{(t)}$. The dashed grey line shows the confidence as defined in Eq. 40 (on an arbitrary scale). **A2.** Three instances of the belief $\pi^{(t)}$ with same expected value, corresponding to different values of $\alpha_{\text{sum}}^{(t)}$ (indicated by arrows). **A3.** The estimate of the parameters $\hat{\theta}^{(t)} = [\hat{p}_{1}^{(t)}, \hat{p}_{2}^{(t)}]$ corresponding to the beliefs in panel A2. **B.** Effect of $\hat{p}_{1}^{(t)}$ on surprise when $\alpha_{\text{sum}}^{(t)} = 10$ is fixed. **B1.** The Shannon (Sh, blue), the Bayesian (Ba, red), and the Confidence Corrected (CC, green) surprise for $Y_{t+1} = 1$ as functions of $\hat{p}_{1}^{(t)}$. The dashed grey curve shows the confidence. We consider $\hat{p}_{1}^{(t)} < 0.5$ as an indication of **wrong** prediction of $Y_{t+1}$ and $\hat{p}_{1}^{(t)} > 0.5$ as an indication of **correct** prediction of $Y_{t+1}$. **B2.** Three instances of the belief $\pi^{(t)}$ corresponding to different values of $\hat{p}_{1}^{(t)}$ with same $\alpha_{\text{sum}}^{(t)}$. **B3.** The estimate of the parameters $\hat{\theta}^{(t)} = [\hat{p}_{1}^{(t)}, \hat{p}_{2}^{(t)}]$ corresponding to the beliefs in panel B2. We note that the qualitative behavior of different measures with respect to $\alpha_{\text{sum}}^{(t)}$ remains the same for other values of $\hat{p}_{1}^{(t)}$, their qualitative behavior with respect to $\hat{p}_{1}^{(t)}$ and remains the same for other values of $\alpha_{\text{sum}}^{(t)}$. To summarize, the three paradigmatic measures of surprise show qualitative differences depending on $\alpha_{\text{sum}}^{(t)}$ given a fixed estimate of the parameter $\hat{\theta}^{(t)}$: The Shannon surprise $S_{\text{Sh}}$ stays constant, the Bayesian surprise $S_{\text{Ba}}$ decreases, and the Confidence Corrected surprise $S_{\text{CC}}$ first decreases and then increases. Similarly, they show qualitative differences depending on the estimate of the parameter $\hat{p}_{1}^{(t)}$ given a fixed $\alpha_{\text{sum}}^{(t)}$. 

---

713 To summarize, the three paradigmatic measures of surprise show qualitative differences depending on $\alpha_{\text{sum}}^{(t)}$ given a fixed estimate of the parameter $\hat{\theta}^{(t)}$: The Shannon surprise $S_{\text{Sh}}$ stays constant, the Bayesian surprise $S_{\text{Ba}}$ decreases, and the Confidence Corrected surprise $S_{\text{CC}}$ first decreases and then increases. Similarly, they show qualitative differences depending on the estimate of the parameter $\hat{p}_{1}^{(t)}$ given a fixed $\alpha_{\text{sum}}^{(t)}$.
Thought experiment: (ii) CEO selection. Faraji et al., 2018 discussed a thought experiment where a few employees are waiting for the outcome of the next CEO selection. The idea is to analyze how surprise upon observing the outcome changes as a function of employees’ confidence in their beliefs. We model the example of Faraji et al., 2018 as a categorical task with a Dirichlet belief.

We assume there are $N = 3$ candidates for the next CEO and model an employee’s expectation about the outcome as a Dirichlet belief parameterized by $\alpha_{\text{sum}}(t)$ and $\hat{\theta}(t) = [\hat{p}_1(t), \hat{p}_2(t), \hat{p}_3(t)]$. All employees have spent the same number of years in the company so that we can assume that $\alpha_{\text{sum}}(t)$ has the same value for all of them. We analyze the dependency of the surprise felt when the 1st candidate is selected (i.e., $Y_{t+1} = 1$) as a function of $\hat{\theta}(t)$. In particular, we follow Faraji et al., 2018 and assume that $\hat{p}_1(t) = 0.3$ is fixed among all employees, i.e., they all estimate a probability of 0.3 for the selection of the 1st candidate, but their estimated selection probabilities for the other candidates vary. Faraji et al., 2018 argued that an employee who puts a probability of $\hat{p}_2(t) = 0.7$ to the selection of the 2nd candidate ($p_2^t = 0.35$) is more surprised when the 1st candidate is selected than another employee who considers the chances for candidates 2 and 3 to be identical ($\hat{p}_2(t) = \hat{p}_3(t) = 0.35$) and close to that of the 1st candidate ($\hat{p}_1(t) = 0.3$). We therefore look at the surprise of the 1st candidate being selected (i.e., $Y_{t+1} = 1$) as a function of the employees’ expectation $\hat{p}_2(t)$ for the 2nd candidate; note that $\hat{p}_3(t) = 1 - \hat{p}_1(t) - \hat{p}_2(t)$.

The Shannon surprise $S_{\text{Sh}}$ and the Bayesian surprise $S_{\text{Ba}}$ are independent of $\hat{p}_2(t)$, which implies that the surprise of observing $Y_{t+1} = 1$ is independent of how estimations $p_2^t$ and $p_3^t$ are distributed across the other candidates (blue and red, Fig. 6A1) – in Appendix C: Methods for case-studies, we show that this result is not limited to the CEO selection experiment but is true for any categorical task. The Confidence Corrected surprise $S_{\text{CC}}$, however, has an interesting U-shape relation with $\hat{p}_2(t)$ (green, Fig. 6A1) – similar to the dependence of confidence on $\hat{p}_2(t)$ (grey, Fig. 6A1). The minimum surprise corresponds to $\hat{p}_2(t) = \hat{p}_3(t) = 0.35$ where confidence goes through its minimum and an employee’s estimate of the outcome probability is as uninformative as possible (Fig. 6A). This means that, given the same expectation for the selection of the 1st candidate, the employees who have a high expectation for the selection of the 2nd or the 3rd candidate are going to be more surprised when hearing that the 1st candidate was selected than the employees who consider all three candidates about equally likely to be selected. This behavior looks plausible and is exactly what $S_{\text{CC}}$ was designed to capture (Faraji et al., 2018). $S_{\text{CC}}$, therefore, explains aspects of surprise perception that are not captured by $S_{\text{Sh}}$ and $S_{\text{Ba}}$.

To summarize, $S_{\text{CC}}$ can capture aspects of surprise perception that are consistent with our intuition and are not captured by $S_{\text{Sh}}$ and $S_{\text{Ba}}$. However, $S_{\text{CC}}$ has also some non-intuitive behavior that appears to contradict common sense – as observed in the coin flipping experiment. Our results suggest that a new definition of surprise may be needed for capturing the intuitively correct behavior in both thought experiments (Fig. 5 and Fig. 6). We address this problem in section Regularized Shannon surprise: A new direction.

Proposed experiment: (i) Classic oddball task ($N = 2$). The different predictions of different surprise measures with respect to changes in $\alpha_{\text{sum}}(t)$ (for a fixed $\hat{p}_1(t)$, Fig. 5A) and $\hat{p}_1(t)$ (for a fixed $\alpha_{\text{sum}}(t)$, Fig. 5B) in the coin flipping experiment can be exploited in a classic oddball task ($N = 2$) that aims at determining which measure of surprise correlates best with a behavioral or physiological measurement $Z$ (Fig. 2A1). In particular, our analyses (Fig. 5) suggest that different surprise measures have a different behavior at the beginning (early phase, Fig. 7A) and at the end (late phase, Fig. 7A) of an oddball task. First, as time passes, participants receive more samples to make their estimations, i.e., $\alpha_{\text{sum}}$ moves towards the right side of Fig. 5A1. Second, with time they learn that the standard stimuli are more likely to happen, i.e., $\hat{p}_1(t)$ moves towards the right side of Fig. 5B1. Therefore, according to the Shannon and the Bayesian surprise, we expect the standard stimuli to be more surprising during the early phase of
Figure 6: CEO selection. A. The Shannon (Sh, blue), the Bayesian (Ba, red), and the Confidence Corrected (CC, green) surprise for $Y_{t+1} = 1$ as functions of $\hat{p}_2(t)$ (with $\alpha_{\text{sum}}(t) = 10$ and $\hat{p}_1(t) = 0.3$ fixed). The dashed grey curves show the confidence as defined in Eq. 40 (on an arbitrary scale). B. Three instances of the belief $\pi(t)$ corresponding to different values of $\hat{p}_2(t)$ in panel A. Note that the space is shown as a triangle because $\theta = [p_1, p_2, p_3]$ lives inside the area specified by $p_1 + p_2 + p_3 = 1$, $p_1 \geq 0$, $p_2 \geq 0$, and $p_3 \geq 0$. C. The estimate of the parameters $\hat{\theta}(t) = [\hat{p}_1(t), \hat{p}_2(t), \hat{p}_3(t)]$ corresponding to the beliefs in panels A. We note that the qualitative behavior of different measures with respect to $\hat{p}_2(t)$ remains the same for other choices of $\alpha_{\text{sum}}(t)$ and $\hat{p}_1(t)$.

The experiment than the late phase (Fig. 7B). The opposite holds for the Confidence Corrected surprise (Fig. 7B). Moreover, the Shannon and the Confidence Corrected surprise predict, as a function of time, an increase in the surprise upon observing a deviant stimulus, whereas the Bayesian surprise predicts the opposite (Fig. 7C). Therefore, we find a double dissociation. Thus, depending on how the measurement $Z$ behaves in the two cases (Fig. 7B and Fig. 7C), we can uniquely determine with which of the three surprise measures it correlates. As a proof of concept, we analyzed a publicly available visual oddball dataset (Robbins et al., 2018) (18 participants) with the EEG recordings at central electrodes, i.e., Cz, C1, and C2. We find that the EEG amplitude at around 450ms after the stimulus onset has a behavior consistent with the predictions of $S_{\text{Sh}}$ but not with those of $S_{\text{Ba}}$ or $S_{\text{CC}}$. See Appendix D: EEG analysis and Fig. 7D-F for details.

Since our proposed approach relies on qualitative differences (such as increase or decrease) of surprise measures, it goes beyond general methods of naive model selection (Gijsen et al., 2021; Kolossa et al.,
Proposed experiment: (ii) Generalized oddball task \((N = 3)\). The U-shape relation between the Confidence Corrected surprise and \(\hat{p}^{(t)}\) in the CEO selection experiment (Fig. 6) can be tested in a generalized oddball task with \(N = 3\) stimuli (similar to Mars et al., 2008). The approach is similar to that of the binary oddball task in Fig. 7: We have a physiological measurement \(Z\) that is thought to be sensitive to surprise; we want to test whether its behavior is consistent with the predictions of \(S_{\text{CC}}\) or with those of \(S_{\text{Sh}}\) and \(S_{\text{Ba}}\).

To do so, we design a sequence of stimuli consisting of two phases separated by an abrupt change (Fig. 8A). The idea is to keep the occurrence frequency of the 1st stimulus \(p_1\) fixed throughout the whole sequence and change the balance between \(p_2\) and \(p_3\) from Phase 1 to Phase 2 (Fig. 8B). In particular, we consider the case that \(p_2 = p_3\) in Phase 1 (before the change) while \(p_2 \ll p_3\) in Phase 2 (after the change). In this case, \(S_{\text{Sh}}\) and \(S_{\text{Ba}}\) predict that the surprise of observing \(Y_t = 1\) is the same in Phase 1 as in Phase 2, whereas \(S_{\text{CC}}\) predicts that observing \(Y_t = 1\) is more surprising in Phase 2 than in Phase 1 (Fig. 8C). Such
Different measures of surprise have different predictions for the early and the late phase of oddball tasks. A. The Shannon (Sh, blue), the Bayesian (Ba, red), and the Confidence Corrected (CC, green) surprise over time for a sequence of 100 binary stimuli (0.1 probability for the deviant stimulus); see Appendix C: Methods for case-studies for simulation details. The surprise sequences were linearly scaled between 0 and 1 before plotting. Each star shows one observation \(Y_t\); the standard (\(Y_t = 1\)) is plotted at \(-0.05\), and the deviant (\(Y_t = 2\)) is plotted at \(1.05\). The early phase consists of the first 25 stimuli and the late phase of the last 25 stimuli. B. Average (over 40 random seeds) surprise values of standard stimuli in the early (first 25 out of 100 trials) and the late (last 25 out of 100 trials) phase of the task. C. Average (over 40 random seeds) surprise values of deviant stimuli in the early and the late phase of the task. D. Event related potential (ERP) averaged over central electrodes (Cz, C1, and C2, averaged over 18 participants after normalization) – raw dataset from Robbins et al., 2018 (see Appendix D: EEG analysis). The red curve corresponds to the deviant ERP and the blue curve to the standard ERP. \(Z\) and \(Z'\) highlight two different physiological candidates for surprise values. Shaded areas show the standard error of the mean (over 18 participants). E. Average (over 18 participants) \(Z\) and \(Z'\) for the standard stimuli in the early (first 75 out of \(\sim 270\) trials) and the late (last 75 out of \(\sim 270\) trials) phase of the task in Robbins et al., 2018. F. Average (over 18 participants) \(Z\) and \(Z'\) for the deviant stimuli in the early and the late phase of the task in Robbins et al., 2018. In all panels, error bars show the standard error of the mean. Overall, \(Z\) matches the prediction of the Shannon surprise, and \(Z'\) can be sensitive to both the Shannon and the Confidence Corrected surprise but not the Bayesian surprise; see Appendix D: EEG analysis for details.

Interestingly, SanMiguel et al., 2021 have recently conducted a very similar experiment, parallel to our study and unknown to us. In a generalized oddball task with \(N = 11\) stimuli, they observed that increasing \(p_2\) from \(1/11\) to \(10/11\) (while keeping \(p_1 = 1/11\) fixed) increases an indicator of surprise in EEG (the Mismatch Negativity amplitude; c.f. Näätänen et al., 2007) in response to \(Y_t = 1\). This observation matches the predictions of \(S_{CC}\) but not those of \(S_{Sh}\) and \(S_{Ba}\).

Case-study 2: Learning in volatile environments

In our second case-study, we investigate the link between surprise and learning. According to previous theoretical (Faraji et al., 2018; Findling et al., 2021; Frémaux & Gerstner, 2016; Gerstner et al., 2018; Liakoni et al., 2021; Yu & Dayan, 2005) and experimental (Behrens et al., 2007; Findling et al., 2021; Heilbron & Meyniel, 2019; Nassar et al., 2012; Nassar et al., 2010; Soltani & Izquierdo, 2019; Xu et al., 2021) studies, surprising events modulate the speed of learning in the brain. The argument is that surprising events are caused by inadequate expectations. Hence, surprise indicates the need for updating the source of those expectations (i.e., the belief).

In volatile environments similar to our generative model (Fig. 1A), an unexpected event can occur either due to an abrupt change in the parameters of the environment (also called unexpected uncertainty (Soltani & Izquierdo, 2019; Yu & Dayan, 2005)) or due to the stochasticity of the environment (i.e., pure randomness, also called expected uncertainty (Soltani & Izquierdo, 2019; Yu & Dayan, 2005)). While in the former case forgetting the old observations and increasing the speed of learning are necessary to explain the rapid adaptive behavior observed in humans (Behrens et al., 2007; Nassar et al., 2012; Nassar et al., 2010; Xu et al., 2021), in the latter case, forgetting the earlier observations leads to unnecessary loss of information. Therefore, not all unexpected events should change the speed of learning (Liakoni et al., 2021; Soltani & Izquierdo, 2019; Yu & Dayan, 2005).

Here, we compare different surprise measures based on how informative they are about environmental
Figure 8: Testing the prediction of the CEO selection experiment in a generalized oddball task with $N = 3$ categorical stimuli. A. The Shannon (Sh, blue), the Bayesian (Ba, red), and the Confidence Corrected (CC, green) surprise over time for a sequence of 300 stimuli; see Appendix C: Methods for case-studies for simulation details. Each star shows one observation $Y_t$ (right vertical axis), and the surprise sequences were linearly scaled between 0 and 1 before plotting. There is an abrupt change in the true distribution of the observations at $t = 151$ dividing the sequence into two phases, i.e., ‘Phase 1’ (1 ≤ $t$ ≤ 150) and ‘Phase 2’ (151 ≤ $t$ ≤ 300). The probability of $Y_t = 1$ remains the same through the whole sequence ($P_1 = 0.3$), but $Y_t = 2$ becomes significantly less likely after the change-point (see panel B). To avoid complications concerning how participants update their belief, we focus on the 2nd half of each phase (indicated by the grey line and called ‘Averaging period’) where they have a relatively accurate estimate of probabilities. B. The true underlying distribution in Phase 1 (left) and Phase 2 (right). C. Average (over 40 random seeds) surprise values of $Y_t = 1$ in Phase 1 (trials 76 to 150) and Phase 2 (trials 226 to 300).

Our goal is to determine which measures are most useful to modulate the speed of learning. To do so, we study a thought experiment and propose an actual experiment where different surprise measures make different predictions on how an abrupt change in the environment influences surprise of upcoming observations.

**General setting: Task formulation.** Both our experiments can be formulated in the same paradigm with a volatile environment ($p_c > 0$) and binary observations ($Y_t \in \{1, 2\}$). At time $t$, an oracle announces the cue $x_t \in [0, 1]$ as a prediction of the probability of the event $Y_t = 1$ (e.g., a cursor is located on an axis between 0 and 1 on a computer screen). A hidden parameter indicates whether the probability predicted by the oracle is correct ($\Theta_t = 1$) or whether the oracle’s prediction is uninformative ($\Theta_t = 0$) and $Y_t$...
comes from a uniform distribution. Formally, we write
\[ P_{Y \mid X}(y = 1 \mid x; \theta = 1) = x \quad \text{and} \quad P_{Y \mid X}(y = 1 \mid x; \theta = 0) = \frac{1}{2}. \] (41)

The belief \( \pi(t)(\theta = 1) = b(t) \in [0, 1] \) shows how much an agent has trust in the oracle, whereas \( \pi(t)(\theta = 0) = 1 - b(t) \) shows how much it believes in the unpredictability of the outcome.

**Thought experiment: Belief in Forecasts.** We started the introduction of the paper by discussing an example of a wrong, and potentially surprising, weather forecast. A farmer who believed in the weather forecast and planned to outdoor work on a sunny morning will be surprised (and possibly angry) when he finds bad weather in the morning. Another farmer who thinks that weather forecasts are uninformative anyway will be less surprised than the first farmer.

Similarly, some people believe in opinion polls or predictions of election outcomes in the media, while others do not. Imagine two citizens A and B who have been living for many years in a country where each year an important election takes place between a blue party \( (y_{t+1} = 1) \) and a red party \( (y_{t+1} = 2) \).

Each year the media announce, a week before the election, the probability \( x_{t+1} \) for the blue party to win \( (Y_{t+1} = 1) \). The media base their predictions on all available cues, including opinion polls, extrapolations from previous years, and elaborate mathematical models; however, nobody knows how reliable these predictions really are. Neither citizen A nor B has other independent cues. The citizens’ trust that the media are correct \( (\Theta_t = 1) \) is represented by \( b(t) \) which indicates how much they believe that the election is predictable and in agreement with the forecast of media. A trust of \( b(t) = 0 \) implies that a citizen believes that the outcome is unpredictable and both parties are equally likely to win. Hence, given the forecast \( x_{t+1} \) by the media, a citizen with the belief \( b(t) \) estimates the probability of the blue party to win as
\[ P(y_{t+1} = 1 \mid x_{t+1}; \pi(t)) = b(t)x_{t+1} + (1 - b(t))\frac{1}{2}. \] (42)

Importantly, the belief \( b(t) \) of a citizen is not fixed but evolves with experience. It could be that for the first 20 years, the media did not know how to interpret data, relied on biased opinion polls, or used a bad mathematical model. Unknown to the citizens, starting in year 21, they may have a change in personnel and thereafter use a near-perfect mathematical model. Our claim, based on the results of earlier studies (Behrens et al., 2007; Faraji et al., 2018; Findling et al., 2021; Frémaux & Gerstner, 2016; Gerstner et al., 2018; Heilbron & Meyniel, 2019; Liakoni et al., 2021; Nassar et al., 2012; Nassar et al., 2010; Soltani & Izquierdo, 2019; Xu et al., 2021; Yu & Dayan, 2005), is that surprise such as ‘oh, this year the prediction was correct!’ is used by citizens to change their belief \( b(t) \) in the trustworthiness of the forecast. Our goal is to find which surprise measure is most useful to indicate the need for such a change.

To do so, let us suppose that citizen A has a high trust \( (e.g., b(t) = 0.9) \) in the forecast and citizen B a low trust \( (e.g., b(t) = 0.05) \). We assume that the media predicted a 90-percent probability of the blue party to win \( (x_{t+1} = 0.9) \) and consider the amount of surprise of citizen A and B at the moment when the results of the election are announced. Under the hypothesis that surprise about the election outcome is used to change the belief, we expect that the following should occur:

- **Expectation 1:** Citizen A who has a strong trust in the media \( (e.g., b(t) = 0.9) \) is more surprised if the actual outcome of the election proves the media wrong \( (\text{red party wins}) \) than if the election’s outcome \( (\text{blue party wins}) \) is consistent with the prediction of the media.

- **Expectation 2:** Citizen B who believes in unpredictability of the election \( (e.g., b(t) = 0.05) \) is more surprised if the actual outcome of the election proves the media right \( (\text{blue party wins as predicted}) \) than if the prediction is wrong.

- **Expectation 3:** If the outcome of the election is against the prediction of the media, then citizen A is more surprised than citizen B.
Note that, in the first two points, we compare the surprise value of two different outcomes for the same belief, whereas, in the last point, we compare the surprise value as a function of the belief $b(t)$.

For the case of citizen A (strong trust in the media) all definitions of surprise match our Expectation 1 (compare the blue and the red curves in Fig. 9 for high values of $b(t)$). If we follow the red curve from high belief to very low belief (i.e., if we decrease $b(t)$), then different measures of surprise show different behaviors. As $b(t)$ decreases, only the Bayes Factor surprise $S_{BF}$ (Fig. 9A1) and the Shannon surprise $S_{Sh}$ (Fig. 9A2) decrease and match our Expectation 3. Importantly, the red and blue curves cross for $S_{BF}$ but not for $S_{Sh}$. Hence, after a certain point ($b(t) < b(0)$), $S_{BF}$ predicts that the media being right is more surprising than the media being wrong – matching our Expectation 2. However, $S_{Sh}$ has a different behavior: as $b(t)$ decreases, the surprise of the media being wrong gets closer to the surprise of the media being right, i.e., the more citizens believe in randomness, the smaller the difference in how surprised they are by either event. In other words, $S_{Sh}$ indicates that when citizens believe in the unpredictability of the outcome, they do not care about the prediction of the media, and as a result, the prediction of the media does not affect their perception of surprise. Therefore, $S_{BF}$ matches all our three Expectations, but $S_{Sh}$ matches only Expectations 1 and 3.

Both $S_{Ba}$ (Fig. 9A3) and $S_{CC}$ (Fig. 9A4) are qualitatively distinguishable from $S_{BF}$ and $S_{Sh}$, since neither $S_{Ba}$ nor $S_{CC}$ has a monotonic relation with the degree of trust in media $b(t)$. Therefore, neither $S_{Ba}$ nor $S_{CC}$ matches our Expectation 3.

To summarize, different surprise measures have qualitatively different predictions for the surprise after an election outcome as a function of the trust $b(t)$ in the media. Importantly, only $S_{BF}$ matches our expectations for a surprise measure that is useful for modulation of the speed of learning (c.f. Proposition 1). The results remain the same if the prior belief $b(0)$ has a different value (Fig. 9B).

**Proposed experiment: Association learning task.** In this section, we propose an actual experiment to test predictions of the different surprise measures analogously to our thought experiment for belief in forecasts. Our proposed experiment is similar to the ones used in studies of association learning (Gershman et al., 2017; Niv et al., 2015): Participants should infer whether the predictions of an oracle $x_{1:t}$ (e.g., cues or conditioned stimuli) are or are not associated with the observations $y_{1:t}$. An alternative would be the design of a go no-go paradigm (Mars et al., 2008; Walz et al., 2015) with a cue variable $x_1$ preceding each stimulus. In both types of experiment, we allow for associations to change abruptly from time to time.

In order to formalize experimental predictions, we consider a task with 150 trials (Fig. 10). At each time $t$, the oracle announces either $X_t = 0.9$ or $X_t = 0.1$, randomly chosen. For the first 50 trials, we assume that the observations are independent of the oracle’s prediction, i.e., $\Theta_t = 0$ for $1 \leq t \leq 50$ and $Y_t \sim \text{Bernoulli}(0.5)$. Then, unknown to the participants, there is an abrupt change at trial 51, and, for the next 50 trials, the observations follow the same distribution as the one predicted by the oracle, i.e., $\Theta_t = 1$ for $51 \leq t \leq 100$ and $Y_t \sim \text{Bernoulli}(X_t)$. Finally after another abrupt change at trial 101, the observations become again independent of the oracle’s predictions, i.e., $\Theta_t = 0$ for $101 \leq t \leq 150$ and $Y_t \sim \text{Bernoulli}(0.5)$. We study the behavior of different surprise measures for both cases of transition from an unpredictable environment to a predictable one (at $t = 51$) and vice versa (at $t = 101$). To model the temporal dynamics of belief in such a volatile environment, we use exact Bayesian inference as in Proposition 1 with change-point probability $p_c = 0.03$ – see Appendix C: Methods for case-studies for details.

Before the first change-point at trial 51, participants have been in an unpredictable environment for a long time and have no trust in the oracle (and rather a strong belief in the unpredictability of the environment), i.e., $b(t) \approx 0$ – a situation similar to the extreme left in the plots of Fig. 9. The opposite is true before the second change-point at trial 101, where participants have been in a predictable environment for a long time.
Figure 9: **Surprise upon election of red or blue party as a function of trust $b(t)$ in the media.**

From left to right: Bayes Factor, Shannon, Bayesian, and Confidence Corrected surprise as functions of $b(t)$, for different prior belief: **A.** $b(0) = 0.2$ and **B.** $b(0) = 0.5$. The prediction of the media for the blue party to win is the same for all panels and equal to $X_t = 0.9$. The red curves correspond to the red party winning the election (i.e., $Y_t = 2$ and the media being wrong), and the blue curves correspond to the blue party winning the election (i.e., $Y_t = 1$ and the media being right). The change-point probability $p_c$ is assumed to be 0.03. However, $S_{BF}$, $S_{Sh}$, and $S_{CC}$ are independent of $p_c$. The qualitative behavior of $S_{Ba}$ remains the same for other values of $p_c$ up to 0.2 (i.e., when there is on average one change-point per 5 time-points), but it becomes similar to the one of $S_{CC}$ for the more volatile environments (i.e., for higher values of $p_c$; c.f. Corollary 3). For all values of $p_c$, the Bayesian surprise $S_{Ba1}$ (equal to $S_{Ba2}$ for $p_c = 0$; c.f. Eq. 23 and Eq. 24) has the same qualitative behavior as $S_{Ba2}$ shown in panel A3 and B3.

After the first change-point at trial 51, there is (on average) a jump in $S_{BF}$ (Fig. 10B first row), signaling a change in the environment. After the change, $S_{BF}$ decreases until the next change at trial 101 where it jumps again, followed by another decreasing period. Thus, $S_{BF}$ efficiently and explicitly indicates environmental changes. This observation is consistent with the statement of Proposition 1. $S_{CC}$ (Fig. 10B last row) has a similar behavior; a minor difference compared to $S_{BF}$ occurs after the first change-point where, after a jump, $S_{CC}$ decreases quickly before it goes up again. This behavior is due to its U-shape relation with $b(t)$ in Fig. 9. $S_{CC}$ is the second most informative measure to detect changes.

The behavior of $S_{Sh}$ and $S_{Ba}$ (Fig. 10B second and third rows, respectively) are different from $S_{BF}$ and $S_{CC}$. One can interpret the temporal average of $S_{Sh}$ as an indicator of the level of overall unpredictability of the environment – without a distinction between expected and unexpected unpredictability (Soltani & Izquierdo, 2019; Yu & Dayan, 2005). After the first change-point, when the environment becomes more predictable, the average $S_{Sh}$ gradually decreases from its baseline value, and after the second change-
Figure 10: **Association learning task. A.** The Bayes Factor (BF, black), the Shannon (Sh, blue), the Bayesian (Ba, red), and the Confidence Corrected (CC, green) surprise over time. The exact same experimental sequence was used for all four surprise measures. The dashed grey curve shows $\theta_t$, and the solid grey curve shows $b(t)$ over time. The oracle’s predictions at time $t$ are either $X_t = 0.9$ or $X_t = 0.1$. The stars show the time-points when the oracle’s predictions were wrong. **B.** The averaged version of panel A (over 500 random seeds). The shaded areas indicate the standard error of the mean. The surprise sequences were linearly normalized between 0 and 1 before plotting in both panels. The change-point probability $p_c$ is assumed to be 0.03. Note that the dependence of different surprise measures on $p_c$ is mainly through its influence on the dynamics of belief updating, in contrast to Fig. 9 where the belief $b(t)$ is assumed to be fixed – see Appendix C: Methods for case-studies for details. The qualitative behavior of $S_{BF}$ and $S_{Sh}$ remains the same for other values of $p_c$. The qualitative behavior of $S_{Ba}$ and $S_{CC}$ also remains the same for other values of $p_c$ up to 0.2 (i.e., when there is on average one change-point per 5 time-points), but they behave similarly to $S_{Sh}$ for the more volatile environments (i.e., for higher values of $p_c$). We note that $p_c > 0.2$ (i.e., when there is on average more than one change-point per 5 time-points) is not quite realistic.

In an actual experiment, a behavioral or biological variable $Z$ can be measured throughout the experiment (Fig. 2A1) – similar to our proposed experiments for the example of the oddball tasks (Fig. 7 and Fig. 8). In order to examine with which measure of surprise $Z$ correlates or whether $Z$ is involved in the biological mechanism behind adaptive learning, one can compute its average over different sequences of stimuli, time-locked to change-points from unpredictable to predictable environments. Because different surprise measures have qualitatively different predictions for this average (c.f. Fig. 10B), we predict that their contributions to $Z$ can be dissociated.
To summarize, experiments similar to our proposed experiment can be used for a theory-driven and principled search for signatures of adaptive learning in humans and animals (c.f. Behrens et al., 2007; Marzecová et al., 2019; Nassar et al., 2012; Soltani and Koechlin, 2021; Yu and Dayan, 2005). Moreover, our results indicate that if surprise is used as a signal to modulate learning speed (Gerstner et al., 2018; Iigaya, 2016; Yu & Dayan, 2005), then the Bayes Factor surprise is the best candidate, followed by the Confidence Corrected surprise. We found that neither Shannon surprise nor Bayesian surprise is a reliable indicator of environmental changes. Since the Bayes Factor surprise is the most informative measure about abrupt changes (c.f. Proposition 1), it should be considered in (i) future experimental studies to dissociate the contributions of different surprise measures to explaining behavioral or biological measurements in volatile environments and (ii) future theoretical studies to find normative interpretations for the biological mechanism behind surprise-modulated learning.

Case-study 3: Exploration by surprise-seeking

In our third case-study, we investigate the link between surprise and decision-making. Seeking surprise has often been interpreted as a strategy for exploring available actions and building an accurate model of the environment (Dubey & Griffiths, 2020; Schulz & Gershman, 2019; Sutton & Barto, 2018). Surprise-seeking exploration strategies have been proven efficient in machine learning (Achiam & Sastry, 2017; Burda et al., 2019; Little & Sommer, 2013; Mobin et al., 2014; Pathak et al., 2017; Schmidhuber, 2010) and have been used to model aspects of human curiosity and exploratory behavior (Dubey & Griffiths, 2020; Gottlieb & Oudeyer, 2018; Gottlieb et al., 2013; Schulz & Gershman, 2019). Here, we study how different surprise measures influence action-selection during surprise-seeking. To do so, we study two thought experiments and propose an actual experiment where seeking surprise gives rise to different exploration strategies, depending on the definition of surprise.

General setting: Task formulation. All our experiments can be formulated in the form of reward-free N-armed bandit tasks (Sutton & Barto, 2018). At time $t$, a participant chooses the arm number $x_t \in \{1, ..., N\}$ and observes a binary outcome $y_t \in \{1, 2\}$. In particular, $y_t$ is similar to a reward-indicator in classic bandit tasks (Behrens et al., 2007), but it has no meaning of money or reward in our task. We assume there is no change in the environment so that $p_c = 0$ and $\Theta = \Theta_1 = \ldots = \Theta_t$. The parameter $\Theta$ is a vector of $N$ real values $[P_1, ..., P_N]$, where $P_i$ (taking value $p_i \in (0, 1)$) is the time-invariant probability of observing $Y_t = 1$ after taking the arm $X_t = i$ similar to the reward probability in classic bandit tasks except that no reward is given. More precisely, the time-invariant conditional distribution $P_{Y|X}$ is characterized by

$$P_{Y|X}(y = 1|x = i; \Theta = [p_1, ..., p_N]) = 1 - P_{Y|X}(y = 2|x = i; \Theta = [p_1, ..., p_N]) = p_i. \quad (43)$$

We consider the prior belief to be a Dirichlet distribution\(^4\)

$$\pi^{(0)}(\theta) = \prod_{i=1}^{N} \text{Dir}(p_i, 1 - p_i; \alpha_{\text{sum},i}^{(0)} \cdot [\hat{p}_i^{(0)}, 1 - \hat{p}_i^{(0)}]). \quad (44)$$

Using exact Bayesian inference, the belief $\pi^{(t)}(\theta)$ at time $t$ is also a Dirichlet distribution

$$\pi^{(t)}(\theta) = \prod_{i=1}^{N} \text{Dir}(p_i, 1 - p_i; \alpha_{\text{sum},i}^{(t)} \cdot [\hat{p}_i^{(t)}, 1 - \hat{p}_i^{(t)}]), \quad (45)$$

\(^4\)Note that $p_i$ is the probability of observing $Y_t = 1$ for the arm $X_t = i$, whereas in the first case-study, $p_i$ was the probability of observing $Y_t = i$ and there was no cue variable $X_t$. Similarly, $\alpha_{\text{sum},i}^{(0)}$ is an arm-specific belief parameter, in contrast to the general belief parameter $\alpha_{\text{sum},i}^{(0)}$ in the first case-study.
where $\alpha_{\text{sum},i}^{(t)}$ shows how many samples the belief for arm $i$ is worth (Efron & Hastie, 2016), and $\hat{p}_i^{(t)}$ is the latest estimate of the probability of observing $Y_{t+1} = 1$ in the next time-step given action $X_{t+1} = i$. If action $i$ has been chosen $|T_i^{(t)}|$ times until time $t$, then we have $\alpha_{\text{sum},i}^{(t)} = \alpha_{\text{sum},i}^{(0)} + |T_i^{(t)}|$. Therefore, $\alpha_{\text{sum},i}^{(t)}$ can also be interpreted as a count of how often action $i$ has been chosen until time $t$. See Appendix C: Methods for case-studies for details.

**Thought experiment: (i) Optimal model-building.** Before turning to surprise-seeking exploration policies, we study the optimal exploration policy in our reward-free bandit task. Imagine that an agent is instructed to always choose its next action $x_{t+1}$ in a way to find the best estimate $\hat{\theta}^{(t+1)}$ of the environment parameter $\Theta$ at the next time-step — i.e., to build the most accurate model of the environment (Schmidhuber, 2010). We define the accuracy of the estimate $\hat{\theta}^{(t+1)}$ as the squared error between $\hat{\theta}^{(t+1)}$ and the true parameter $\Theta$. Then, to obtain (on average) the best estimate of the parameter at the next time-step $t+1$, the optimal exploration policy is by definition to choose the action $x_{t+1} = i$ that has in expectation the lowest mean squared error

$$\text{MSE}_{i}^{(t+1)} = \mathbb{E}_{p^{(t)}(Y_{t+1}, \Theta | X_{t+1} = i)} \left[ \left| \hat{\theta}^{(t+1)} - \Theta \right|^2 \right],$$

(46)

where $\| \cdot \|_2$ stands for the $\ell_2$-norm. Note that $\Theta$ is a random variable, and $\hat{\theta}^{(t+1)}$ is the agent’s estimate of parameter $\Theta$ at the next time-step $t+1$ which depends on $Y_{t+1}$ and is itself also a random variable. Thus, the expectation in Eq. 46 is taken over all values of the parameter $\Theta$ and the next observation $Y_{t+1}$, and it is conditioned on the previous actions $x_{1:t}$, the previous observations $y_{1:t}$, and the hypothetical choice $i$ of the next action $X_{t+1}$. The optimal policy can be re-written as choosing the action that maximizes an optimal gain function $g^* (\hat{p}_i^{(t)}, \alpha_{\text{sum},i}^{(t)})$, i.e.,

$$x_{t+1} = f^*(x_{1:t}, y_{1:t}) = \arg \min_{i \in \{1, \ldots, N\}} \text{MSE}_{i}^{(t+1)} = \arg \max_{i \in \{1, \ldots, N\}} g^* (\hat{p}_i^{(t)}, \alpha_{\text{sum},i}^{(t)}),$$

(47)

where (c.f. Appendix C: Methods for case-studies and Fig. 11A)

$$g^* (\hat{p}_i^{(t)}, \alpha_{\text{sum},i}^{(t)}) = \frac{\hat{p}_i^{(t)} (1 - \hat{p}_i^{(t)})}{(\alpha_{\text{sum},i}^{(t)} + 1)^2}.$$  

(48)

The optimal gain function $g^*$ indicates which action should be chosen (Fig. 11A). To grasp the idea of the optimal policy, let us first imagine that we have chosen arm 1 and 2 ten times each. If we observed $Y_t = 1$ after every single time we chose arm 1 but after only 50% of times when we chose arm 2, then we would naturally be more confident about our estimate of $\hat{p}_1$ than our estimate of $\hat{p}_2$. Thus, in order to increase the precision of our estimates, we should keep choosing arm 2. Consistent with this intuition, the optimal gain function $g^*$ has an inverted-U-relation with the estimated probability $\hat{p}_i^{(t)}$ with its maximum at $\hat{p}_i^{(t)} = 0.5$ (Fig. 11A1). This means that the optimal policy is to pick the arm with the highest stochasticity in its outcome distribution (among the arms with the same $\alpha_{\text{sum},i}^{(t)}$).

Second, let us imagine that we have chosen arm 1 only once but have chosen arm 2 ten times. In this case, whatever the actual observations, we would still be less confident about our estimate of $\hat{p}_1$ than our estimate of $\hat{p}_2$. Thus, in order to increase the precision of our estimates, we should choose arm 1 more often. Consistent with this intuition, the optimal gain function $g^*$ is a decreasing function of $\alpha_{\text{sum},i}^{(t)}$, implying that the more often an agent chooses arm $i$, the less informative it becomes (Fig. 11A2).

Moreover, for any two arms $i$ and $j$ with different estimated probabilities $\hat{p}_i^{(t)}$ and $\hat{p}_j^{(t)}$, if arm $i$ is less stochastic than the arm $j$ (i.e., if $\hat{p}_j^{(t)}$ is closer to 0.5), then there are always choice counts $\alpha_{\text{sum},i}^{(t)} < \alpha_{\text{sum},j}^{(t)}$ such that the arm $i$ with less stochastic outcome gets selected according to $g^*$ — for example, in Fig. 11A1,
the maximum of the blue curve (for \( \alpha_{\text{sum},j}^{(t)} = 10 \)) at \( \tilde{p}_j^{(t)} = 0.5 \) is below the minimums of the red curve (\( \alpha_{\text{sum},j}^{(t)} = 3 \)) at \( \tilde{p}_i^{(t)} = 0.1 \) and \( \tilde{p}_i^{(t)} = 0.9 \). This implies that also the least stochastic arm gets selected occasionally and all arms get selected infinitely many times in the limit \( t \to \infty \). We note that \( g^* \) is independent of the prior parameters \( \alpha_{\text{sum},i}^{(0)} \) and \( \tilde{p}_i^{(0)} \) (Fig. 11A3 and A4).

To summarize, the optimal exploration policy prefers arms that (i) are more stochastic and (ii) have been chosen less often. See Appendix C: Methods for case-studies for details and proofs.

**Thought experiment: (ii) Surprise-seeking.** Now, imagine that the primary goal of an agent, instead of building the most accurate model of the environment, is to choose its next action \( x_{t+1} \) in a way to be maximally surprised at the next time-step \( t + 1 \) (Burda et al., 2019; Little & Sommer, 2013; Mobin et al., 2014; Pathak et al., 2017; Storck et al., 1995). Our goal is to study different surprise-seeking exploration policies and compare them with the optimal policy for model-building discussed in the first thought experiment.

Given a measure of surprise \( S \), a classic surprise-seeking behavior is to choose the action \( i \) which maximizes the expected \( S \) of the next observation (Burda et al., 2019; Little & Sommer, 2013; Mobin et al., 2014; Pathak et al., 2017; Storck et al., 1995). Formally, we write

\[
x_{t+1} = f_S(x_1:t, y_1:t) = \arg \max_{i \in \{1, \ldots, N\}} \mathbb{E}_{p(\cdot|x=i; \pi(t))}[S(Y_{t+1}|x_{t+1} = i; \pi(t))].
\]

(49)

It is important to note that the policy in Eq. 49 is not invariant with respect to non-linear transformations of surprise measures. This is in contrast to our previous case-studies where non-linear transformations of surprise measures did not change their qualitative behavior (c.f. Fig. 2A). Eq. 49 will be applied to all surprise measures except \( S_{\text{BF}} \). For the Bayes Factor surprise, we define the surprise-seeking policy as

\[
x_{t+1} = f_{S_{\text{BF}}}(x_1:t, y_1:t) = \arg \max_{i \in \{1, \ldots, N\}} \mathbb{E}_{p(\cdot|x=i; \pi(t))}[\log S_{\text{BF}}(Y_{t+1}|x_{t+1} = i; \pi(t))],
\]

(50)

which is the same as seeking the difference in Shannon surprise (c.f. Proposition 2). Note that, because \( \mathbb{E}_{p(\cdot|x=i; \pi(t))}[S_{\text{BF}}(Y_{t+1}|x_{t+1} = i; \pi(t))] \) is always and by definition equal to 1, classic surprise-seeking (Eq. 49) with the Bayes Factor surprise is equivalent to uniformly random exploration.

Similar to the case of the optimal exploration policy (Eq. 47), we can define a gain function for each measure of surprise and write the corresponding surprise-seeking policy as the one that maximizes that gain function, i.e., in general, for a measure of surprise \( S \), the surprise-seeking policy can be written as

\[
f_S(x_1:t, y_1:t) = \arg \max_{i \in \{1, \ldots, N\}} g_S(p_i^{(t)}, \alpha_{\text{sum},i}^{(t)}, \tilde{p}_i^{(0)}, \alpha_{\text{sum},i}^{(0)}),
\]

(51)

with \( g_S \) the corresponding gain function; see Appendix C: Methods for case-studies for details. Different surprise measures give rise to different gain functions and show different preferences over actions (Fig. 11).

According to \( g_{\text{Sh}} \) (Fig. 11B1), arms with more stochastic outcomes (i.e., with \( \tilde{p}_i^{(t)} \) closer to 0.5) are preferred to arms with more deterministic outcomes (i.e., with \( \tilde{p}_i^{(t)} \) closer to 0 or 1), in agreement with the optimal gain function \( g^* \) (Fig. 11A1). However, according to \( g_{\text{Ba}} \) the opposite is true (Fig. 11C1). The reason is that improbable events lead to huge changes in the agent’s belief such that more deterministic arms have higher expected \( S_{\text{Ba}} \). On the other hand, \( g_{\text{Ba}} \) is a decreasing function of \( \alpha_{\text{sum},i}^{(t)} \) (Fig. 11C2) and in agreement with \( g^* \) (Fig. 11A2). The preference of \( g_{\text{Ba}} \) for deterministic arms also decreases with increasing \( \alpha_{\text{sum},i}^{(t)} \ – \) e.g., the maximum of the blue curve in Fig. 11C1 has a smaller value than the minimum of the red curve. This means that independently of the difference in the stochasticity of their outcomes, \( g_{\text{Ba}} \) will eventually choose arms that have been chosen less often, in agreement with \( g^* \). However, since \( S_{\text{Sh}} \) is a probabilistic mismatch surprise measure (Fig. 3), \( g_{\text{Sh}} \) is independent of \( \alpha_{\text{sum},i}^{(t)} \): How often an arm
Figure 11: **Gain functions for the reward-free bandit task.** The optimal exploration policy and different surprise-seeking policies can be written as maximizing a gain function that depends on the current belief parameters $\hat{p}_i(t)$ and $\alpha_{\text{sum,}i}^{(t)}$ and the prior belief parameters $\hat{p}_i^{(0)}$ and $\alpha_{\text{sum,}i}^{(0)}$. For two different choices of parameters, the blue ($\hat{p}_i^{(t)} = 0.2, \alpha_{\text{sum,}i}^{(t)} = 10, \hat{p}_i^{(0)} = 0.5, \alpha_{\text{sum,}i}^{(0)} = 2$) and the red ($\hat{p}_i^{(t)} = 0.5, \alpha_{\text{sum,}i}^{(t)} = 3, \hat{p}_i^{(0)} = 0.2, \alpha_{\text{sum,}i}^{(0)} = 5$) curves show the gain functions for A. the optimal policy and the policy based on seeking B. the Shannon (B1: the red curve covers the blue one), C. the Bayesian, D. the Bayes Factor, and E. the Confidence Corrected surprise as functions of $\hat{p}_i^{(t)}$ (left column, A1-E1), $\alpha_{\text{sum,}i}^{(t)}$ (A2-E2), $\hat{p}_i^{(0)}$ (A3-E3), and $\alpha_{\text{sum,}i}^{(0)}$ (A4-E4). For the solid red curves in the left column (A1-E1), $\alpha_{\text{sum,}i}^{(t)}$, $\hat{p}_i^{(0)}$, and $\alpha_{\text{sum,}i}^{(0)}$ are fixed at their values corresponding to the red dashed lines in the other columns. The same rule applies to the blue curves and to the other columns. Faded panels indicate independence, i.e., when the gain function does not depend on (is constant with respect to) the corresponding parameter. See Appendix C: Methods for case-studies for details.

This observation is consistent with different behaviors of $S_{\text{Sh}}$ and $S_{\text{Ba}}$ with respect to confidence in our first case-study (Fig. 5). Therefore, seeking $S_{\text{Ba}}$ has an asymptotic behavior similar to that of the optimal policy, whereas seeking $S_{\text{Sh}}$ remains systematically different from the optimal policy (Fig. 11A-C).
closer their prediction \(\hat{\theta}(t)\) and the explicit reward, they are informed that their monetary payoff at the end of the experiment increases the actual experiment to study human exploratory behavior and test predictions of different exploration policies depending on the definition of surprise. Seeking the Bayesian surprise gets eventually closer to the optimal policy with \(\alpha\) sub-optimal. To summarize, seeking surprise gives rise to different exploration policies depending on the definition of surprise. Seeking the Bayesian surprise gets eventually closer to the optimal policy with increasing \(\alpha_{\text{sum,}i}\), but seeking the Shannon or the Bayes Factor surprise is always systematically different from the optimal policy. The preferences when seeking the Confidence Corrected surprise are often opposite to the preferences of the optimal policy.

Proposed experiment: Exploration in reward-free 10-armed bandit. In this section, we propose an actual experiment to study human exploratory behavior and test predictions of different exploration policies. We consider a 10-armed bandit task. To motivate participants to explore in the absence of any explicit reward, they are informed that their monetary payoff at the end of the experiment increases the closer their prediction \(\hat{\theta}(t) = [\hat{p}_1(t), ..., \hat{p}_N(t)]\) matches the true parameter \(\theta = [p_1, ..., p_N]\):

\[
\text{monetary payoff } \propto 1 - \frac{||\hat{\theta}(t) - \theta||^2}{N}.
\]  

Therefore, to maximize payoff, participants need to choose actions \(x_{t+1}\) so as to best update their estimates \([\hat{p}_1(t), ..., \hat{p}_N(t)]\) of the parameter. In order to dissociate learning and memory from decision-making, the Bayes-optimal estimate for \([\hat{p}_1(t), ..., \hat{p}_N(t)]\) and its corresponding confidence can be updated automatically and shown on a computer screen (e.g., see Kobayashi et al., 2019; Prat-Carrabin et al., 2021). Then the only task of participants is to choose a strategy for action-selection – i.e., to choose the next action \(x_{t+1}\).

Our goal is to study (i) how well each policy performs in terms of model-building, and (ii) whether we can dissociate between different policies if we can only observe a finite number of their action choices. To do so, we first (i) simulate different policies and measure their performance over time. Then, we (ii) apply model-recovery (Wilson & Collins, 2019) to the simulated data in order to see if we can recover the true policy from a finite number of action-choices – i.e., to see if our task can distinguish between different policies on the level of action-choices. See Appendix C: Methods for case-studies for details. We consider two scenarios: in the first scenario, the prior outcome probability \(\hat{p}_i(0) = 0.5\) is the same for all arms – with \(\alpha_{\text{sum,}i} = 2\); in the second scenario, \(\hat{p}_i(0)\) is chosen from a wide interval from 0.05 to 0.95 – with
Figure 12: **Exploration performance and model-recovery in a reward-free 10-armed bandit task.**

A. The mean squared error (MSE) of the estimate of the optimal policy ($g^*$) and different surprise-seeking policies as a function of time $t$ – averaged over 500 random seeds. The prior parameters $\alpha_{\text{sum},i}^{(0)}$ and $\hat{p}_i^{(0)}$ were fixed for all arms at 2 and 0.5, respectively. The shaded areas indicate the standard error of the mean.

B. Confusion matrix for model recovery (Wilson & Collins, 2019) using data shown in panel A. We used classic Bayesian model selection (Efron & Hastie, 2016; Kass & Raftery, 1995) and, given a sequence of action-choices and observations (for one random seed), computed the posterior probability of different models. Each cell shows the posterior probability of a candidate model (corresponding column) given that the action-choices were made by one of the true models (corresponding row) – averaged over 500 random seeds.

C. and D. The same as panels A and B, respectively, except that the prior parameter $\alpha_{\text{sum},i}^{(0)}$ was fixed at 10 for all arms, but $\hat{p}_i^{(0)}$ for each arm was chosen randomly between 0.05 and 0.95. Seeking Bayesian surprise ($g_{\text{Ba}}$, red) remains closest to the optimal policy also for other choices and distributions of the prior parameters $\alpha_{\text{sum},i}^{(0)}$ and $\hat{p}_i^{(0)}$. See Appendix C: Methods for case-studies for details.

In the first scenario, the effect of $\alpha_{\text{sum},i}^{(0)}$ on the policies is dominant, whereas in the second scenario, $\hat{p}_i^{(0)}$ also plays a role in action-selection.
In the second scenario, maximizing $g_{Sh}$ differs from maximizing $g_{BF}$ (Fig. 12C). By constantly comparing $\hat{p}_i(t)$ with $\hat{p}_i(0)$, seeking $S_{BF}$ has an indirect preference for arms that have been chosen less often; therefore, seeking $S_{BF}$ has a better performance than maximizing $g_{Sh}$. While the best surprise-seeking policy is still to maximize $g_{Ba}$, its difference with the optimal policy becomes more obvious in the second scenario – when $\hat{p}_i(t)$ matters. In this scenario, $g_{CC}$ is again the worst policy and all policies can be distinguished from each other given their action-choices (Fig. 12D).

To summarize, our proposed experiment can distinguish different exploration policies both at the level of (i) performance and (ii) action-choices with as few as 100 trials. We found that seeking the Bayesian surprise leads to an exploration policy closest to the optimal policy for model-building. This explains the popularity of the Bayesian surprise in the field of machine learning (Little & Sommer, 2013; Mobin et al., 2014; Schmidhuber, 2010; Storck et al., 1995). Seeking the Shannon (or Bayes Factor) surprise differs more from the optimal policy as time passes – as a result of their indifference with respect to $\alpha_{\text{sum},t}^{(t)}$. Seeking the Confidence Corrected surprise leads to a policy that favors situations that differ most from prior expectations, and the resulting policy is often opposite to the optimal policy. As a result, a reward-free multi-armed bandit task as described above can be used to study the degree of suboptimality of human exploration strategies and be used to compare them with different models of human curiosity (Dubey & Griffiths, 2020; Gottlieb & Oudeyer, 2018) and surprise-seeking policies.

Summary and intermediate discussion (ii)

In the three case-studies, we showed that the 4 different categories of surprise in Fig. 4 have different predictions regarding (i) the perception of surprise, (ii) adaptive learning by surprise, and (iii) exploration by surprise-seeking. We designed experiments where these predictions can be tested in practice. Our theoretical analyses, therefore, can be used to correlate physiological variables with different measures of surprise, search for signatures of adaptive learning in the brain, and study human curiosity and exploratory behavior.

Our results suggest that the Bayes Factor surprise is the most suitable measure of surprise to modulate learning in volatile environments (Frémaux & Gerstner, 2016; Gerstner et al., 2018; Iigaya, 2016; Liakoni et al., 2021; Soltani & Izquierdo, 2019; Yu & Dayan, 2005) and the Bayesian surprise is the most suitable measure of surprise to model curiosity as an innate mechanism in humans for exploration – whenever model-building is crucial for better future performance (Dubey & Griffiths, 2019, 2020; Schmidhuber, 2010; Singh, Lewis, & Barto, 2010; Singh, Lewis, Barto, & Sorg, 2010; Storck et al., 1995). In particular, we showed that a surprise measure that is useful to modulate the learning speed should necessarily compare the current belief with the prior belief (Fig. 9 and Fig. 10), while we also showed that such a comparison leads to suboptimal exploration in surprise-seeking policies (Fig. 11 and Fig. 12). Similarly, we showed that a surprise measure that is useful for exploration should increase with increasing uncertainty (Fig. 11 and Fig. 12), whereas we also showed that such a behavior is slow and suboptimal in detecting change-points (Fig. 9 and Fig. 10). In other words, our results show that the very features that make the Bayes Factor surprise an appropriate measure for learning make it an unsuitable measure for exploration and vice versa for the Bayesian surprise.

Unlike the Bayes Factor and the Bayesian surprise, the Shannon surprise (c.f. Eq. 9 and Eq. 10) considers less likely events always as more surprising. Surprise in natural language is defined as ‘the feeling or emotion excited by something unexpected’ (Oxford English Dictionary, n.d.). If we focus on the term ‘unexpected’, identify it with ‘unlikely under the current belief’, and neglect the terms ‘feeling’ and ‘emotion’, then our results suggest that the Shannon surprise measures a quality closely related to the definition of surprise in natural language (i.e., the dictionary definition of surprise). However, we observed that the Confidence Corrected surprise has a more intuitive behavior in some cases where confidence (or commitment to a belief) plays an important role (Fig. 6), despite its counter-intuitive behavior in some
Other situations (Fig. 5).

**Regularized Shannon surprise: A new direction**

We argued that the Confidence Corrected surprise has a more intuitive behavior than the other measures in the CEO selection experiment because it explicitly accounts for confidence (Fig. 6). However, because it treats the confidence for a *correct* prediction in the same way as the confidence for a *wrong* prediction, its behavior is against common sense in some other experiments (Fig. 5 and Fig. 7). In this section, we overcome this issue by providing a modified definition for the Shannon surprise which only penalizes the confidence for a *wrong* prediction.

![Figure 13: (Caption next page.)](image-url)
Figure 13: (Previous page.) **The behavior of the regularized Shannon surprise for different case-studies.** Each panel corresponds to one panel in Fig. 5-Fig. 12. Sh (dark blue) corresponds to the Shannon surprise and ShR (light blue) corresponds to the regularized Shannon surprise. The regularized Shannon surprise has the same behavior as the Shannon surprise in panels A-E as well as in panels I-K, but it has the same behavior as the Confidence Corrected surprise in panels F-H. **A.** Coin flipping experiment: Surprise of $Y_{t+1} = 1$ as a function of $\alpha_{\text{sum}}^{(t)}$ (c.f. Fig. 5A1). **B.** Coin flipping experiment: Surprise of $Y_{t+1} = 1$ as a function of $\hat{p}_1^{(t)}$ (c.f. Fig. 5B1). **C.** Classic oddball task: Surprise over time for standard (shown at -0.05) and deviant (shown at 1.05) stimuli (c.f. Fig. 7A). **D.** Classic oddball task: Average surprise of standard stimuli in the early and the late phase of the task (c.f. Fig. 7B). **E.** Classic oddball task: Average surprise of deviant stimuli in the early and the late phase of the task (c.f. Fig. 7C). **F.** CEO selection experiment: Surprise of $Y_{t+1} = 1$ as a function of $\hat{p}_2^{(t)}$ (c.f. Fig. 6A1). **G.** CEO selection experiment: Surprise over time (c.f. Fig. 8A). **H.** Generalized oddball task: Average surprise values of $Y_{t+1} = 1$ in Phase 1 and Phase 2 (c.f. Fig. 8C). **I.** Belief in forecasts experiment: The regularized Shannon surprise as a function of trust in media $b(t)$ for media being right (blue) or wrong (red); $b(0) = 0.5$ (c.f. Fig. 9B). **J.** Association learning experiment: Surprise over time for one random seed; stars indicate when oracle’s prediction has been wrong (c.f. Fig. 10A). **K.** Association learning experiment: Surprise over time (as in panel J) averaged over 500 random seeds (c.f. Fig. 10B). **L.** Bandit experiment: The gain function $g_{\text{ShR}}$ corresponding to the policy of seeking ShR as a function of $\hat{p}_i^{(t)}$ (c.f. Fig. 11). $g_{\text{ShR}}$ is independent of $\alpha_{\text{sum},i}^{(t)}$, $\alpha_{\text{sum},i}^{(0)}$, and $\hat{p}_i^{(0)}$. **M-N.** Bandit experiment: Exploration performance for seeking ShR (light blue) compared to seeking Sh (dark blue) and the optimal policy $g^*$ (black) for M. $\alpha_{\text{sum},i}^{(0)} = 2$ and $\hat{p}_i^{(0)} = 0.5$ (c.f. Fig. 12A) and N. $\alpha_{\text{sum},i}^{(0)} = 10$ and $\hat{p}_i^{(0)} \sim \text{Uniform}(0.05, 0.95)$ (c.f. Fig. 12B). See Appendix E: Regularized Shannon surprise for details.

We call the modified measure the **Regularized Shannon surprise** and define it as

$$S_{\text{ShR}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Sh}}(y_{t+1}|x_{t+1}; \pi^{(t)}) + R \left( S_{\text{Sh}}(y_{t+1}|x_{t+1}; \pi^{(t)}) - \min_y S_{\text{Sh}}(y|x_{t+1}; \pi^{(t)}) \right), \quad (53)$$

where $R : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ can be any continuous function that satisfies: (i) $R(0) = 0$, and (ii) $R(z)$ is an increasing function of $z$ for all $z \in \mathbb{R}^+$. As a result, we have the following two properties for the regularized Shannon surprise:

1. Whenever $y_{t+1}$ is the most expected observation, i.e., $y_{t+1} = \arg\min_y S_{\text{Sh}}(y|x_{t+1}; \pi^{(t)})$, we have $S_{\text{ShR}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Sh}}(y_{t+1}|x_{t+1}; \pi^{(t)})$. This means that the confidence for a correct prediction does not penalize surprise.

2. Whenever $y_{t+1}$ is not the most expected observation, the difference between the regularized Shannon surprise and the Shannon surprise, i.e., $S_{\text{ShR}}(y_{t+1}|x_{t+1}; \pi^{(t)}) - S_{\text{Sh}}(y_{t+1}|x_{t+1}; \pi^{(t)})$, is a decreasing function of the Shannon surprise of the most likely observation $y_{t+1}^* = \arg\min_y S_{\text{Sh}}(y|x_{t+1}; \pi^{(t)})$. This means that for a fixed Shannon surprise of $y_{t+1}$, the more we expect another observation $y_{t+1}^*$, the more surprised we are by observing $y_{t+1} \neq y_{t+1}^*$, consistent with our expectation for the CEO selection experiment (Fig. 6A1).

As a simple choice, let us consider $R(z) = z$, i.e.,

$$S_{\text{ShR}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = 2S_{\text{Sh}}(y_{t+1}|x_{t+1}; \pi^{(t)}) - \min_y S_{\text{Sh}}(y|x_{t+1}; \pi^{(t)}). \quad (54)$$

Then, the predictions of the regularized Shannon surprise $S_{\text{ShR}}$ are the same as the predictions of the Shannon surprise $S_{\text{Sh}}$ for the cases where $S_{\text{Sh}}$ in the coin flipping thought experiment and the classic
oddball task (compare Fig. 13A-E with Fig. 5 and Fig. 7A-C). Importantly, the EEG amplitude at
around 450ms in the visual oddball task we analyzed has a behavior consistent with the predictions of
\( S_{SR} \) but not with the predictions of \( S_{CC} \) (compare Fig. 13D-E with Fig. 7). In the belief in forecast and
association learning experiments, \( S_{SR} \), similar to \( S_{Sh} \), can be interpreted as a measure of unpredictability
and unexpectedness in the environment (Fig. 13I-K). In this regard, \( S_{SR} \) is also a good model of the
dictionary definition of surprise. On the other hand, in the CEO selection experiment and the generalized
oddball task, the confidence regularization in \( S_{SR} \) makes it more similar to \( S_{CC} \) than \( S_{Sh} \) (compare
Fig. 13F-H with Fig. 6 and Fig. 8): Given a fixed estimated probability \( \hat{p}_1(t) \) for the selection of the 1st
candidate \( Y_{t+1} = 1 \), the regularized Shannon surprise for \( Y_{t+1} = 1 \) increases with increasing expectation
for the selection of another candidate (e.g., \( Y_{t+1} = 2 \)) – in agreement with the results of SanMiguel et al.,
2021. Therefore, \( S_{SR} \) captures those features of \( S_{Sh} \) that make it similar to the dictionary definition of
surprise but adds a dependence upon confidence.

Finally, our results show that, similar to seeking \( S_{Sh} \) or \( S_{CC} \), seeking \( S_{SR} \) is a sub-optimal exploration
policy and has a poor performance for model-building (Fig. 13L-N and Appendix C: Methods for case-
studies).

General discussion

What does it formally mean to be surprised? How can we experimentally distinguish between different
surprise measures? How do different surprise measures relate to different computational roles of surprise
in the brain? And how do they relate to the word ‘surprise’ in natural language? To address these
question, we reviewed 16 surprise measures in a common mathematical framework and studied their
links to perception, learning, and decision-making. We identified the conditions under which they are
indistinguishable (Fig. 2) and provided a technical (Fig. 3) and a conceptual (Fig. 4) categorization of
these measures.

Our results suggest that the class of prediction surprise measures (Fig. 4) is the closest one to the
dictionary definition of surprise (see Summary and intermediate discussion (ii)). However, following Faraji
et al., 2018, we also suggested that a surprise measure should explicitly capture the notion of confidence
in order to explain intuitive aspects of surprise perception. We therefore proposed the ‘Regularized
Shannon surprise’ (Eq. 53) and show how its predictions can be tested experimentally. Our predictions
are supported by recent experimental evidence (SanMiguel et al., 2021).

We found that the very features that make a surprise measure suitable for adaptive learning are in conflict
with the ones that make it suitable for exploration. In particular, adaptive behavior as observed in humans
is achieved only if a surprise from the class of change-point detection measures modulates learning (Fig. 4),
whereas a close-to-optimal exploration strategy is achieved only if a surprise from the class of information-
gain surprise measures drives action-selection (Fig. 4). Change-point detection surprise measures compare
the probability of an event under the prior belief with that under the current belief, which is why they are
useful in adaptive learning (Liakoni et al., 2021; Soltani & Izquierdo, 2019; Yu & Dayan, 2005). In the
sense of ‘unexpected under the current belief’, change-point detection surprise measures are also similar
to the dictionary definition of surprise. Information-gain surprise measures, however, essentially differ
from the definition of surprise in natural language. We, therefore, suggest to avoid the term ‘surprise’
when referring to these measures. Specifically, we suggest the names ‘Bayesian information gain’ and
‘Postdictive belief-update’ instead of ‘Bayesian surprise’ and ‘Postdictive surprise’ – similar to Kolossa
et al., 2015.

Based on our theoretical results, we proposed experimental paradigms where different surprise categories
make different predictions. A natural direction for future studies is to test these predictions experimentally.
A wide range of questions can be addressed, either directly or indirectly, such as ‘which surprise-seeking
policy is most similar to human exploratory behavior in different experiments?’ (Dubey & Griffiths, 2020; Gottlieb & Oudeyer, 2018), ‘which surprise measure correlates best with pupil diameter (Antony et al., 2021; Nassar et al., 2012; Preuschoff et al., 2011), with different EEG components (Gijsen et al., 2021; Kolossa et al., 2015; Mousavi et al., 2020; Visalli et al., 2021), or with fMRI bold activity (Gläscher et al., 2010; Konovalov & Krajbich, 2018; Loued-Khenissi & Preuschoff, 2020)?’, and even ‘which surprise measure explains best the neural activity in the locus coeruleus?’ (Sara, 2009).

Some important open questions on the mechanistic level (Marr, 1982) are ‘which neural circuits are involved in computation of different surprise measures?’ (see Fiser et al., 2010; Knill and Pouget, 2004; Soltani and Wang, 2010 for examples of neural models of probabilistic inference), ‘how can surprise-modulated adaptive learning be implemented at the level of synaptic plasticity?’ (Berlemont & Nadal, 2021; Gerstner et al., 2018; Iigaya, 2016; Illing et al., 2021), and ‘how can surprise-seeking exploration strategies be implemented in the brain?’ (see Basanisi et al., 2020 for an example). There are also important open questions on the computational and algorithmic levels (Marr, 1982) such as ‘how can we model the influence of attention on surprise perception?’ (Gottlieb & Oudeyer, 2018), ‘how can we formulate other surprise theories like Palm’s theory of surprise (Palm, 2012) or the theory of surprise as a compression measure (Schmidhuber, 2010) in our mathematical framework?’, and ‘how do different surprise measures contribute to formation (Rouhani & Niv, 2021), segmentation (Antony et al., 2021; Rouhani et al., 2020), and modification (Gershman et al., 2017; Sinclair & Barense, 2018) of memory?’.

In conclusion, our results articulate and unify many of the existing paradigms for the study of surprise and suggests various new directions for further theoretical and experimental studies of surprise and its roles in the brain function.

References

Abramowitz, M., & Stegun, I. A. (1948). Handbook of mathematical functions with formulas, graphs, and mathematical tables. US Government printing office.

Achiam, J., & Sastry, S. (2017). Surprise-based intrinsic motivation for deep reinforcement learning. arXiv preprint arXiv:1703.01732.

Adams, R. P., & MacKay, D. J. (2007). Bayesian online changepoint detection. arXiv preprint arXiv:0710.3742.

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hassun, U., McDougle, S. D., & Norman, K. A. (2021). Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. Neuron, 109(2), 377–390.e7. https://doi.org/10.1016/j.neuron.2020.10.029

Baldi, P. (2002). A computational theory of surprise. In M. Blaum, P. G. Farrell, & H. C. A. van Tilborg (Eds.), Information, coding and mathematics: Proceedings of workshop honoring prof. bob mceliece on his 60th birthday (pp. 1–25). Springer US. https://doi.org/10.1007/978-1-4757-3585-7_1

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

Barber, D. (2012). Bayesian reasoning and machine learning. Cambridge University Press. https://doi.org/10.1017/CBO9780511804779

Barto, A., Miorilli, M., & Baldassarre, G. (2013). Novelty or surprise? Frontiers in Psychology, 4, 907. https://doi.org/10.3389/fpsyg.2013.00907

Basanisi, R., Brovelli, A., Cartoni, E., & Baldassarre, G. (2020). A generative spiking neural-network model of goal-directed behaviour and one-step planning. PLOS Computational Biology, 16(12), 1–32. https://doi.org/10.1371/journal.pcbi.1007579

Bayarri, M., & Berger, J. O. (1997). Measures of surprise in bayesian analysis. Duke University.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. Nature neuroscience, 10, 1214–1221. https://doi.org/10.1038/nn1954
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.

Berlemont, K., & Nadal, J.-P. (2021). Confidence-controlled hebbian learning efficiently extracts category membership from stimuli encoded in view of a categorization task. *bioRxiv*. https://doi.org/10.1101/2020.08.06.239533

Blagouchine, I. V. (2016). Three notes on ser’s and hasse’s representations for the zeta-functions. *arXiv preprint arXiv:1606.02044*.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2019). Large-scale study of curiosity-driven learning. *International Conference on Learning Representations*.

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons.

Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.

da Silva, C. F., & Hare, T. A. (2020). Humans primarily use model-based inference in the two-stage task. *Nature Human Behaviour*, 4, 1053–1066. https://doi.org/10.1038/s41562-020-0905-y

Daw, N., & Courville, A. (2008). The pigeon as particle filter. *Advances in neural information processing systems*, 20, 369–376.

Daw, N., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans’ choices and striatal prediction errors. *Neuron*, 69(6), 1204–1215. https://doi.org/j.neuron.2011.02.027

Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, 8, 1704–1711. https://doi.org/10.1038/nn1560

Dubey, R., & Griffiths, T. L. (2019). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review*, 127(3), 455–476. https://doi.org/10.1037/rev0000175

Dubey, R., & Griffiths, T. L. (2020). Understanding exploration in humans and machines by formalizing the function of curiosity. *Current Opinion in Behavioral Sciences*, 35, 118–124. https://doi.org/10.1016/j.cobeha.2020.07.008

Dubey, R., Ho, M. K., Mehta, H., & Griffiths, T. (2021). Aha! moments correspond to meta-cognitive prediction errors. *PsyrXiv*. https://doi.org/10.31234/osf.io/c5v42

Efron, B., & Hastie, T. (2016). *Computer age statistical inference*. Cambridge University Press.

Esmaeili, V., Tamura, K., Muscinelli, S. P., Modirshanechi, A., Boscaglia, M., Lee, A. B., Oryshchuk, A., Foustoukos, G., Liu, Y., Crochet, S., et al. (2021). Rapid suppression and sustained activation of distinct cortical regions for a delayed sensory-triggered motor response. *Neuron*, 109(13), 2183–2201.e9. https://doi.org/10.1016/j.neuron.2021.05.005

Exton-McGuinness, M. T., Lee, J. L., & Reichelt, A. C. (2015). Updating memories—the role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, 278, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011

Faraji, M., Preuschhof, K., & Gerstner, W. (2018). Balancing new against old information: The role of puzzlement surprise in learning. *Neural computation*, 30(1), 34–83. https://doi.org/10.1162/neco_a_01025

Fearnhead, P., & Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4), 589–605. https://doi.org/10.1111/j.1467-9868.2007.00601.x

Findling, C., Chopin, N., & Koechlin, E. (2021). Imprecise neural computations as a source of adaptive behaviour in volatile environments. *Nature Human Behaviour*, 5, 99–112. https://doi.org/10.1038/s41562-020-00971-z
Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Sciences*, 14(3), 119–130. https://doi.org/10.1016/j.tics.2010.01.003

Frémaux, N., & Gerstner, W. (2016). Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9, 85. https://doi.org/10.3389/fncir.2015.00085

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Gershman, S. J. (2019). What does the free energy principle tell us about the brain? *Neurons, Behavior, Data analysis, and Theory*, 2(3), 1–10.

Gershman, S. J., & Blei, D. M. (2012). A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1), 1–12. https://doi.org/10.1016/j.jmp.2011.08.004

Gershman, S. J., Monfils, M.-H., Norman, K. A., & Niv, Y. (2017). Statistical computations underlying the dynamics of memory updating. *PLOS Computational Biology*, 10, 1–13. https://doi.org/10.1371/journal.pcbi.1003939

Gerstner, W., Lehmann, M., Liakoni, V., Corneil, D., & Brea, J. (2018). Eligibility traces and plasticity on behavioral time scales: Experimental support of neohebbian three-factor learning rules. *Frontiers in Neural Circuits*, 12. https://doi.org/10.3389/fncir.2018.00053

Ghahramani, Z. (2013). Bayesian non-parametrics and the probabilistic approach to modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984), 20110553. https://doi.org/10.1098/rsta.2011.0553

Gijsen, S., Grundei, M., Lange, R. T., Ostwald, D., & Blankenburg, F. (2021). Neural surprise in somatosensory bayesian learning. *PLOS Computational Biology*, 17(2), 1–36. https://doi.org/10.1371/journal.pcbi.1008068

Gläscher, J., Daw, N., Dayan, P., & O’Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4), 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Glaze, C. M., Kable, J. W., & Gold, J. I. (2015). Normative evidence accumulation in unpredictable environments. *Elife*, 4, e08825. https://doi.org/10.7554/eLife.08825.001

Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, 19, 758–770. https://doi.org/10.1038/s41583-018-0078-0

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, 17(11), 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Hairston, W. D., Whitaker, K. W., Ries, A. J., Vettel, J. M., Bradford, J. C., Kerick, S. E., & McDowell, K. (2014). Usability of four commercially-oriented EEG systems. *Journal of Neural Engineering*, 11(4), 046018. https://doi.org/10.1088/1741-2560/11/4/046018

Heilbron, M., & Meyniel, F. (2019). Confidence resets reveal hierarchical adaptive learning in humans. *PLoS computational biology*, 15(4), e1006972. https://doi.org/10.1371/journal.pcbi.1006972

Heydari, S., & Holroyd, C. B. (2016). Reward positivity: Reward prediction error or salience prediction error? *Psychophysiology*, 53(8), 1185–1192. https://doi.org/10.1111/psyp.12673

Horvath, L., Colcombe, S., Milham, M., Ray, S., Schwartenbeck, P., & Ostwald, D. (2021). Human belief state-based exploration and exploitation in an information-selective symmetric reversal bandit task. *Computational Brain & Behavior*. https://doi.org/10.1007/s42113-021-00112-3
Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature neuroscience, 5*(5), 485–490. https://doi.org/10.1038/nn841

Hurley, M. M., Dennett, D. C., Adams Jr, R. B., & Adams, R. B. (2011). *Inside jokes: Using humor to reverse-engineer the mind.* MIT press.

Huys, Q. J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., Dayan, P., & Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences, 112*(10), 3098–3103. https://doi.org/10.1073/pnas.1414219112

Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system (N. Uchida, Ed.). *eLife, 5*, e18073. https://doi.org/10.7554/eLife.18073

Illing, B., Ventura, J., Bellec, G., & Gerstner, W. (2021). Local plasticity rules can learn deep representations using self-supervised contrastive predictions. *arXiv preprint arXiv:2010.08262*.

Imada, T., Hari, R., Loveless, N., McEvoy, L., & Sams, M. (1993). Determinants of the auditory mismatch response. *Electroencephalography and Clinical Neurophysiology, 87*(3), 144–153. https://doi.org/10.1016/0013-4694(93)90120-K

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association, 90*(430), 773–795.

Knill, D. C., & Pouget, A. (2004). The bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Kobayashi, K., Ravaioli, S., Baranes, A., Woodford, M., & Gottlieb, J. (2019). Diverse motives for human curiosity. *Nature human behaviour, 3*, 587–595. https://doi.org/10.1038/s41562-019-0589-3

Kolossa, A., Kopp, B., & Fingscheidt, T. (2015). A computational analysis of the neural bases of bayesian inference. *NeuroImage, 106*, 222–237. https://doi.org/10.1016/j.neuroimage.2014.11.007

Konovalov, A., & Krajbich, I. (2018). Neurocomputational dynamics of sequence learning. *Neuron, 98*(6), 1282–1293.e4. https://doi.org/10.1016/j.neuron.2018.05.013

Kopp, B., & Lange, F. (2013). Electrophysiological indicators of surprise and entropy in dynamic task-switching environments. *Frontiers in Human Neuroscience, 7*, 300. https://doi.org/10.3389/fnhum.2013.00300

Kounios, J., & Beeman, M. (2009). The aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science, 18*(4), 210–216. https://doi.org/10.1111/j.1467-8721.2009.01638.x

Kuhn, T. S. (1962). *The structure of scientific revolutions.* University of Chicago press.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The p300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792–795. https://doi.org/10.1126/science.887923

Lehmann, M. P., Xu, H. A., Liakoni, V., Herzog, M. H., Gerstner, W., & Preuschoff, K. (2019). One-shot learning and behavioral eligibility traces in sequential decision making. *Elife, 8*, e47463. https://doi.org/10.7554/eLife.47463

Liakoni, V., Modirshanechi, A., Gerstner, W., & Brea, J. (2021). Learning in volatile environments with the bayes factor surprise. *Neural Computation, 33*(2), 1–72. https://doi.org/10.1162/neco_a_01352

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

Little, D. Y.-J., & Sommer, F. T. (2013). Learning and exploration in action-perception loops. *Frontiers in Neural Circuits, 7*, 37. https://doi.org/10.3389/fncir.2013.00037
Loued-Khenissi, L., & Preuschoff, K. (2020). Information theoretic characterization of uncertainty distinguishes surprise from accuracy signals in the brain. *Frontiers in Artificial Intelligence, 3, 5*. https://doi.org/10.3389/frai.2020.00005

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Magliero, A., Bashore, T. R., Coles, M. G., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21*(2), 171–186. https://doi.org/10.1111/j.1469-8986.1984.tb00201.x

Maheu, M., Dehaene, S., & Meyniel, F. (2019). Brain signatures of a multiscale process of sequence learning in humans. *Elife, 8*, e41541. https://doi.org/10.7554/eLife.41541

Markovic, D., Stojic, H., Schoebel, S., & Kiebel, S. J. (2021). An empirical evaluation of active inference in multi-armed bandits. *Neural Networks, 144*, 229–246. https://doi.org/10.1016/j.neunet.2021.08.018

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt; Co., Inc.

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *Journal of Neuroscience, 28*(47), 12539–12545. https://doi.org/10.1523/JNEUROSCI.2925-08.2008

Marzecová, A., Van den Bussche, E., & Verguts, T. (2019). Attentional orienting relies on bayesian estimates of expected and unexpected uncertainty. 2019 Conference on Cognitive Computational Neuroscience. https://doi.org/10.32470/CCN.2019.1203-0

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

Meyniel, F. (2020). Brain dynamics for confidence-weighted learning. *PLOS Computational Biology, 16*, 1–27. https://doi.org/10.1371/journal.pcbi.1007935

Meyniel, F., Maheu, M., & Dehaene, S. (2016). Human inferences about sequences: A minimal transition probability model. *PLOS Computational Biology, 12*, 1–26. https://doi.org/10.1371/journal.pcbi.1005260

Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science, 160*(3827), 554–555. https://doi.org/10.1126/science.160.3827.554

Mobin, S. A., Arnemann, J. A., & Sommer, F. (2014). Information-based learning by agents in unbounded state spaces. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc.

Modirshanechi, A., Kiani, M. M., & Aghajan, H. (2019). Trial-by-trial surprise-decoding model for visual and auditory binary oddball tasks. *NeuroImage, 196*, 302–317. https://doi.org/10.1016/j.neuroimage.2019.04.028

Mousavi, Z., Kiani, M. M., & Aghajan, H. (2020). Brain signatures of surprise in eeg and meg data. *bioRxiv*. https://doi.org/10.1101/2020.01.06.985664

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). *Clinical Neurophysiology, 118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Nassar, M. R., & Frank, M. J. (2016). Taming the beast: Extracting generalizable knowledge from computational models of cognition. *Current opinion in behavioral sciences, 11*, 49–54. https://doi.org/10.1016/j.cobeha.2016.04.003

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience, 15*(7), 1040–1046. https://doi.org/10.1038/nn.3130
Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience, 30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Neath, A. A., & Cavanaugh, J. E. (2012). The bayesian information criterion: Background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics, 4*(2), 199–203. https://doi.org/10.1002/wics.199

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience, 35*(21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T. T., Kiebel, S. J., & Blankenburg, F. (2012). Evidence for neural encoding of bayesian surprise in human somatosensation. *NeuroImage, 62*(1), 177–188. https://doi.org/10.1016/j.neuroimage.2012.04.050

Oudeyer, P.-Y., & Kaplan, F. (2009). What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurorobotics, 1*, 6. https://doi.org/10.3389/neuro.12.006.2007

Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016). Chapter 11 - intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In B. Studer & S. Knecht (Eds.), *Motivation* (pp. 257–284). Elsevier. https://doi.org/10.1016/bs.pbr.2016.05.005

Oxford English Dictionary. (n.d.). “surprise, n.”. https://www.oed.com/view/Entry/194999

Palm, G. (2012). *Novelty, information and surprise*. Springer Science & Business Media.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, 2778–2787.

Prat-Carrabin, A., Wilson, R. C., Cohen, J. D., & Da Silveira, R. A. (2021). Human inference in changing environments with temporal structure. *arXiv preprint arXiv:2101.11143*.

Preuschoff, K., t Hart, B. M., & Einhauser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline’s role in decision making. *Frontiers in Neuroscience, 5*, 115. https://doi.org/10.3389/fnins.2011.00115

Robbins, K., Su, K.-m., & Hairston, W. D. (2018). An 18-subject eeg data collection using a visual-oddball task, designed for benchmarking algorithms and headset performance comparisons. *Data in Brief, 16*, 227–230. https://doi.org/10.1016/j.dib.2017.11.032

Rouhani, N., & Niv, Y. (2021). Signed and unsigned reward prediction errors dynamically enhance learning and memory. (T. Kahn & C. Büchel, Eds.). *eLife, 10*, e61077. https://doi.org/10.7554/eLife.61077

Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(9), 1430–1443. https://doi.org/10.1037/xlm0000518

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition, 203*, 104269. https://doi.org/10.1016/j.cognition.2020.104269

Rubin, J., Ulansovsky, N., Nelken, I., & Tishby, N. (2016). The representation of prediction error in auditory cortex. *PLoS computational biology, 12*(8), e1005058. https://doi.org/10.1371/journal.pcbi.1005058

Ryali, C., Reddy, G., & Yu, A. J. (2018). Demystifying excessively volatile human learning: A bayesian persistent prior and a neural approximation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc.

Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. (2021). Active Inference: Demystified and Compared. *Neural Computation, 33*(3), 674–712. https://doi.org/10.1162/neco_a_01357

SanMiguel, I., Costa-Faidella, J., Lugo, Z. R., Vilella, E., & Escera, C. (2021). Standard tone stability as a manipulation of precision in the oddball paradigm: Modulation of prediction error responses to fixed-probability deviants. *Frontiers in Human Neuroscience, 15*, 577. https://doi.org/10.3389/fnhum.2021.734200
Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature reviews neuroscience, 10*(3), 211–223. https://doi.org/10.1038/nrn2573

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development, 2*(3), 230–247. https://doi.org/10.1109/TAMD.2010.2056368

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Frontiers in Psychology, 4*, 710. https://doi.org/10.3389/fpsyg.2013.00710

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning & Memory, 25*(8), 369–381. https://doi.org/10.1101/lm.046912.117

Singh, S., Lewis, R. L., & Barto, A. G. (2010). Where do rewards come from? *Proceedings of the annual conference of the cognitive science society, 2601–2606.*

Singh, S., Lewis, R. L., Barto, A. G., & Sorg, J. (2010). Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Transactions on Autonomous Mental Development, 2*(2), 70–82. https://doi.org/10.1109/TAMD.2010.2051031

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience, 20*(10), 635–644. https://doi.org/10.1038/s41583-019-0180-y

Soltani, A., & Koechlin, E. (2021). Computational models of adaptive behavior and prefrontal cortex. *Neuropsychopharmacology.* https://doi.org/10.1038/s41386-021-01123-1

Soltani, A., & Wang, X.-J. (2010). Synaptic computation underlying probabilistic inference. *Nature neuroscience, 13*(1), 112–119. https://doi.org/10.1038/nn.2450

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science, 193*(4258), 1142–1146. https://doi.org/10.1126/science.959831

Storck, J., Hochreiter, S., & Schmidhuber, J. (1995). Reinforcement driven information acquisition in non-deterministic environments. *Proceedings of the international conference on artificial neural networks, Paris, 2, 159–164.*

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

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience, 33*(19), 8264–8269. https://doi.org/10.1523/JNEUROSCI.5695-12.2013

Tribus, M. (1961). *Thermostatics and thermodynamics: An introduction to energy, information and states of matter, with engineering applications.* D. Van Nostrand.

Visalli, A., Capizzi, M., Ambrosini, E., Kopp, B., & Vallesi, A. (2021). Electroencephalographic correlates of temporal bayesian belief updating and surprise. *NeuroImage, 231*, 117867. https://doi.org/10.1016/j.neuroimage.2021.117867

Walz, J. M., Goldman, R. I., Carapezza, M., Muraskin, J., Brown, T. R., & Sajda, P. (2015). Prestimulus eeg alpha oscillations modulate task-related fmri bold responses to auditory stimuli. *NeuroImage, 113*, 153–163. https://doi.org/10.1016/j.neuroimage.2015.03.028

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife, 8*, e49547. https://doi.org/10.7554/eLife.49547
Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS computational biology, 9*(7), e1003150. https://doi.org/10.1371/journal.pcbi.1003150

Xu, H. A., Modirshanechi, A., Lehmann, M. P., Gerstner, W., & Herzog, M. H. (2021). Novelty is not surprise: Human exploratory and adaptive behavior in sequential decision-making. *PLOS Computational Biology, 17*(6). https://doi.org/10.1371/journal.pcbi.1009070

Yu, A. J., & Cohen, J. D. (2009). Sequential effects: Superstition or rational behavior? In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026
Appendix A: Special cases and links to related works

Several existing models and experimental paradigms are special cases of Eq. 1 and Eq. 2. The standard generative models for studying passive learning in volatile environments (Adams & MacKay, 2007; Fearnhead & Liu, 2007; Liakoni et al., 2021; Nassar et al., 2012; Nassar et al., 2010; Wilson et al., 2013) is obtained if we remove the cue variables $X_{1:t}$ (Fig. 1B). For example, in the Gaussian experiment of Nassar et al., 2012; Nassar et al., 2010, $Y_t$ is a sample from a Gaussian distribution with a mean equal to $\Theta_t$ and a known variance, and $\pi^{(0)}$ is a very broad uniform distribution. Liakoni et al., 2021 extensively studied this generative model and showed that both exact and approximate Bayesian inference in this generative model lead to surprise-modulated learning rules. In the next sections, we show that the same holds true for our more general generative model.

Variants of bandit and reversal bandit tasks (Behrens et al., 2007; Findling et al., 2021; Horvath et al., 2021) can be modeled by considering the cue variables $X_{1:t}$ as actions $A_{1:t}$ (Fig. 1C). For example, in the experiment of Behrens et al., 2007, $X_t = A_t$ is one of the two possible actions that participants can choose, $Y_t$ is the indicator of whether they are rewarded or not, and $\Theta_t$ indicates which action is rewarded with higher probability. In this setting, $\mathbb{P}(x_t | x_{t-1}, y_{t-1}) = \mathbb{P}(x_t)$ is the probability that participants take action $x_t$, independently of the dynamics of the environment.

The minimal model of human inferences about binary sequences of Meyniel et al., 2016 (Fig. 1D) is obtained if the cue variable $X_t$ is equal to the previous observation $Y_{t-1}$. There, $Y_t$, conditioned on $Y_{t-1}$, is a sample from a Bernoulli distribution with parameter $\Theta_t$. In this setting, we have $\mathbb{P}(x_t | x_{t-1}, y_{t-1}) = \delta_{x_t,y_{t-1}}$. This class of generative models have also been successfully used to explain the neural signatures of surprise via encoding (Gijsen et al., 2021; Maheu et al., 2019; Meyniel, 2020) and decoding (Modirshanechi et al., 2019; Mousavi et al., 2020) models.

Classic Markov Decision Processes (MDPs) (Sutton & Barto, 2018) can also be written in the form of our generative model. To reduce our generative model to an MDP, we set $p_c = 0$, consider the observation $Y_t$ as the pair of the current state and immediate reward value, and consider the cue variable $X_t$ as the previous pair of action and observation (or state) $(A_{t-1}, Y_{t-1})$ (Fig. 1E). In this setting, we have $\mathbb{P}(X_t = (a_{t-1}, y) | x_{t-1}, y_{t-1}) = \delta_{y,y_{t-1}} \mathbb{P}(a_{t-1} | y_{t-1})$, where $\mathbb{P}(a_{t-1} | y_{t-1})$ is called the action selection policy in Reinforcement Learning theory (Sutton & Barto, 2018) and is independent of the dynamics of the environment. The theory of Reinforcement Learning for MDPs has been frequently used in neuroscience and psychology to model human reward-driven decision-making (Daw et al., 2011; Gläscher et al., 2010; Huys et al., 2015; Lehmann et al., 2019; Niv et al., 2015; Xu et al., 2021).

Appendix B: Proofs

In this appendix, we provide proofs for our Propositions and Corollaries mentioned in section Theories of surprise: A technical review in the main text. We also provide further results for the Bayesian and postdictive surprise in Lemma 1 and 2 and Remark 1.

Proof of Proposition 1

The proof is in essence the same as the proof of Proposition 1 of Liakoni et al., 2021. We write

$$\pi^{(t+1)}(\theta) = \mathbb{P}^{(t+1)}(\Theta_{t+1} = \theta)$$

$$= \mathbb{P}^{(t+1)}(\Theta_{t+1} = \theta | C_{t+1} = 0) \mathbb{P}^{(t+1)}(C_{t+1} = 0) + \mathbb{P}^{(t+1)}(\Theta_{t+1} = \theta | C_{t+1} = 1) \mathbb{P}^{(t+1)}(C_{t+1} = 1).$$

(55)
We use Bayes’ rule and write \( P^{(t+1)}(\Theta_{t+1} = \theta|C_{t+1} = 0) \) (c.f. the 1st term in Eq. 55) as
\[
P^{(t+1)}(\Theta_{t+1} = \theta|C_{t+1} = 0) = \frac{P^{(t)}(\Theta_{t+1} = \theta|C_{t+1} = 0, x_{t+1}, y_{t+1})}{P^{(t)}(y_{t+1}|C_{t+1} = 0, x_{t+1})}
\]
\[
= \frac{P^{(t)}(y_{t+1}|C_{t+1} = 0, x_{t+1}, \Theta_{t+1} = \theta)P^{(t)}(\Theta_{t+1} = \theta|C_{t+1} = 0, x_{t+1})}{P^{(t)}(y_{t+1}|C_{t+1} = 0, x_{t+1})}
\]
\[
= \frac{P_{Y}|X(y_{t+1}|x_{t+1}; \theta)\pi^{(t)}(\theta)}{P(y_{t+1}|x_{t+1}; \pi^{(t)})} = \gamma^{(t+1)}_{\text{integration}}(\theta),
\]
and similarly
\[
P^{(t+1)}(\Theta_{t+1} = \theta|C_{t+1} = 1) = \frac{P_{Y}|X(y_{t+1}|x_{t+1}; \theta)\pi^{(0)}(\theta)}{P(y_{t+1}|x_{t+1}; \pi^{(0)})} = \gamma^{(t+1)}_{\text{reset}}(\theta).
\]
Then, for \( P^{(t+1)}(C_{t+1} = 1) \) and \( P^{(t+1)}(C_{t+1} = 0) \) we have
\[
P^{(t+1)}(C_{t+1} = 1) = P^{(t)}(C_{t+1} = 1|x_{t+1}, y_{t+1})
\]
\[
= \frac{p_c P^{(t)}(y_{t+1}|x_{t+1}, C_{t+1} = 1)}{(1 - p_c)P^{(t)}(y_{t+1}|x_{t+1}, C_{t+1} = 0) + p_c P^{(t)}(y_{t+1}|x_{t+1}, C_{t+1} = 1)}
\]
\[
= \frac{m S_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)})}{1 + m S_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)})} = \gamma_{t+1}
\]
with \( m = \frac{p_c}{1 - p_c} \). Therefore, the proof is complete by substituting these terms in Eq. 55.

**Proof of Proposition 2**

Based on the definition of the adaptation rate \( \gamma_{t+1} \) (c.f. Proposition 1), we have
\[
S_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \frac{1 - p_c}{p_c} \frac{\gamma_{t+1}}{1 - \gamma_{t+1}}
\]
For the difference in the 1st definition of the Shannon surprise (c.f. Eq. 9), we can write
\[
\Delta S_{\text{Sh1}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Sh1}}(y_{t+1}|x_{t+1}; \pi^{(t)}) - S_{\text{Sh1}}(y_{t+1}|x_{t+1}; \pi^{(0)})
\]
\[
= \log \left( \frac{P(y_{t+1}|x_{t+1}; \pi^{(0)})}{p_c P(y_{t+1}|x_{t+1}; \pi^{(0)})} \right) = \log \frac{\gamma_{t+1}}{p_c}.
\]
As a result, we have \( \gamma_{t+1} = p_c \exp \Delta S_{\text{Sh1}}(y_{t+1}|x_{t+1}; \pi^{(t)}) \) and hence
\[
S_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \frac{(1 - p_c) \exp \Delta S_{\text{Sh1}}(y_{t+1}|x_{t+1}; \pi^{(t)})}{1 - p_c \exp \Delta S_{\text{Sh1}}(y_{t+1}|x_{t+1}; \pi^{(t)})}.
\]
The proof is more straightforward for the difference in the 2nd definition (c.f. Eq. 10) where we have
\[
\Delta S_{\text{Sh2}}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{\text{Sh2}}(y_{t+1}|x_{t+1}; \pi^{(t)}) - S_{\text{Sh2}}(y_{t+1}|x_{t+1}; \pi^{(0)})
\]
\[
= \log \left( \frac{P(y_{t+1}|x_{t+1}; \pi^{(0)})}{P(y_{t+1}|x_{t+1}; \pi^{(t)})} \right) = \log S_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)}).
\]
Therefore, the proof is complete.
Proof of Proposition 3

Based on the definitions of the two versions of the Shannon surprise (c.f. Eq. 9 and Eq. 10), we have
\[
\mathbb{P}^{(t)}(y_{t+1}|x_{t+1}) = \exp\left(-S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)})\right),
\]
\[
P(y_{t+1}|x_{t+1}; \pi^{(t)}) = \exp\left(-S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)})\right).
\]
(63)

The proof is complete by using these equations and replacing the probabilities in Eq. 14 and Eq. 15. ■

Proof of Proposition 4

For a categorical task with \(N\) categories and one-hot coded observations, we have (c.f. Eq. 17 and Eq. 18)
\[
E_1[Y_{t+1}] = \left[p_i P(n|x_{t+1}; \pi^{(0)}) + (1 - p_i) P(n|x_{t+1}; \pi^{(t)})\right]_{n=1}^N
\]
\[
E_2[Y_{t+1}] = \left[P(n|x_{t+1}; \pi^{(t)})\right]_{n=1}^N
\]
(64)

where \(z = [z_n]_{n=1}^N\) is an \(N\)-dimensional vector with \(z_n\) the \(n\)th element. To be able to prove the proposition for \(E_1[Y_{t+1}]\) and \(E_2[Y_{t+1}]\) simultaneously, we define \(E_i[Y_{t+1}] = [p_i, n]_{n=1}^N\), where \(p_1, n = p_i P(n|x_{t+1}; \pi^{(0)}) + (1 - p_i) P(n|x_{t+1}; \pi^{(t)})\) and \(p_2, n = P(n|x_{t+1}; \pi^{(t)})\).

We show the one-hot coded vector corresponding to category \(m \in \{1, ..., N\}\) by \(e_m\). For the absolute error surprise, we have (c.f. Eq. 19)
\[
S_{\text{Abs}}(y_{t+1} = e_m|x_{t+1}; \pi^{(t)}) = \sum_{n=1}^N |\delta_{m,n} - p_{i,n}| = |1 - p_{i,m}| + \sum_{n=1, n \neq m}^N p_{i,n} = 2(1 - p_{i,m}),
\]
(65)

which is the same as \(2S_{\text{SPEi}}(y_{t+1} = e_m|x_{t+1}; \pi^{(t)})\) (c.f. Eq. 14 and Eq. 15).

For the squared error surprise, we have (c.f. Eq. 19)
\[
S_{\text{Sqi}}(y_{t+1} = e_m|x_{t+1}; \pi^{(t)}) = \sum_{n=1}^N (\delta_{m,n} - p_{i,n})^2 = (1 - p_{i,m})^2 + \sum_{n=1, n \neq m}^N p_{i,n}^2
\]
\[
= 2(1 - p_{i,m}) + ||p_{i,:}||^2_2 - 1,
\]
(66)

where we have \(2(1 - p_{i,m}) = 2S_{\text{SPEi}}(y_{t+1} = e_m|x_{t+1}; \pi^{(t)})\) and
\[
\text{Conf.}[P(.|x_{t+1}; \pi^{(t)})] = ||p_{i,:}||^2_2 - 1
\]
(67)

shows the \(\ell_2\)-norm of the estimate vector \(p_{i,:}\) as a measure of confidence; \(||p_{i,:}||^2_2\) takes its maximum value when the prediction has a probability of 1 for one category and zero for the rest, and \(||p_{i,:}||^2_2\) takes its minimum when it is distributed uniformly over all categories. Therefore, the proof is complete. ■

Proof of Proposition 5

Assume that \(Y_{t+1} \in \mathbb{R}^N\) given the cue \(x_{t+1}\) and the belief \(\pi^{(t)}\) has a Gaussian distribution with a covariance matrix \(\sigma^2 I\), i.e.,
\[
P(y_{t+1}|x_{t+1}; \pi^{(t)}) = \mathcal{N}(y_{t+1}; E_2[Y_{t+1}], \sigma I).
\]
(68)
We then have
\[ S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = -\log \mathcal{N}(y_{t+1}; E_2[Y_{t+1}], \sigma I) \]
\[ = - \frac{N}{2} \log (2\pi\sigma) + \frac{||y_{t+1} - E_2[Y_{t+1}]||^2}{2\sigma^2} \]
\[ = a + bS_{\text{Sh}2}(y_{t+1}) = e_m|x_{t+1}; \pi^{(t)}), \]
where \( a = -N \log (2\pi\sigma)/2 \) and \( b = 1/(2\sigma^2) \). Therefore, the proof is complete.

Theoretical results for the Bayesian surprise

**Lemma 1.** (Relation between the Bayesian surprise and the Shannon surprise) In the generative model of Definition 1, the Bayesian surprise can be written as
\[ S_{\text{Ba}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = p_cE_{\pi^{(t)}} \left[ S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \delta \theta) \right] + (1 - p_c)E_{\pi^{(t)}} \left[ S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \delta \theta) \right] \]
\[ - S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)}), \]
and
\[ S_{\text{Ba}2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = E_{\pi^{(t)}} \left[ S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \delta \theta) \right] - S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) \]
\[ + D_{KL} \left[ \pi^{(t)}||p_c\pi^{(0)} + (1 - p_c)\pi^{(t)} \right], \]
where \( \delta \theta \) is a delta-distribution at \( \theta \).

**Proof:** For the 1st definition of the Bayesian surprise \( \text{(c.f. Eq. 23)} \), we have
\[ S_{\text{Ba}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL} \left[ \mathbb{P}_{\Theta_{t+1}}^{(t+1)} || \mathbb{P}_{\Theta_{t+1}}^{(t+1)} \right] = E_{\pi^{(t)}} \left[ \log \frac{\mathbb{P}^{(t)}(\Theta_{t+1})}{\mathbb{P}^{(t+1)}(\Theta_{t+1})} \right]. \]

We know
\[ \mathbb{P}_{\Theta^{(t)}}^{(t+1)} = p_c\pi^{(0)} + (1 - p_c)\pi^{(t)}, \]
and
\[ \mathbb{P}_{\Theta^{(t+1)}}^{(t+1)}(\Theta_{t+1}) = \frac{\mathbb{P}^{(t)}(\Theta_{t+1})P_{Y|X}(y_{t+1}|x_{t+1}; \Theta_{t+1})}{\mathbb{P}^{(t)}(y_{t+1}|x_{t+1})} \Rightarrow \frac{\mathbb{P}^{(t+1)}(\Theta_{t+1})}{\mathbb{P}^{(t)}(\Theta_{t+1})} = \frac{P_{Y|X}(y_{t+1}|x_{t+1}; \Theta_{t+1})}{\mathbb{P}^{(t)}(y_{t+1}|x_{t+1})}. \]

We, therefore, have
\[ S_{\text{Ba}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = - p_cE_{\pi^{(t)}} \left[ \log P_{Y|X}(y_{t+1}|x_{t+1}; \Theta) \right] \]
\[ - (1 - p_c)E_{\pi^{(t)}} \left[ \log P_{Y|X}(y_{t+1}|x_{t+1}; \Theta) \right] + \log \mathbb{P}^{(t)}(y_{t+1}|x_{t+1}), \]
which is equivalent to \( \text{(c.f. Eq. 9 and Eq. 10)} \)
\[ S_{\text{Ba}1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = p_cE_{\pi^{(t)}} \left[ S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \delta \theta) \right] + (1 - p_c)E_{\pi^{(t)}} \left[ S_{\text{Sh}2}(y_{t+1}|x_{t+1}; \delta \theta) \right] \]
\[ - S_{\text{Sh}1}(y_{t+1}|x_{t+1}; \pi^{(t)}). \]

For the 2nd definition of the Bayesian surprise \( \text{(c.f. Eq. 24)} \), we have
\[ S_{\text{Ba}2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL} \left[ \pi^{(t)}||\pi^{(t+1)} \right] = E_{\pi^{(t)}} \left[ \log \frac{\pi^{(t)}(\Theta)}{\pi^{(t+1)}(\Theta)} \right]. \]
We use Eq. 22 and Eq. 74 and write
\[
S_{Ba2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = -\mathbb{E}_{\pi^{(t)}}\left[ \log P_{Y|X}(y_{t+1}|x_{t+1}; \Theta) \right] \\
+ \log \left[ \mathbb{P}^{(t)}(y_{t+1}|x_{t+1}) \right] \\
+ \mathbb{E}_{\pi^{(t)}}\left[ \log \frac{\pi^{(t)}(\Theta)}{p_c \pi_0^{(0)}(\Theta) + (1 - p_c) \pi^{(t)}(\Theta)} \right],
\]

which is equivalent to (c.f. Eq. 9 and Eq. 10)
\[
S_{Ba2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \mathbb{E}_{\pi^{(t)}}\left[ S_{Sh2}(y_{t+1}|x_{t+1}; \delta_\Theta) \right] - S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)}) \\
+ D_{KL}\left[ \pi^{(t)}||p_c \pi_0^{(0)} + (1 - p_c) \pi^{(t)} \right].
\]

Therefore, the proof is complete.

**Remark 1.** When the change point probability is zero, i.e. \( p_c = 0 \), the Bayesian surprise is equal to the expected Shannon surprise minus the Shannon surprise, i.e.,
\[
S_{Ba}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \mathbb{E}_{\pi^{(t)}}\left[ S_{Sh}(y_{t+1}|x_{t+1}; \delta_\Theta) \right] - S_{Sh}(y_{t+1}|x_{t+1}; \pi^{(t)}),
\]
where \( S_{Ba} = S_{Ba1} = S_{Ba2} \) and \( S_{Sh} = S_{Sh1} = S_{Sh2} \).

**Proof:** The remark is the direct consequence of Lemma 1.

**Theoretical results for the postdictive surprise**

**Lemma 2.** (Relation between the postdictive surprise and the Shannon surprise) In the generative model of Definition 1, the postdictive surprise can be written as
\[
S_{Po1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \mathbb{E}_P\left( |x_{t+1}; \mathbb{P}^{(t)}_{\Theta_{t+1}} \right) \left[ S_{Sh2}(y_{t+1}|x_{t+1}; \mathbb{P}(\Theta_{t+1}|Y_{t+1} = Y, x_{t+1})) \right] \\
- S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)})
\]

and
\[
S_{Po2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \mathbb{E}_P\left( |x_{t+1}; \pi^{(t)} \right) \left[ S_{Sh2}(y_{t+1}|x_{t+1}; \mathbb{P}(\Theta_{t+1}|Y_{t+1} = Y, x_{t+1})) \right] \\
- S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)}) \\
+ D_{KL}\left[ P(.|x_{t+1}; \pi^{(t)}) || P(.|x_{t+1}; \mathbb{P}^{(t)}_{\Theta_{t+1}}) \right],
\]

**Proof:** We first prove the equality for \( S_{Po1} \) for which we have (c.f. Eq. 25)
\[
S_{Po1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL}\left[ P(.|x_{t+1}; \mathbb{P}^{(t)}_{\Theta_{t+1}}) || P(.|x_{t+1}; \pi^{(t+1)}) \right] \\
= \mathbb{E}_P\left( |x_{t+1}; \mathbb{P}^{(t)}_{\Theta_{t+1}} \right) \left[ \log \frac{P(Y|x_{t+1}; \mathbb{P}^{(t)}_{\Theta_{t+1}})}{P(Y|x_{t+1}; \pi^{(t+1)})} \right],
\]
where
\[
P(y|x_{t+1}; \mathbb{P}^{(t)}_{\Theta_{t+1}}) = \int P_Y|X(y|x_{t+1}; \theta) \mathbb{P}^{(t)}(\Theta_{t+1} = \theta) d\theta,
\]
and, using Bayes’ rule,

\[ P(y|x_{t+1}; \pi^{(t+1)}) = \int P_{Y|X}(y|x_{t+1}; \theta) \pi^{(t+1)}(\theta) d\theta \]

\[ = \int P_{Y|X}(y|x_{t+1}; \theta) \frac{P^{(t)}(\Theta_{t+1} = \theta) P_{Y|X}(y|x_{t+1}; \theta)}{P(y_{t+1}|x_{t+1}; P^{(t)}_{\Theta_{t+1}})} d\theta \]  \hspace{1cm} (85)

Using the Bayes’ rule and the definition of the marginal probability (c.f. Eq. 4), we can find

\[ \frac{P(y|x_{t+1}; \pi^{(t+1)})}{P(y|x_{t+1}; \pi^{(t+1)})} = \frac{1}{P(y_{t+1}|x_{t+1}; P^{(t)}_{\Theta_{t+1}})} \times \int P_{Y|X}(y|x_{t+1}; \theta) \frac{P^{(t)}(\Theta_{t+1} = \theta) P_{Y|X}(y|x_{t+1}; \theta)}{P(y|x_{t+1}; P^{(t)}_{\Theta_{t+1}})} d\theta \]

\[ = \frac{1}{P(y_{t+1}|x_{t+1}; P^{(t)}_{\Theta_{t+1}})} \times \int P_{Y|X}(y|x_{t+1}; \theta) P^{(t)}(\Theta_{t+1} = \theta|Y_{t+1} = y, x_{t+1}) d\theta \]

\[ = \frac{P(y_{t+1}|x_{t+1}; P^{(t)}(\Theta_{t+1}|Y_{t+1} = y, x_{t+1}))}{P(y_{t+1}|x_{t+1}; P^{(t)}_{\Theta_{t+1}})}, \hspace{1cm} (86) \]

and as a result (using Eq. 9 and Eq. 10)

\[ \log \frac{P(y|x_{t+1}; P^{(t)}_{\Theta_{t+1}})}{P(y|x_{t+1}; \pi^{(t+1)})} = -\log P(y_{t+1}|x_{t+1}; P^{(t)}(\Theta_{t+1}|Y_{t+1} = y, x_{t+1})) \]

\[ + \log P(y_{t+1}|x_{t+1}; P^{(t)}_{\Theta_{t+1}}) \]

\[ = S_{Sh2}(y_{t+1}|x_{t+1}; P^{(t)}(\Theta_{t+1}|Y_{t+1} = y, x_{t+1})) - S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)}), \hspace{1cm} (87) \]

which, using Eq. 83, makes the proof complete.

To prove the 2nd equality, we note that (c.f. Eq. 26)

\[ S_{Po2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL}[P(.|x_{t+1}; \pi^{(t)}) || P(.|x_{t+1}; \pi^{(t+1)})] \]

\[ = \mathbb{E}_{P(.|x_{t+1}; \pi^{(t)})} \left[ \log P(Y|x_{t+1}; \pi^{(t)}) \right] \]

\[ - \log P(Y|x_{t+1}; \pi^{(t+1)}) \]

\[ \hspace{1cm} (88) \]

and

\[ \log \frac{P(y|x_{t+1}; \pi^{(t)})}{P(y|x_{t+1}; \pi^{(t+1)})} = \log \frac{P(y|x_{t+1}; P^{(t)}_{\Theta_{t+1}})}{P(y|x_{t+1}; \pi^{(t+1)})} + \log \frac{P(y|x_{t+1}; \pi^{(t)})}{P(y|x_{t+1}; P^{(t)}_{\Theta_{t+1}})}. \hspace{1cm} (89) \]

Therefore, using Eq. 83 and the definition of $D_{KL}$, the proof is complete.

\section*{Proof of Proposition 6}

First, we prove the statement for the 2nd definition of the Confidence Corrected surprise (c.f. Eq. 31) for which we have

\[ S_{CC2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = D_{KL}[\pi^{(t)} || \pi^{(t+1)}_{\text{reset}}] = \mathbb{E}_{\pi^{(t)}} \left[ \log \frac{\pi^{(t)}(\Theta)}{\pi^{(t+1)}_{\text{reset}}(\Theta)} \right]. \hspace{1cm} (90) \]
Using the definition of $\pi_{\text{reset}}^{(t+1)}$ in Proposition 1, we can write

$$S_{CC2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = -E_{\pi(t)} \left[ \log P_{Y|X}(y_{t+1}|x_{t+1}; \Theta) \right]$$

$$+ \log P(y_{t+1}|x_{t+1}; \pi^{(0)})$$

$$+ E_{\pi(t)} \left[ \log \frac{\pi(t)(\Theta)}{\pi^{(0)}(\Theta)} \right],$$

(91)

which is equivalent to (c.f. Eq. 9 and Eq. 10)

$$S_{CC2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \mathbb{E}_{\pi(t)} \left[ S_{Sh2}(y_{t+1}|x_{t+1}; \delta \Theta) \right] - S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(0)})$$

$$+ D_{KL} \left[ \pi^{(t)} || \pi^{(0)} \right].$$

(92)

Now, we can replace $E_{\pi(t)} \left[ S_{Sh2}(y_{t+1}|x_{t+1}; \delta \Theta) \right]$ by using Eq. 79 and have

$$S_{CC2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)}) - S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(0)})$$

$$+ S_{Ba2}(y_{t+1}|x_{t+1}; \pi^{(t)}) - D_{KL} \left[ \pi^{(t)} || p_c \pi^{(0)} + (1-p_c) \pi^{(t)} \right]$$

$$+ D_{KL} \left[ \pi^{(t)} || \pi^{(0)} \right],$$

(93)

which is the same as Eq. 33. For the 1st definition of the Confidence Corrected surprise (c.f. Eq. 28), we can repeat all steps to have

$$S_{CC1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)}) - S_{Sh1}(y_{t+1}|x_{t+1}; \pi_{\text{flat}})$$

$$+ S_{Ba2}(y_{t+1}|x_{t+1}; \pi^{(t)}) - D_{KL} \left[ \pi^{(t)} || p_c \pi^{(0)} + (1-p_c) \pi^{(t)} \right]$$

$$+ D_{KL} \left[ \pi^{(t)} || \pi_{\text{flat}} \right].$$

(94)

If $\pi^{(t)}$ is absolutely continuous with respect to $\pi_{\text{flat}}$, then we have $D_{KL} \left[ \pi^{(t)} || \pi_{\text{flat}} \right] = C \left[ \pi^{(t)} \right] - C \left[ \pi_{\text{flat}} \right],$

which completes the proof.

**Proof of Corollary 1**

The corollary is the direct conclusion of Eq. 60 and Eq. 62.

**Proof of Corollary 2**

Let us show the set of possible observations by $\mathcal{Y}$. We assume that $\mathcal{Y}$ is bounded, i.e., $|\mathcal{Y}| < \infty$. By assumption, we have $P(y_{t+1}|x_{t+1}; \pi^{(0)}) = 1/|\mathcal{Y}|$. We therefore (using Eq. 5, Eq. 9, and Eq. 10) have

$$S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \log \frac{m S_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)})}{1 + m S_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)})} + \log \frac{|\mathcal{Y}|}{p_c},$$

(95)

$$S_{Sh2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = \log S_{BF}(y_{t+1}|x_{t+1}; \pi^{(t)}) + \log |\mathcal{Y}|.$$

Both relations are invertible. Therefore, the proof is complete.

**Proof of Corollary 3**

In the limit of $p_c \to 1$, we have $S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(0)})$ (c.f. Eq. 9) which implies that $\Delta S_{Sh1}(y_{t+1}|x_{t+1}; \pi^{(t)})$ (c.f. Proposition 2) in Eq. 33 is equal to 0. Similarly, in the limit of $p_c \to 1$, we have $D_{KL} \left[ \pi^{(t)} || p_c \pi^{(0)} + (1-p_c) \pi^{(t)} \right] = D_{KL} \left[ \pi^{(t)} || \pi^{(0)} \right]$. Therefore, in the limit of $p_c \to 1$ and given Eq. 33, we have $S_{CC2}(y_{t+1}|x_{t+1}; \pi^{(t)}) = S_{Ba2}(y_{t+1}|x_{t+1}; \pi^{(t)})$. 

56
Appendix C: Methods for case-studies

In this appendix, we provide methods for the analyses of section Different definitions make different predictions: Case studies in the main text.

Methods for case-study 1

The formulas for surprise calculations in Fig. 5-Fig. 8 are given in this section. Note that we always have \( \alpha_i^{(t)} = \alpha_{\text{sum}}^{(t)} P_i^{(t)} \geq 1 \) (since \( \alpha^{(0)} = 1_{N \times 1} \)); this bound determines axes’ boundaries in Fig. 5 and Fig. 6.

**Surprise formulas.** The Shannon surprise (c.f. Eq. 10) in the categorical is

\[
S_{\text{Sh}}(y = i; \pi^{(t)}) = -\log P_i^{(t)}. \tag{96}
\]

To compute the Bayesian and the Confidence Corrected surprise, we need to use the general formula for the KL-divergence between two Dirichlet distributions parameterized by \( \alpha = [\alpha_1, ..., \alpha_N] \) and \( \beta = [\beta_1, ..., \beta_N] \) as \( (\text{Gijsen et al., 2021}) \)

\[
D_{\text{KL}}[\alpha \parallel \beta] = \log \frac{\Gamma(\alpha_{\text{sum}})}{\Gamma(\beta_{\text{sum}})} - \sum_{n=1}^{N} \log \frac{\Gamma(\alpha_n)}{\Gamma(\beta_n)} + \sum_{n=1}^{N} (\alpha_n - \beta_n)(\psi(\alpha_n) - \psi(\alpha_{\text{sum}})), \tag{97}
\]

where we define \( \alpha_{\text{sum}} = \sum_{n=1}^{N} \alpha_n \) and \( \beta_{\text{sum}} = \sum_{n=1}^{N} \beta_n \), \( \Gamma \) is the gamma function, and \( \psi \) is the digamma function \( (\text{Abramowitz & Stegun, 1948}) \).

Using the general formula, the Bayesian surprise (c.f. Eq. 24) can be computed as

\[
S_{\text{Ba}}(y = i; \pi^{(t)}) = \log \frac{\Gamma(\alpha_{\text{sum}})}{\Gamma(\alpha_{\text{sum}} + 1)} - \sum_{n=1}^{N} \log \frac{\Gamma(\alpha_{\text{sum}}^{(t)})}{\Gamma(\alpha_{n}^{(t)} + 1)} \sum_{n=1}^{N} \delta_{i,n} (\psi(\alpha_{n}^{(t)}) - \psi(\alpha_{\text{sum}}^{(t)})) \tag{98}
\]

where we used \( \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha \) \( (\text{Abramowitz & Stegun, 1948}) \) to simplify the expression. We, therefore, have

\[
S_{\text{Ba}}(y = i; \pi^{(t)}) = \log P_i^{(t)} + \psi(\alpha_{\text{sum}}^{(t)}) - \psi(\alpha_{i}^{(t)}). \tag{99}
\]

We can also rewrite Eq. 99 by using the following series representation of the digamma function for any \( x > 0 \) \( (\text{Blagouniche, 2016}) \)

\[
\psi(x) = \log x - \sum_{n=1}^{\infty} \frac{|G_n|(n-1)!}{(x)_n}, \tag{100}
\]

where \( G_n \) are ‘Gregory coefficients’ and \( (x)_n = x(x+1)...(x+n-1) \) is rising factorial. As a result, the Bayesian surprise is

\[
S_{\text{Ba}}(y = i; \pi^{(t)}) = \sum_{n=1}^{\infty} |G_n|(n-1)! \left[ \frac{1}{(\alpha_{\text{sum}} P_i^{(t)})_n} - \frac{1}{(\alpha_{\text{sum}})_n} \right]. \tag{101}
\]

We use this representation in some of our proofs below.

The formula for the Confidence Corrected surprise can be written as

\[
S_{\text{CC}}(y = i; \pi^{(t)}) = \log \frac{\Gamma(\alpha_{\text{sum}}^{(t)})}{\Gamma(N+1)} - \sum_{n=1}^{N} \log \frac{\Gamma(\alpha_{\text{sum}}^{(t)} P_n^{(t)})}{\Gamma(1 + \delta_{i,n})} + \sum_{n=1}^{N} (\alpha_{\text{sum}}^{(t)} P_n^{(t)} - 1 - \delta_{i,n}) \left( \psi(\alpha_{\text{sum}}^{(t)} P_n^{(t)}) - \psi(\alpha_{\text{sum}}^{(t)}) \right). \tag{102}
\]
Influence of $\alpha^{(t)}_{\text{sum}}$. When $\hat{\theta}^{(t)}$ is fixed, then the Shannon surprise (Eq. 96) is constant with respect to $\alpha^{(t)}_{\text{sum}}$. For the Bayesian surprise (Eq. 99), it is straightforward to show, using Eq. 101 and the fact that $\hat{p}_i^{(t)} < 1$, that

$$\frac{\partial S_{\text{Ba}}(y = i; \pi^{(t)})}{\partial \alpha^{(t)}_{\text{sum}}} < 0.$$  

(103)

Therefore, the Bayesian surprise $S_{\text{Ba}}(y = i; \pi^{(t)})$ is always a decreasing function of $\alpha^{(t)}_{\text{sum}}$. Moreover, using asymptotic (for $\alpha^{(t)}_{\text{sum}} \to \infty$) approximation of the digamma function (Abramowitz & Stegun, 1948), we have

$$S_{\text{Ba}}(y = i; \pi^{(t)}) \approx \frac{1}{2\alpha^{(t)}_{\text{sum}}} \frac{1 - \hat{p}_i^{(t)}}{\hat{p}_i^{(t)}},$$  

(104)

which also concludes that $\lim_{\alpha^{(t)}_{\text{sum}} \to \infty} S_{\text{Ba}}(y = i; \pi^{(t)}) = 0$.

We could not make any general statement about the relation of the Confidence Corrected surprise with $\alpha^{(t)}_{\text{sum}}$ (c.f. Fig. 5 and Eq. 102). However, using asymptotic (for $\alpha^{(t)}_{\text{sum}} \to \infty$) approximation of the gamma and digamma functions (Abramowitz & Stegun, 1948), we have

$$S_{\text{CC}}(y = i; \pi^{(t)}) \approx \frac{N - 1}{2} \log \alpha^{(t)}_{\text{sum}} + \text{const.}(\hat{\theta}^{(t)}),$$  

(105)

which concludes that $\lim_{\alpha^{(t)}_{\text{sum}} \to \infty} S_{\text{CC}}(y = i; \pi^{(t)}) = \infty$.

Influence of $\hat{\theta}^{(t)}$. When $\alpha^{(t)}_{\text{sum}}$ is fixed, then the Shannon surprise (Eq. 96) is decreasing with respect to $\hat{p}_i^{(t)}$. According to Eq. 101, the Bayesian surprise is also always a decreasing function of $\hat{p}_i^{(t)}$. Moreover, both Shannon and Bayesian surprise are constant with respect to $\hat{p}_j^{(t)}$ for any $j \neq i$.

We could not make any general statement about the relation of the Confidence Corrected surprise with $\hat{\theta}^{(t)}$ (c.f. Fig. 5, Fig. 6, and Eq. 102).

Simulation details. For Fig. 7A-C, for a given random seed, we sampled $Y_t \sim \text{Cat}([0.9, 0.1])$ for $1 \leq t \leq 100$. To model the temporal dynamics of surprise values, we use the common assumption (Gijsen et al., 2021; Maheu et al., 2019; Meyniel et al., 2016; Modirshanechi et al., 2019; Yu & Cohen, 2009) that participants use a simple leaky integration for the update of their beliefs, i.e.,

$$\alpha_i^{(t+1)} = \kappa \alpha_i^{(t)} + \delta_{i,t+1},$$  

(106)

where $\kappa \in [0, 1]$ is the leak parameter that determines how fast old observations are forgotten, and $\delta$ is the Kronecker delta function. With this update rule, we have the guarantee that $\alpha_i^{(t+1)}$ does not grow unlimited and participants are never too sure about their predictions (as long as $\kappa < 1$). To extract surprise values, we put $\kappa = \exp(-\frac{1}{20})$ equivalent to an integration over a window of approximately 20 observations (Meyniel et al., 2016). However, note that our predictions are independent of the exact value of $\kappa$, and we found the same qualitative behavior of different surprise measures for $\kappa \in [\exp(-\frac{1}{5}), \exp(-\frac{1}{100})]$. For Fig. 7B and Fig. 7C, we first averaged the surprise of $Y_t = 1$ (i.e., standard stimulus) and $Y_t = 2$ (i.e., deviant stimulus) in the early and the late phase of the task separately for each random seed. The values showed in the figure correspond to the mean and the standard error of the mean (over 40 different seeds) of the average-surprise of $Y_t = 1$ (Fig. 7B) and the average-surprise of $Y_t = 2$ (Fig. 7C).

For Fig. 8, for a given random seed, we sampled $Y_t \sim \text{Cat}([0.3, 0.35, 0.35])$ for $1 \leq t \leq 150$ and $Y_t \sim \text{Cat}([0.3, 0.1, 0.9])$ for $151 \leq t \leq 300$. To extract surprise values, we updated the belief at each time using leaky integration (c.f. Eq. 106) with $\kappa = \exp(-\frac{1}{20})$. However, note that our predictions are independent
of the exact value of $\kappa$, and we found the same qualitative behavior of different surprise measures for $\kappa \in [\exp(-\frac{1}{2}), \exp(-\frac{1}{100})]$. For Fig. 8C, we first averaged the surprise of $Y_t = 1$ in the averaging-period of phase 1 ($76 \leq t \leq 150$) and the averaging-period of phase 2 ($226 \leq t \leq 300$) separately for each random seed. The values showed in the figure correspond to the mean and the standard error of the mean (over 40 different seeds) of the average-surprise of $Y_t = 1$.

### Methods for case-study 2

The formulas for surprise calculations and the temporal update rule for the belief for data shown in Fig. 9 and Fig. 10 are given in this section.

**Surprise formulas.** Different measures of surprise for observing $y_{t+1} = 1$ given the cue $x_{t+1} \in [0,1]$ are calculated below. To compute the surprise of $y_{t+1} = 2$ given the cue $x_{t+1} \in [0,1]$, one only needs to replace all $x_{t+1}$ by $1 - x_{t+1}$.

Using Eq. 42, the Bayes Factor surprise (c.f. Eq. 5) and the Shannon surprise (c.f. Eq. 10) can be written as

$$S_{BF}(y_{t+1} = 1|x_{t+1}; b^{(t)}) = \frac{1}{2} + b^{(0)} \left( x_{t+1} - \frac{1}{2} \right),$$

and

$$S_{Sh}(y_{t+1} = 1|x_{t+1}; b^{(t)}) = - \log \left[ \frac{1}{2} + b^{(t)} \left( x_{t+1} - \frac{1}{2} \right) \right].$$

We define $c^{(t)} = p_c b^{(0)} + (1 - p_c) b^{(t)}$ and $D_{KL}[b^{(t)}||c^{(t)}] = b^{(t)} \log \frac{b^{(t)}}{c^{(t)}} + (1 - b^{(t)}) \log \frac{1 - b^{(t)}}{1 - c^{(t)}}$. To compute the Bayesian surprise, we use the result of Lemma 1; to do so, we compute

$$E_{\exp(t)} \left[ S_{Sh2}(y_{t+1} = 1|x_{t+1}; \delta_{0}) \right] = -b^{(t)} \log 2x_{t+1} + \log 2,$$

$$S_{Sh1}(y_{t+1} = 1|x_{t+1}; b^{(t)}) = - \log \left[ \frac{1}{2} + c^{(t)} \left( x_{t+1} - \frac{1}{2} \right) \right].$$

We, therefore, have

$$S_{Ba}(y_{t+1} = 1|x_{t+1}; b^{(t)}) = \log \left[ 1 - c^{(t)} (1 - 2x_{t+1}) \right] - b^{(t)} \log 2x_{t+1} + D_{KL}[b^{(t)}||c^{(t)}].$$

With similar tricks, the Confidence Corrected surprise can be computed as

$$S_{CC}(y_{t+1} = 1|x_{t+1}; b^{(t)}) = \log \left[ 1 - b^{(0)} (1 - 2x_{t+1}) \right] - b^{(t)} \log 2x_{t+1} + D_{KL}[b^{(t)}||b^{(0)}],$$

where we define $D_{KL}[b^{(t)}||b^{(0)}] = b^{(t)} \log \frac{b^{(t)}}{b^{(0)}} + (1 - b^{(t)}) \log \frac{1 - b^{(t)}}{1 - b^{(0)}}$.

**The update rule.** Using Proposition 1 for the update of the belief, we have

$$b^{(t+1)} = \gamma_{t+1} b^{(t+1)}_{\text{reset}} + (1 - \gamma_{t+1}) b^{(t+1)}_{\text{integration}},$$

where $\gamma_{t+1}$ is the adaptation rate as define in Proposition 1, and

$$b^{(t+1)}_{\text{integration}} = 1 - \frac{(1 - b^{(t)}) \frac{1}{2}}{(1 - b^{(t)}) \frac{1}{2} + b^{(t)} x_{t+1}^{2-y_{t+1}} (1 - x_{t+1})^{y_{t+1} - 1}},$$

and similarly

$$b^{(t+1)}_{\text{reset}} = 1 - \frac{(1 - b^{(0)}) \frac{1}{2}}{(1 - b^{(0)}) \frac{1}{2} + b^{(0)} x_{t+1}^{2-y_{t+1}} (1 - x_{t+1})^{y_{t+1} - 1}}.$$
Simulation details. We fixed the parameters \( \theta_{1:150} \) as shown in Fig. 10: \( \theta_{1:50} = \theta_{101:150} = 0 \) and \( \theta_{51:100} = 1 \). Then, for each random seed, we randomly sampled the cue variables \( x_{1:150} \) independently: \( X_t \sim \text{Uniform}\{0.1, 0.9\} \), i.e., at each point, the oracle chooses one of the possible outcomes and assigns a probability of 0.9 to it. We then, given the same random seed, sampled the observations \( y_{1:150} \) as described before: \( Y_t \sim \text{Cat}\{0.5, 0.5\} \) whenever \( \Theta_t = 0 \), and \( Y_t \sim \text{Cat}\{x_t, 1 - x_t\} \) whenever \( \Theta_t = 1 \), where \( \text{Cat} \) stands for categorical distribution. Fig. 10A shows data generated for one random seed, and Fig. 10B shows the average belief and surprise over 500 random seeds.

Methods for case-study 3
The formulas for the gain functions and the details of the update rules for Fig. 11 and Fig. 12 are given in this section.

To simplify the notation, we define, for all \( p \in (0, 1) \), \( q \in (0, 1) \), and \( Y \in \{1, 2\} \)
\[
H(p, q) = -p \log q - (1 - p) \log (1 - q),
\]
and
\[
P_{\text{Band}}(Y, p) = p^2 - Y (1 - p)^{Y - 1}.
\]

The belief and the update rule. Given the prior belief as in Eq. 44, the belief at time \( t \) is
\[
\pi^{(t)}(\theta) = \prod_{i=1}^{N} \text{Dir}\left([p_i, 1 - p_i]; \alpha^{(t)}_{\text{sum}, i}, \hat{p}^{(t)}_i, 1 - \hat{p}^{(t)}_i \right),
\]
where
\[
\alpha^{(t)}_{\text{sum}, i} = \alpha^{(0)}_{\text{sum}, i} + |T^{(t)}_i| \quad \text{and} \quad \hat{p}^{(t)}_i = \frac{\alpha^{(0)}_{\text{sum}, i} \hat{p}^{(0)}_i + |T^{(t)}_i|}{\alpha^{(0)}_{\text{sum}, i} + |T^{(t)}_i|} = \frac{\alpha^{(0)}_{\text{sum}, i} \hat{p}^{(0)}_i + |T^{(t)}_{i, 1}|}{\alpha^{(0)}_{\text{sum}, i}},
\]
where \( T^{(t)}_i = \{\tau \leq t \mid x_\tau = i\} \) is the set of time points when arm \( i \) has been chosen, \( T^{(t)}_{i, 1} = \{\tau \leq t \mid x_\tau = i, y_\tau = 1\} \) is the set of time points when arm \( i \) has been chosen and \( y_\tau = 1 \) has been observed, and \( |T^{(t)}_i| \) and \( |T^{(t)}_{i, 1}| \) are the size of \( T^{(t)}_i \) and \( T^{(t)}_{i, 1} \), respectively. The marginal probability is
\[
P(y = 1 \mid x = i; \pi^{(t)}) = \hat{p}^{(t)}_i.
\]

The optimal gain function. Using Eq. 117 and Eq. 118, we can re-write Eq. 46
\[
\text{MSE}^{(t+1)}_i = \mathbb{E}_{\pi^{(t)}(Y_{t+1}, \Theta_t \mid X_{t+1} = i)} \left[ \sum_{j=1}^{N} (\hat{p}^{(t)}_j - \hat{p}^{(t)}_i)^2 \right] = \frac{\hat{p}^{(t)}_i (1 - \hat{p}^{(t)}_i) \alpha^{(t)}_{\text{sum}, i}}{(\alpha^{(t)}_{\text{sum}, i} + 1)^2} + \sum_{j=1, j \neq i}^{N} \frac{\hat{p}^{(t)}_j (1 - \hat{p}^{(t)}_j)}{\alpha^{(t)}_{\text{sum}, j} + 1}.
\]

Therefore, the optimal strategy is
\[
x_{t+1} = f^*(x_{1:t}, y_{1:t}) = \arg \min_{i \in \{1, \ldots, N\}} \text{MSE}^{(t+1)}_i = \arg \max_{i \in \{1, \ldots, N\}} g^*(\hat{p}^{(t)}_i, \alpha^{(t)}_{\text{sum}, i}),
\]
with the gain function
\[
g^*(\hat{p}^{(t)}_i, \alpha^{(t)}_{\text{sum}, i}) = \frac{\hat{p}^{(t)}_i (1 - \hat{p}^{(t)}_i)}{(\alpha^{(t)}_{\text{sum}, i} + 1)^2}.
\]

\( g^* \) is independent of \( \hat{p}^{(0)}_i \) and \( \alpha^{(0)}_{\text{sum}, i} \), and its behavior with respect to the other two variables is as follows:
• With respect to $\hat{\alpha}_i^{(t)}$: for any fixed $\alpha_{\text{sum},i}^{(t)}$, we have $g^*(\hat{\alpha}_i^{(t)}, \alpha_{\text{sum},i}^{(t)}) \propto \hat{\alpha}_i^{(t)}(1-\hat{\alpha}_i^{(t)}) = 0.25 - (\hat{\alpha}_i^{(t)} - 0.5)^2$

which has an inverted-U-shape, is an even function with respect to $\hat{\alpha}_i^{(t)} - 0.5$, and has its maximum at $\hat{\alpha}_i^{(t)} = 0.5$. Moreover, consider a combination of $\hat{\alpha}_i^{(t)} \in (0, 1)$, $\delta p \in (-1, 1)$, and $\alpha_{\text{sum},i}^{(t)} \geq 1$

for which we have $\hat{\alpha}_i^{(t)} + \delta p \in (0, 1)$ and $g^*(\hat{\alpha}_i^{(t)}, \alpha_{\text{sum},i}^{(t)}) < g^*(\hat{\alpha}_i^{(t)} + \delta p, \alpha_{\text{sum},i}^{(t)})$. Then, there always exists a $\delta \alpha > 0$ such that $g^*(\hat{\alpha}_i^{(t)}, \alpha_{\text{sum},i}^{(t)}) > g^*(\hat{\alpha}_i^{(t)} + \delta p, \alpha_{\text{sum},i}^{(t)} + \delta \alpha)$. Therefore, independent of the difference in the stochasticity level (i.e. $\delta p$), as an arm gets to be chosen more often (as $\delta \alpha$ increases) it eventually becomes less informative.

• With respect to $\alpha_{\text{sum},i}^{(t)}$: $g^*$ is always decreasing.

**Gain function for the Bayes Factor surprise:** For the Bayes Factor surprise, following Eq. 5 and Eq. 50, we have

$$x_{t+1} = f_{\text{BF}}(x_{1:t}, y_{1:t}) = \arg \max_{i \in \{1,...,N\}} -D_{\text{KL}}\left[P(.|x_{t+1} = i; \pi(t))||P(.|x_{t+1} = i; \pi(0))\right]$$

$$= \arg \max_{i \in \{1,...,N\}} g_{BF}(\hat{\alpha}_i^{(t)}, \hat{\alpha}_i^{(0)})$$

with the corresponding gain function (c.f. Eq. 115)

$$g_{BF}(\hat{\alpha}_i^{(t)}, \hat{\alpha}_i^{(0)}) = H\left[\hat{\alpha}_i^{(t)}, \hat{\alpha}_i^{(0)}\right] - H\left[\alpha_i^{(t)}, \alpha_i^{(0)}\right].$$

$g_{BF}$ is independent of $\alpha_{\text{sum},i}^{(0)}$ and $\alpha_{\text{sum},i}^{(t)}$, and its behavior with respect to the other two variables is as follows:

• With respect to $\hat{\alpha}_i^{(t)}$: it has an inverted-U-shape with its maximum at $\hat{\alpha}_i^{(t)} = \hat{\alpha}_i^{(0)}$.

• With respect to $\hat{\alpha}_i^{(0)}$: it has an inverted-U-shape with its maximum at $\hat{\alpha}_i^{(0)} = \hat{\alpha}_i^{(t)}$.

**Gain function for the Shannon surprise:** For the Shannon surprise (c.f. Eq. 10) we have

$$x_{t+1} = f_{\text{Sh}}(x_{1:t}, y_{1:t}) = \arg \max_{i \in \{1,...,N\}} E_{P(.|x_{t+1} = i; \pi(t))}[-\log P(Y_{t+1}|x_{t+1} = i; \pi(t))]$$

$$= \arg \max_{i \in \{1,...,N\}} g_{Sh}(\hat{\alpha}_i^{(t)}),$$

with the corresponding gain function (c.f. Eq. 115)

$$g_{Sh}(\hat{\alpha}_i^{(t)}) = H\left[\hat{\alpha}_i^{(t)}, \hat{\alpha}_i^{(0)}\right].$$

$g_{Sh}$ is independent of $\alpha_{\text{sum},i}^{(0)}$, $\hat{\alpha}_i^{(0)}$, and $\alpha_{\text{sum},i}^{(t)}$, and it has an inverted-U-relation with $\hat{\alpha}_i^{(t)}$ with its maximum at $\hat{\alpha}_i^{(t)} = 0.5$.

**Gain function for the Bayesian surprise:** For the Bayesian surprise Eq. 24 we have (using Eq. 97, Eq. 99, and Eq. 116)

$$f_{\text{Ba}}(x_{1:t}, y_{1:t}) = \arg \max_{i \in \{1,...,N\}} E_{P(.|x_{t+1} = i; \pi(t))}\left[D_{\text{KL}}\left[\pi(t)||\psi(t)(\Theta|Y_{t+1}, X_{t+1} = i)\right]\right]$$

$$= \arg \max_{i \in \{1,...,N\}} E_{P(.|x_{t+1} = i; \pi(t))}\left[\log P_{\text{Band}}(Y_{t+1}, \hat{\alpha}_i^{(t)}) + \psi(\alpha_{\text{sum},i}) - \psi(\alpha_{\text{sum},i}P_{\text{Band}}(Y_{t+1}, \hat{\alpha}_i^{(t)}))\right]$$

$$= \arg \max_{i \in \{1,...,N\}} g_{Ba}(\hat{\alpha}_i^{(t)}, \alpha_{\text{sum},i}^{(t)})$$

(127)
with the corresponding gain function

\[ g_{\text{Ba}}(\hat{p}_i(t), \alpha_{\text{sum},i}) = \psi(\alpha_{\text{sum},i}) - [\hat{p}_i(t) \psi(\alpha_{\text{sum},i} \hat{p}_i(t)) + (1 - \hat{p}_i(t)) \psi(\alpha_{\text{sum},i}(1 - \hat{p}_i(t)))] - H[\hat{p}_i(t), \hat{p}_i(0)]. \] (128)

Using the expanded representation of \( S_{\text{Ba}} \) in Eq. 101, we can rewrite the gain function as

\[ g_{\text{Ba}}(\hat{p}_i(t), \alpha_{\text{sum},i}) = \frac{1}{2 \alpha_{\text{sum},i}} + \sum_{n=1}^{\infty} \frac{|G_{n+1}| n!}{\alpha_{\text{sum},i}^{\alpha(t)}(\hat{p}_i(t) + 1)^n} + \frac{1}{\alpha_{\text{sum},i}(1 - \hat{p}_i(t) + 1)^n} - \frac{1}{(\alpha_{\text{sum},i} + 1)^n}. \] (129)

g_{\text{Ba}} is independent of \( \hat{p}_i(0) \) and \( \alpha_{\text{sum},i}^{(0)} \), and its behavior with respect to the other two variables is as follows:

- With respect to \( \hat{p}_i(t) \): Using Eq. 129 and the fact that \( \hat{p}_i(t) \in (0, 1) \), it is straightforward to show that the partial derivative of \( g_{\text{Ba}} \) with respect to \( \hat{p}_i(t) \) is negative for \( \hat{p}_i(t) < 0.5 \), is positive for \( \hat{p}_i(t) > 0.5 \), and is equal to 0 for \( \hat{p}_i(t) = 0.5 \). Therefore, \( g_{\text{Ba}} \) has a U-relation with respect to \( \hat{p}_i(t) \) with its minimum at \( \hat{p}_i(t) = 0.5 \). However, using asymptotic approximation of the digamma function (Abramowitz & Stegun, 1948), we have \( g_{\text{Ba}} \approx 1/(2 \alpha_{\text{sum},i}^{\alpha(t)}) \) for large values of \( \alpha_{\text{sum},i}^{(t)} \), which is independent of \( \hat{p}_i(t) \). Therefore, seeking Bayesian surprise becomes independent of \( \hat{p}_i(t) \) as \( \alpha_{\text{sum},i}^{(t)} \) increases.

- With respect to \( \alpha_{\text{sum},i}^{(t)} \): using the fact that \( S_{\text{Ba}} \) is always a decreasing function of \( \alpha_{\text{sum},i}^{(t)} \) (c.f. Eq. 103) independent of the observation \( Y_{t+1} \) and its probability \( \hat{p}_i(t) \), we can conclude that \( g_{\text{Ba}} \) is also a decreasing function of \( \alpha_{\text{sum},i}^{(t)} \).

**Gain function for the Confidence Corrected surprise:** For the Confidence Corrected surprise, by using Proposition 6, we have

\[ f_{\text{CC}}(x_{1:t}, y_{1:t}) = \arg \max_{i \in \{1, ..., N\}} \mathbb{E}_{P(\cdot|x_{t+1}=i; \pi(t))} [S_{\text{CC}}(Y_{t+1}; x_{t+1} = i, \pi(t))] \]

\[ = \arg \max_{i \in \{1, ..., N\}} g_{\text{CC}}(\hat{p}_i(t), \alpha_{\text{sum},i}, \hat{p}_i(0), \alpha_{\text{sum},i}) \] (130)

with the corresponding gain function

\[ g_{\text{CC}}(\hat{p}_i(t), \alpha_{\text{sum},i}, \hat{p}_i(0), \alpha_{\text{sum},i}) = g_{\text{BF}}(\hat{p}_i(t), \hat{p}_i(0)) + g_{\text{Ba}}(\hat{p}_i(t), \alpha_{\text{sum},i}, \hat{p}_i(0), \alpha_{\text{sum},i}) + D_{\text{KL}}[\pi(t)||\pi(0)]. \] (131)

g_{\text{CC}} depends on all our 4 variables \( \hat{p}_i(t), \alpha_{\text{sum},i}, \hat{p}_i(0), \) and \( \alpha_{\text{sum},i}^{(0)} \). We could not make any general statement about the relation of \( g_{\text{CC}} \) with these variables. However, using asymptotic approximation of the gamma and digamma functions (Abramowitz & Stegun, 1948), we have

\[ g_{\text{CC}} \approx \frac{1}{2} \log \alpha_{\text{sum}} + \text{Const. w.r.t. } \alpha_{\text{sum}}^{(t)} \] (132)

for large values of \( \alpha_{\text{sum}}^{(t)} \). Therefore, \( g_{\text{CC}} \) is an increasing function of \( \alpha_{\text{sum}}^{(t)} \) for large values of \( \alpha_{\text{sum}}^{(t)} \).

**Simulation details.** We first put \( \alpha_{\text{sum},i}^{(0)} \) equal to 2 and 10 for scenarios 1 and 2, respectively. Then, for a given random seed, we randomly sampled prior parameter \( \hat{p}_i^{(0)} \) for \( i \in \{1, ..., 10\} \), differently for two scenarios: \( \hat{p}_i^{(0)} \sim \delta_{0.5} \) for scenario 1 and \( \hat{p}_i^{(0)} \sim \text{Uniform}(0.05, 0.95) \) for scenario 2. Given the same random seed, we build the environment by sampling \( p_i \sim \text{Dir}(\alpha_{\text{sum},i}^{(0)}, [\hat{p}_i^{(0)}, 1 - \hat{p}_i^{(0)}]) \) independently for
1916 $i \in \{1, \ldots, 10\}$ – note that $p_i$s are not known to the agents. We then ran 6 different algorithms separately, corresponding to (1) $g^*$, (2) $g_{Sh}$, (3) $g_{BF}$, (4) $g_{Ba}$, (5) $g_{CC}$, and (6) $g_{SHR}$. At time $t$, each algorithm computed its corresponding gain function $g$ for all actions. Then, it chose the action with highest gain as $x_t$ – when there was a tie, the action with the smaller index was chosen, e.g., action 2 was preferred to action 6. Given their actions $x_t$, different algorithms observed different observations $y_t$. Then, each algorithm updated its belief, and this procedure was repeated until $t = 100$. At time $t$, the mean-squared error $MSE_t = \sum_{i=1}^{N} (p_i - \hat{p}_i(t))^2$ was computed as a measure of performance. Fig. 12A, Fig. 12C, and Fig. 13M-N show average $MSE_t$ over 500 random seeds as a function of time.

**Model recovery.** To perform model-recovery, we characterized each exploration policy as a stochastic model with a softmax policy over a gain function: We assumed that the probability of taking action $x_{t+1} = i$ by the agent that uses the gain function $g$ is

$$p(x_{t+1} = i | x_{1:t}, y_{1:t}; g, \beta) \propto \exp \left[ \beta \cdot g(p_i(t), \alpha_{\text{sum},i}^{(0)}, \alpha_i^{(0)}) \right],$$

where $\beta$ is a free parameter. Then, given a sequence of actions $x_{1:t}$ and observations $y_{1:t}$ (for one random seed) and for a gain function $g$, we approximated $p(x_{1:t} | y_{1:t}; g)$ by $\max_\beta p(x_{1:t} | y_{1:t}; g, \beta)$. Because we have the same number of free parameters for all models, this approximation is equivalent to using Schwarz approximation – also known as Bayesian Information Criterion; BIC (Efron & Hastie, 2016; Neath & Carvalho, 2012; Schwarz, 1978). We then considered a uniform prior over policies and found the posterior probability of each policy given an observed sequence of action-choices: $p(g | x_{1:t}, y_{1:t}) \propto p(x_{1:t} | y_{1:t}; g)$. The confusion matrices in Fig. 12B and Fig. 12D show average $p(g | x_{1:t}, y_{1:t})$ over 500 random seeds.

**Appendix D: EEG analysis**

In this appendix, we provide the methods and results for the EEG analysis reported in Fig. 7.

**Dataset**

We used a publicly available dataset published in Robbins et al., 2018. The dataset has been originally collected for the study reported in Hairston et al., 2014 and includes the behavioral and the EEG data of 18 adult participants in a visual oddball experiment. The details of the experiment are provided in Robbins et al., 2018; we briefly review the main features here.

Participants were instructed to respond to the standard and deviant stimuli separately by pressing different buttons. The standard stimulus was an image of a US soldier, and the deviant stimulus was an image of an enemy combatant. At each trial, one of the two images was randomly chosen (with roughly 1/8 probability of choosing the deviant image (Modirshanechi et al., 2019)) and presented for 150ms on a computer screen. Each participant completed three blocks of (approximately) 89 trials each. EEG signals were sampled at a rate of 512Hz in a 64-channel electrode space. The data was preprocessed to remove line noise, detect and interpolate bad electrodes, and remove artifacts (see Robbins et al., 2018 for details).

We used the preprocessed data.

**Further pre-processing**

For each participant, we averaged the EEG signals over the central electrodes (Cz, C1, and C2) which we then further smoothed (moving averaging with the window of length 50ms) and standardized (zero mean and unit variance). We extracted EEG trials from 100ms before to 900ms after the stimulus onset and removed the baseline activity by subtracting the mean calculated over the first 100ms. We excluded error trials (i.e., the trials where participants either pressed the wrong button or did not press any button, error rate: 3.8% ± 2.4%, range: 0.4% to 11.3%) from further analyses.
Analysis

For each participant, we computed the event related potentials (ERPs) separately for the standard and the deviant stimuli, averaged over all trials. Fig. 7D shows the mean and the standard error of the mean (over participants) of the standard and the deviant ERPs. We used one-sample t-test (FDR controlled by 0.1 (Benjamini & Hochberg, 1995)) and found time windows where the standard and the deviant ERPs were significantly different. Then, for participant $i \in \{1, \ldots, 18\}$, we found the time-points where the deviant ERP had its maximum in time windows of 450-600ms (which we call $T_i$) and 600-800ms (which we call $T_i'$) – see Kolossa et al., 2015; Kopp and Lange, 2013 for similar approaches. Then, for participant $i$ and trial $t$, we averaged the ERP amplitudes at time windows of $T_i \pm 20$ms (called $z_{i,t}$) and $T_i' \pm 20$ms (called $z_{i,t}'$) of the standard and the deviant stimuli in the early and the late phase of the task separately for each participant. The values showed in Fig. 7E and Fig. 7F corresponds to the mean and the standard error of the mean (over 18 participants) of the average-Z (or the average-Z') at standard trials (Fig. 7E) and the average-Z (or the average-Z') at deviant trials (Fig. 7F).

Results

$Z$ is a candidate for the amplitude of the EEG P3a component and $Z'$ for the amplitude of the EEG P3b component (Kolossa et al., 2015; Kopp & Lange, 2013; Visalli et al., 2021). It has been reported that the latency of the EEG P300 component varies depending on task features (Kutas et al., 1977; Magliero et al., 1984), but we also note that the latency of $Z$ and $Z'$ is different from the latency of P3a and P3b in previous surprise studies (Kolossa et al., 2015; Kopp & Lange, 2013; Visalli et al., 2021). Therefore, we consider our results complementary to the earlier results.

Our results show that $Z$ has a behavior consistent with the Shannon surprise. In contrast to previous studies that found P3a (in tasks different from oddball task) to be a signature of the Bayesian surprise (Kolossa et al., 2015; Visalli et al., 2021), our results provide strong evidence against a correlation between $Z$ and the Bayesian or the Confidence Corrected surprise (Fig. 7E and Fig. 7F). Our results show that the prediction of the Bayesian surprise is not consistent with the behavior of $Z'$ either (Fig. 7F). If $Z'$ is interpreted as the P3b amplitude, then this result supports earlier findings that P3b reflects the Shannon surprise (Kolossa et al., 2015; Kopp & Lange, 2013; Visalli et al., 2021). However, our results cannot determine whether the Shannon or the Confidence Corrected surprise is a better fit for $Z'$ and suggest a mixed effect (e.g., a correlation between $Z'$ and a linear combination of the Shannon or the Confidence Corrected surprise) which can be tested by further and more advanced analyses.

Appendix E: Regularized Shannon surprise

In this Appendix, we consider $R(z) = z$ and derive the formulas used in Fig. 13.

Case-study 1

Using Eq. 96, we can write

$$S_{ShR}(y = i; \pi^{(t)}) = -\log \hat{p}_i^{(t)} + \left( \log \hat{p}_i^{(t)} - \log \hat{p}_i^{(t)}_{\text{max}} \right)$$

$$= -2 \log \hat{p}_i^{(t)} + \log \hat{p}_i^{(t)}_{\text{max}},$$

where

$$\hat{p}_i^{(t)}_{\text{max}} = \max_j \hat{p}_j^{(t)}.$$
As a result, we have

\[
\frac{\partial S_{\text{ShR}}(y = i; \pi(t))}{\partial \hat{p}_k^{(t)}} = \begin{cases} 
-\frac{2}{\hat{p}_k^{(t)}} & < 0 \quad \text{if } i = k \neq \arg \max_j \hat{p}_j^{(t)} \\
-\frac{1}{\hat{p}_k^{(t)}} & < 0 \quad \text{if } i = k = \arg \max_j \hat{p}_j^{(t)} \\
\frac{1}{\hat{p}_k^{(t)}} & > 0 \quad \text{if } i \neq k = \arg \max_j \hat{p}_j^{(t)} \\
0 & \text{otherwise},
\end{cases}
\]

(136)

The regularized Shannon surprise is constant with respect to \(\alpha_{\text{sum}}^{(t)}\).

**Case-study 2**

We define

\[ P^{(t)} = b^{(t)}x_{t+1} + \frac{(1 - b^{(t)})}{2} \]

(137)

\[ P_{\text{max}}^{(t)} = b^{(t)} \max \{ x_{t+1}, 1 - x_{t+1} \} + \frac{(1 - b^{(t)})}{2}. \]

Using Eq. 108, we have

\[ S_{\text{ShR}}(y_{t+1} = 1|x_{t+1}; b^{(t)}) = -\log P^{(t)} + \left( \log P_{\text{max}}^{(t)} - \log P^{(t)} \right). \]

(138)

**Case-study 3**

Using the general form of the surprise seeking strategies (c.f. Eq. 49) and the definition of the regularized Shannon surprise (c.f. Eq. 53) we can write

\[ x_{t+1} = f_{\text{ShR}}(x_{1:t}, y_{1:t}) \]

\[ = \arg \max_{i \in \{1, \ldots, N\}} \mathbb{E}[P_{(i), x_{t+1} = i; \pi(t)}] \left[ -2 \log P(Y_{t+1}|x_{t+1} = i; \pi(t)) + \log \max (\hat{p}_i^{(t)}, 1 - \hat{p}_i^{(t)}) \right] \]

(139)

with the corresponding gain function

\[ g_{\text{ShR}}(\hat{p}_i^{(t)}) = 2H(\hat{p}_i^{(t)}, \hat{p}_i^{(t)}) + \log \max (\hat{p}_i^{(t)}, 1 - \hat{p}_i^{(t)}), \]

(140)

which is an even function of \(\hat{p}_i^{(t)} - 0.5\). For \(\hat{p}_i^{(t)} > 0.5\), the derivative of the gain function is

\[ \frac{\partial g_{\text{ShR}}}{\partial \hat{p}_i^{(t)}} = 2 \left[ \frac{1}{2 \hat{p}_i^{(t)}} - \log \frac{\hat{p}_i^{(t)}}{1 - \hat{p}_i^{(t)}} \right]. \]

(141)

The derivative is positive for \(0.5 < \hat{p}_i^{(t)} < p^*\), is negative for \(p^* < \hat{p}_i^{(t)} < 1\), and is 0 for \(\hat{p}_i^{(t)} = p^*\), where \(p^* \approx 0.68\). Therefore, the gain function has an inverted-W-relation with \(\hat{p}_i^{(t)}\) with its maximum at \(\hat{p}_i^{(t)} = 1 - p^* \approx 0.32\) and \(\hat{p}_i^{(t)} = p^* \approx 0.68\) – corresponding to the most preferred arms.