*ARTICLE* **Communicated by Stefan Kiebel**

# Active Inference, Curiosity and Insight

**Karl J. Friston**
*k.friston@ucl.ac.uk*
**Marco Lin**
*marco.lin91@gmail.com*
*Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London WC1N 3BG, U.K.*

**Christopher D. Frith**
*c.frith@ucl.ac.uk*
*Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London WC1N 3BG, and Institute of Philosophy, School of Advanced Studies, University of London EC1E 7HU, U.K.*

**Giovanni Pezzulo**
*giovanni.pezzulo@gmail.com*
*Institute of Cognitive Sciences and Technologies, National Research Council, 7-00185 Rome, Italy*

**J. Allan Hobson**
*allan_hobson@hms.harvard.edu*
*Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London WC1N 3BG, U.K., and Division of Sleep Medicine, Harvard Medical School, Boston, MA 02215, U.S.A.*

**Sasha Ondobaka**
*s.ondobaka@ucl.ac.uk*
*Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London WC1N 3BG, U.K.*

**This article offers a formal account of curiosity and insight in terms of active (Bayesian) inference. It deals with the dual problem of inferring states of the world and learning its statistical structure. In contrast to current trends in machine learning (e.g., deep learning), we focus on how people attain insight and understanding using just a handful of observations, which are solicited through curious behavior. We use simulations of abstract rule learning and approximate Bayesian inference to show that minimizing (expected) variational free energy leads to active sampling of novel contingencies. This epistemic behavior closes explanatory gaps in generative models of the world, thereby reducing uncertainty and**

*Neural Computation* **29, 2633–2683** (2017) 
doi:10.1162/NECO_a_00999

**satisfying curiosity. We then move from epistemic learning to model selection or structure learning to show how abductive processes emerge when agents test plausible hypotheses about symmetries (i.e., invariances or rules) in their generative models. The ensuing Bayesian model reduction evinces mechanisms associated with sleep and has all the hallmarks of "aha" moments. This formulation moves toward a computational account of consciousness in the pre-Cartesian sense of sharable knowledge (i.e., *con*: "together"; *scire*: "to know").**

## 1 Introduction

This article presents a formal (computational) description of epistemic behavior that calls on two themes in theoretical neurobiology. The first is the use of Bayesian principles for understanding the nature of intelligent and purposeful behavior (Koechlin, Ody, & Kouneiher, 2003; Oaksford & Chater, 2003; Coltheart, Menzies, & Sutton, 2010; Nelson, McKenzie, Cottrell, & Sejnowski, 2010; Collins & Koechlin, 2012; Solway & Botvinick, 2012; Donoso, Collins, & Koechlin, 2014; Seth, 2014; Koechlin, 2015; Lu, Rojas, Beckers, & Yuille, 2016). The second is the role of self-modeling, reflection, and sleep (Metzinger, 2003; Hobson, 2009). In particular, we formulate curiosity and insight in terms of inference—namely, the updating of beliefs about how our sensations are caused. Our focus is on the transitions from states of ignorance to states of insight—namely, states with (i.e., *con*) awareness (i.e., *scire*) of causal contingencies. We associate these epistemic transitions with the process of Bayesian model selection and the emergence of insight. In short, we try to show that resolving uncertainty about the world, through active inference, necessarily entails curious behavior and consequent 'aha' or eureka moments.

The basic theme of this article is that one can cast learning, inference, and decision making as processes that resolve uncertainty about the world. This theme is central to many issues in psychology, cognitive neuroscience, neuroeconomics, and theoretical neurobiology, which we consider in terms of *curiosity* and *insight*. The purpose of this article is not to review the large literature in these fields or provide a synthesis of established ideas (e.g., Schmidhuber, 1991; Oaksford & Chater, 2001; Koechlin et al., 2003; Botvinick & An, 2008; Nelson et al., 2010; Navarro & Perfors, 2011; Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Botvinick & Toussaint, 2012; Collins & Koechlin, 2012; Solway & Botvinick, 2012; Donoso et al., 2014). Our purpose is to show that the issues this diverse literature addresses can be accommodated by a single imperative (minimization of expected free energy, or resolution of uncertainty) that already explains many other phenomena–for example, decision making under uncertainty, stochastic optimal control, evidence accumulation, addiction, dopaminergic responses, habit learning, reversal learning, devaluation, saccadic searches,

scene construction, place cell activity, omission-related responses, mismatch negativity, P300 responses, phase-precession, and theta-gamma coupling (Friston, FitzGerald et al., 2016; Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017). In what follows, we ask how the resolution of uncertainty might explain curiosity and insight.

**1.1 Curiosity.** Curiosity is an important concept in many fields, including psychology (Berlyne, 1950, 1954; Loewenstein, 1994), computational neuroscience, and robotics (Schmidhuber, 1991; Oaksford & Chater, 2001). Much of neural development can be understood as learning contingencies about the world and how we can act on the world (Saegusa, Metta, Sandini, Sakka, 2009; Nelson et al., 2010; Nelson, Divjak, Gudmundsdottir, Martignon, & Meder, 2014). This learning rests on intrinsically motivated curious behavior that enables us to predict the consequences of our actions: as nicely summarized by Still and Precup (2012), "A learner should choose a policy that also maximizes the learner's predictive power. This makes the world both interesting and exploitable." This epistemic, world-disclosing perspective speaks to the notion of optimal data selection and important questions about how rational or optimal we are in querying our world (Oaksford, Chater, Larkin, 2000; Oaksford & Chater, 2003). Clearly, the epistemic imperatives behind curiosity are especially prescient in developmental psychology and beyond: "In the absence of external reward, babies and scientists and others explore their world. Using some sort of adaptive predictive world model, they improve their ability to answer questions such as what happens if I do this or that?" (Schmidhuber, 2006). In neurorobotics, these imperatives are often addressed in terms of active learning (Markant & Gureckis, 2014; Markant, Settles, & Gureckis, 2016), with a focus on intrinsic motivation (Baranes & Oudeyer, 2009). Active learning and intrinsic motivation are also key concepts in educational psychology, where they play an important role in enabling insight and understanding (Eccles & Wigfield, 2002).

**1.2 Insight and Eureka Moments.** The Eureka effect (Auble, Franks, & Soraci, 1979) was introduced to psychology by comparing the recall for sentences that were initially confusing but subsequently understood. The implicit resolution of confusion appears to be the main determinant of recall and the emotional concomitants of insight (Shen, Yuan, Liu, & Luo, 2016). Several psychological theories for solving insight problems have been proposed—for example, progress monitoring and representational change theory (Knoblich, Ohlsson, & Raney, 2001; MacGregor, Ormerod, & Chronicle, 2001). Both enjoy empirical support, largely from eye movement studies (Jones, 2003). Furthermore, several psychophysical and neuroimaging studies have attempted to clarify the functional anatomy of insight (see Bowden, Jung-Beeman, Fleck, & Kounios, 2005), for a psychological review and Dresler et al., 2015, for a review of the neural correlates of

insight in dreaming and psychosis). In what follows, we offer a normative framework that complements psychological theories by describing how curiosity engenders insight. Our treatment is framed by two questions posed by Berlyne (1954) in his seminal treatment of curiosity: "The first question is why human beings devote so much time and effort to the acquisition of knowledge. . . . The second question is why, out of the infinite range of knowable items in the universe, certain pieces of knowledge are more ardently sought and more readily retained than others?" (p. 180).

In brief, we will try to show that the acquisition of knowledge and its retention are emergent properties of active inference—specifically, that curiosity manifests as an active sampling of the world to minimize uncertainty about hypotheses—or explanations—for states of the world, while retention of knowledge entails the Bayesian model selection of the most plausible explanation. The first process rests on curious, evidence-accumulating, uncertainty-resolving behavior, while the second operates on knowledge structures (i.e., generative models) after evidence has been accumulated.

Our approach rests on the free energy principle, which asserts that any sentient creature must minimize the entropy of its sensory exchanges with the world. Mathematically, entropy is uncertainty or expected surprise, where surprise can be expressed as a free energy function of sensations and (Bayesian) beliefs about their causes. This suggests that creatures are compelled to minimize uncertainty or expected free energy. In what follows, we will see that resolving different sorts of uncertainty furnishes principled explanations for different sorts of behavior. These levels of uncertainty pertain to plausible states of the world, plausible policies that change those states, and plausible models of those changes.

The first level of uncertainty is about the causes of sensory outcomes under a particular policy (i.e., sequence of actions). Reducing this sort of uncertainty corresponds to perceptual inference (a.k.a. state estimation). In other words, the first thing we need to do is infer the current state of the world and the context in which we are operating. We then have to contend with uncertainty about policies per se that can be cast in terms of uncertainty about future states of the world, outcomes, and the probabilistic contingencies that bind them. We will see that minimizing these three forms of expected surprise—by choosing an uncertainty resolving policy—corresponds to information-seeking epistemic behavior, goal-seeking pragmatic behavior, and novelty-seeking curious behavior, respectively. In short, by pursuing the best policy, we accumulate experience and reduce uncertainty about probabilistic contingencies through epistemic learning—namely, inferring (the parameters of our models of) how outcomes are generated.

Finally, curious, novelty-seeking policies enable us to reduce our uncertainty about our generative models per se, leading to structure learning, insight, and understanding. Here, a generative model constitutes a hypothesis about how observable outcomes are generated, where we entertain

Table 1: Sources of Uncertainty Scored by (Expected) Free Energy and the Behaviors Entailed by Its Minimization (Resolution of Uncertainty through Approximate Bayesian Inference).

| Source of Uncertainty | Free Energy (Surprise) | Minimization | Active Inference |
|---|---|---|---|
| Uncertainty about hidden states given a policy | $F(\pi) = F(\tilde{o}, \mathbf{s}_\tau^\pi, \mathbf{a}\|\pi)$ | With respect to expected states $\mathbf{s}_\tau^\pi$ | Perceptual inference (*state estimation*) |
| Uncertainty about policies in terms of expected:<br>Future states (*intrinsic value*)<br>Future outcomes (*extrinsic value*)<br>Model parameters (*novelty*) | $G(\pi) = G(\mathbf{s}_\tau^\pi, \mathbf{a}\|\pi) =$<br>$\mathbf{o}_\tau^\pi \cdot \widehat{\mathbf{o}}_\tau^\pi + \mathbf{H} \cdot \mathbf{s}_\tau^\pi +$<br>$\mathbf{o}_\tau^\pi \cdot \mathbf{C}_\tau +$<br>$\mathbf{o}_\tau^\pi \cdot \mathbf{W} \cdot \mathbf{s}_\tau^\pi$ | With respect to policies $\boldsymbol{\pi}$ | Epistemic planning<br>Intrinsic motivation<br>Extrinsic motivation<br>Curiosity |
| Uncertainty about model parameters given a model | $F(\tilde{o}, \mathbf{s}_\tau^\pi, \boldsymbol{\pi}, \mathbf{a}\|m)$ | With respect to parameters $\mathbf{a}$ | Epistemic learning (*active learning*) |
| Uncertainty about the model | $F(\tilde{o}, \mathbf{s}_\tau^\pi, \boldsymbol{\pi}, \mathbf{a}\|m)$ | With respect to model $m$ | Structure learning (*insight and understanding*) |

competing hypotheses that are, a priori, equally plausible. In short, the last level of uncertainty reduction entails the selection of models that render outcomes the least surprising, having suppressed all other forms of uncertainty. All but the last process require experience to resolve uncertainty about either the states (inference) or parameters (learning) of a particular model. However, optimization of the model per se can proceed in a fact-free, or outcome-free, fashion, using experience accumulated to date. In other words, no further facts or outcomes are necessary for this last level of optimization: facts and outcomes are constitutive of the experience on which this optimization relies. It is this Bayesian model selection we associate with fact-free learning (Aragones, Gilboa, Postlewaite, & Schmeidler, 2005) and the emergence of insight (Bowden et al., 2005).

Table 1 provides a summary of these uncertainty-reducing processes, where uncertainty is associated with free energy formulations of surprise such that uncertainty-resolving behavior reduces expected free energy. To motivate and illustrate this formalism, we set ourselves the task of simulating a curious agent that spontaneously learned rules—governing the sensory consequences of her action—from limited and ambiguous sensory evidence (Lu et al., 2016; Tervo, Tenenbaum, & Gershman, 2016). We chose abstract rule learning to illustrate how conceptual knowledge could be

accumulated through experience (Botvinick & Toussaint, 2012; Zhang & Maloney, 2012; Koechlin, 2015) and how implicit Bayesian belief updating can be accelerated by applying Bayesian principles not to sensory samples but to beliefs based on those samples. This structure learning (Tenenbaum et al., 2011; Tervo et al., 2016) is based on recent developments in Bayesian model selection, namely, Bayesian model reduction (Friston, Litvak et al., 2016). Bayesian model reduction refers to the evaluation of reduced forms of a full model to find simpler (reduced) models using only posterior beliefs (Friston & Penny, 2011). Reduced models furnish parsimonious explanations for sensory contingencies that are inherently more generalizable (Navarro & Perfors, 2011; Lu et al., 2016) and, as we will see, provide for simpler and more efficient inference. In brief, we use simulations of abstract rule learning to show that context-sensitive contingencies, which are manifest in a high-dimensional space of latent or hidden states, can be learned using straightforward variational principles (i.e., minimization of free energy). This speaks to the notion that people "use their knowledge of real-world environmental statistics to guide their search behavior" (Nelson et al., 2014). We then show that Bayesian model reduction adds an extra level of inference, which rests on testing plausible hypotheses about the structure of internal or generative models. We will see that this process is remarkably similar to physiological processes in sleep, where redundant (synaptic) model parameters are eliminated to minimize model complexity (Hobson & Friston, 2012). We then show that qualitative changes in model structure emerge when Bayesian model reduction operates online during the assimilation of experience. The ensuing optimization of model evidence provides a plausible (Bayesian) account of abductive reasoning that looks very much like an "aha" moment. To simulate something akin to an aha moment requires a formalism that deals explicitly with probabilistic beliefs about states of the world and its causal structure. This contrasts with the sort of structure or manifold learning that predominates in machine learning (e.g., deep learning; LeCun, Bengio, & Hinton, 2015), where the objective is to discover structure in large data sets by learning the parameters of neural networks. This article asks whether abstract rules can be identified using active (Bayesian) inference, following a handful of observations and plausible, uncertainty-reducing hypotheses about how sensory outcomes are generated.

**1.3 Active Inference and the Resolution of Uncertainty.** Active inference is a corollary of the free energy principle that tries to explain action and perception in terms of minimizing variational free energy. Variational free energy is a proxy for surprise or (negative) Bayesian model evidence. This means that minimizing free energy corresponds to avoiding surprises or maximizing model evidence, and minimizing expected free energy corresponds to resolving uncertainty. The active aspect of active inference emphasizes that we are the embodied authors of our sensations. This means

that the consequences of action must themselves be inferred (Baker, Saxe, & Tenenbaum, 2009). In turn, this implies that we have (prior) beliefs about our behavior. Active inference assumes that the only (self-consistent) prior belief is that we will minimize free energy; in other words, we (believe we) will resolve uncertainty through active sampling of the world (Friston, Mattout, & Kilner, 2011; Friston et al., 2015). Alternative prior beliefs can be discounted by *reductio ad absurdum*: if we do not believe that we will resolve uncertainty through active inference, and active inference realizes beliefs by minimizing uncertainty (i.e., fulfilling expectations), then active inference will not minimize uncertainty.

From a technical perspective, this article introduces generalizations of active inference for discrete state-space models (i.e., hidden Markov models and Markov decision processes) along two lines, both concerning the parameters of generative models that encode probabilistic contingencies. First, posterior beliefs about both hidden states and parameters are included in expected free energy, leading to epistemic or exploratory behavior that tries to resolve ignorance, in addition to risk and ambiguity. In other words, policies acquire epistemic value in virtue of resolving uncertainty about states and outcomes (risk and ambiguity) or resolving uncertainty about contingencies (ignorance)—in other words, "what happens if I do this or that?" (Schmidhuber, 2006). Second, we consider minimizing the free energy of the model per se (as opposed to model parameters), in terms of prior beliefs about which parameters are necessary to explain observed outcomes and which parameters are redundant and can be eliminated. As with our previous treatments of active inference, we pay special attention to biological plausibility and try to link optimization to neuronal processes. These developments can be regarded as rolling back the implications of minimizing variational free energy under a generic internal or generative model of the world.

**1.4 Overview.** This article has three sections. The first briefly reviews active inference and relates the underlying objective function (expected free energy) to established notions like utility, mutual information, and Bayesian surprise. The second describes the paradigm used in this article. In brief, we require agents to learn an abstract rule, in which the correct response is determined by the color of a cue whose location is determined by another cue. By transcribing task instructions into the prior beliefs of a simulated subject, we examine how quickly the rule can be learned—and how this epistemic learning depends on curious, uncertainty-reducing behavior that resolves ignorance (about the meaning of cues), ambiguity (about the context or rule in play), and risk (of making a mistake). In the third section, we turn to Bayesian model reduction or structure learning and consider the improvement in free energy—and performance—when competing hypotheses about the mapping between hidden states and outcomes are tested against the evidence of experience (Nelson et al., 2010). This evidence is accumulated by posterior beliefs over parameters and can be examined

offline to simulate sleep and the emergence of eureka moments. We conclude with a brief illustration of communicating prior beliefs to others (i.e., sharing of knowledge) and discuss the implications for active inference and artificial intelligence.

## 2 Active Inference and Free Energy

Active inference assumes that every characteristic (variable) of an agent minimizes variational free energy (Friston, 2013). This leads to some surprisingly simple update rules for perception, planning, and learning. In principle, the active inference scheme described in this section can be applied to any paradigm or choice behavior. It has been used to model waiting games (Friston et al., 2013), two-step maze tasks (Friston et al., 2015), evidence accumulation in the urn task (FitzGerald, Schwartenbeck, Moutoussis, Dolan, & Friston, 2015), trust games from behavioral economics (Moutoussis, Trujillo-Barreto, El-Deredy, Dolan, & Friston, 2014), addictive behavior (Schwartenbeck, FitzGerald, Mathys, Dolan, Wurst, Kronbichler, & Friston, 2015), saccadic eye movements in scene construction (Mirza, Adams, Mathys, & Friston, 2016), and engineering benchmarks such as the mountain car problem (Friston, Adams, & Montague, 2012). It has also been used with computational fMRI (Schwartenbeck, FitzGerald, Mathys, Dolan, & Friston, 2015). In short, the simulations used to illustrate the emergence of curiosity and insight below follow from a single principle: the minimization of free energy (i.e., surprise) or maximization of model evidence.

Active inference rests on a generative model of observed outcomes. This model is used to infer the most likely causes of outcomes in terms of expected states of the world. These states are called latent, or hidden because they can only be inferred through observations that are usually limited. Crucially, observations depend on action (e.g., where you are looking), which requires the generative model to entertain expectations about outcomes under different sequences of action (i.e., policies). Because the model generates the consequences of action, it must have expectations about future states. These expectations are optimized by minimizing variational free energy, which renders them the most likely states of the world given current observations. Crucially, the prior probability of a policy depends on the free energy expected when pursuing that policy. The (expected) free energy is a proxy for uncertainty and has a number of familiar special cases, including expected utility, epistemic value, Bayesian surprise, and mutual information. After evaluating the expected free energy of each policy; and implicitly their posterior probabilities, the most likely action can be selected. This action generates a new outcome, and the (perception action) cycle starts again.

The resulting behavior represents a principled sampling of sensory cues that has both epistemic and pragmatic aspects. Generally, behavior in an

ambiguous context is dominated by epistemic imperatives until there is no further uncertainty to resolve and pragmatic (prior) preferences predominate. At this point, explorative behavior gives way to exploitative behavior. In this article, we are interested in epistemic behavior, and use prior preferences only to establish a task or instruction set—namely, report a choice when sufficiently confident.

The formal description of active inference that follows introduces many terms and expressions that might appear a bit daunting at first reading. However, most of the technical material represents a standard treatment of Markov decision processes in terms of belief propagation or variational message passing that has been described in a series of previous papers. Furthermore, the simulations reported in this article and previous papers use exactly the same routines (see the software note at the end of the article). We have therefore tried to focus on the essential ideas (and variables) to provide an accessible and basic account of active inference, so that we can focus on curiosity (epistemic novelty-seeking behavior) and insight (Bayesian model reduction). People who want a more detailed account of the basic active inference scheme can refer to Table 2 (for a full glossary of terms described in the appendix) and Friston, FitzGerald et al. (2016) and Friston et al. (2017).

**2.1 The Generative Model.** Figure 1 provides a schematic specification of the generative model used for the sorts of problems considered in this article. This model is described in more detail in the appendix. In brief, outcomes at any particular time depend on hidden states, while hidden states evolve in a way that depends on action. The generative model is specified by two sets of high-dimensional matrices or arrays. The first, $A^m$, maps from hidden states to the $m$th outcome or modality—for example, exteroceptive (e.g., visual) or proprioceptive (e.g., eye position) modalities. These parameters encode the likelihood of an outcome given their hidden causes. The second set, $\mathbf{B}^n$, prescribes transitions among the $n$th factor of hidden states, under an action specified by the current policy.[1] These hidden factors correspond to different attributes of the world, like the location, color, or category of an object.[2] The remaining parameters encode prior beliefs about future outcomes $\mathbf{C}^m$ and initial states $\mathbf{D}^n$. The probabilistic mappings or contingencies are generally parameterized as Dirichlet distributions, whose sufficient statistics are concentration parameters. Concentration parameters can be thought of as counting the number of times a particular combination of

[1] Parameter matrices in bold denote known parameters. In this article, we consider that all model parameters are known (or have been learned), with the exception of the likelihood mapping; namely, the $A$ parameters.

[2] Implicit in this notation is the factorization of hidden states into factors, whose transitions can be modeled with separate probability transition matrices. This means that the transitions among the levels or states of one factor do not depend on another factor. For example, the way an object moves does not depend on its color.

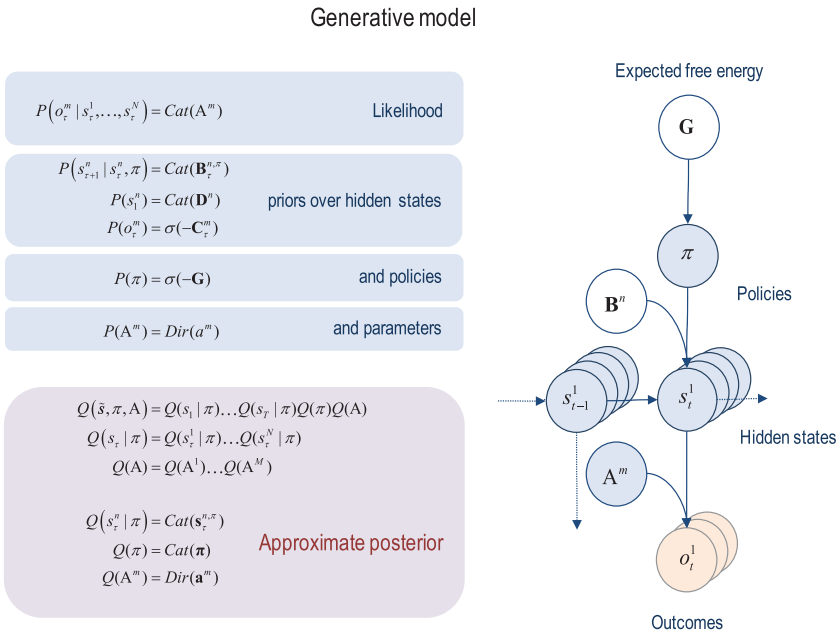



Figure 1: Generative model and (approximate) posterior. A generative model specifies the joint probability of outcomes or consequences and their (latent or hidden) causes. Usually the model is expressed in terms of a likelihood (the probability of consequences given causes) and priors over causes. When a prior depends on a random variable, it is called an *empirical prior*. Here, the likelihood is specified by a high-dimensional array **A** whose components are the probability of an outcome under every combination of hidden states. The empirical priors in this instance pertain to transitions among hidden states **B** that may depend on action, where actions are determined probabilistically in terms of policies (sequences of actions denoted by $\pi$). The key aspect of this generative model is that policies are more probable a priori if they minimize the (path integral of) expected free energy **G**. Bayesian model inversion refers to the inverse mapping from consequences to causes—estimating the hidden states and other variables that cause outcomes. In variational Bayesian inversion, one has to specify the form of an approximate posterior distribution, which is provided in the lower panel. This particular form uses a mean-field approximation in which posterior beliefs are approximated by the product of marginal distributions over hidden states or factors. Here, a mean-field approximation is applied both to posterior beliefs at different points in time and factors. (See the appendix and Table 2 for a detailed explanation of the variables.) The Bayesian network (right panel) provides a graphical representation of the dependencies implied by the equations on the left. Here (and in subsequent figures), $t$ denotes the current time point, and $\tau$ indexes all possible time points.

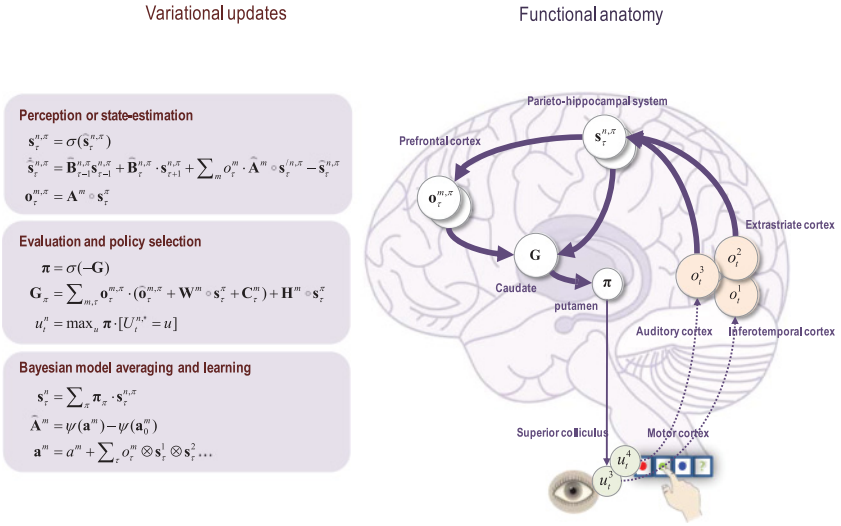



Figure 2: Schematic overview of belief updating. The left panel lists the belief updates mediating perception (i.e., state estimation), policy selection, and learning; while the right panel assigns the updates to various brain areas. This attribution is purely schematic and serves to illustrate a crude functional anatomy. Here, we have assigned observed outcomes to visual representations in the occipital cortex, with visual (*what*) modalities entering a ventral stream and proprioceptive (*where*) modalities originating a dorsal stream. Auditory feedback is associated with the auditory cortex. Hidden states encoding context have been associated with the hippocampal formation and association (parietal) cortex. The evaluation of policies, in terms of their (expected) free energy, has been placed in the caudate. Expectations about policies, assigned to the putamen, are used to create Bayesian model averages of future outcomes (e.g., in the frontal eye fields and supplementary motor area). Finally, expected policies specify the most likely action (e.g., via the deep layers of the superior colliculus). The arrows denote message passing among the sufficient statistics of each factor or marginal. The appendix and Table 2 explain the equations and variables.

states and outcomes has been observed. In this article, we focus on learning the likelihood model and therefore assume that state transitions and initial states are known (or have been learned).

The generative model in Figure 1 means that outcomes are generated in the following way. First, a policy is selected using a softmax function of the expected free energy for each policy. Sequences of hidden states are generated using the probability transitions specified by the selected policy. Finally, these hidden states generate outcomes in one or more modalities. Figure 2 (left panel) provides a graphical summary of the dependencies

implied by the generative model in Figure 1. Perception or inference about hidden states (i.e., state estimation) corresponds to inverting a generative model given a sequence of outcomes, while learning corresponds to updating the parameters of the model. Perception therefore corresponds to optimizing expectations of hidden states and policies with respect to variational free energy, while learning corresponds to accumulating concentration parameters. These constitute the sufficient statistics of posterior beliefs, usually denoted by the probability distribution $Q(x)$, where $x = \tilde{s}, \pi, A$ are hidden or unknown quantities.

**2.2 Variational Free Energy and Inference.** In variational Bayesian inference, model inversion entails minimizing variational free energy with respect to the sufficient statistics of approximate posterior beliefs. These beliefs are approximate because they assume the posterior can be factorized into marginal distributions—here, over hidden states at each point in time, policies, and parameters. This is known as a mean-field assumption (see the factorization of the approximate posterior in the lower right panel of Figure 1). The ensuing minimization of free energy with respect to posterior beliefs can be expressed as follows (see Table 2 for a glossary of expressions):

$$\begin{aligned} Q(x) &= \arg\min_{Q(x)} F \\ &\approx P(x|\tilde{o}), \\ F &= E_Q[\ln Q(x) - \ln P(\tilde{o}, x)], \\ &= \underbrace{D[Q(x)||P(x|\tilde{o})]}_{divergence} - \underbrace{\ln P(\tilde{o})}_{log\ evidence} \\ &= \underbrace{D[Q(x)||P(x)]}_{complexity} - \underbrace{E_Q[\ln P(\tilde{o}|x)]}_{accuracy}, \end{aligned} \tag{2.1}$$

where $\tilde{o} = (o_1, \ldots, o_t)$ denotes observations up to the current time. Because the (Kullback-Leibler, KL) divergence cannot be less than zero, the penultimate equality means that free energy is minimized when the approximate posterior is the true posterior. At this point, the free energy becomes the negative log evidence for the generative model (Beal, 2003). This means that minimizing free energy is equivalent to maximizing model evidence, which is equivalent to minimizing the complexity of accurate explanations for observed outcomes.

Minimizing free energy ensures expectations encode posterior beliefs, given observed outcomes. However, beliefs about policies rest on future outcomes. This means that policies should, a priori, minimize the free energy expected in the future (Friston et al., 2015). This can be formalized as

follows (see the appendix):

$$
\begin{aligned}
P(\pi) &= \sigma(-G(\pi)),\\
G(\pi) &= \sum_{\tau} G(\pi,\tau),\\
G(\pi,\tau) &= E_{\tilde{Q}}[\ln Q(\mathbf{A}, s_\tau|\pi) - \ln P(o_\tau, \mathbf{A}, s_\tau|\tilde{o},\pi)]\\
&= \underbrace{E_{\tilde{Q}}[\ln Q(\mathbf{A}) - \ln Q(\mathbf{A}|s_\tau, o_\tau,\pi)]}_{(negative)\ novelty}\\
&\quad + \underbrace{E_{\tilde{Q}}[\ln Q(o_\tau|\pi) - \ln Q(o_\tau|s_\tau,\pi)]}_{(negative)\ intrinsic\ or\ epistemic\ value} - \underbrace{E_{\tilde{Q}}[\ln P(o_\tau)]}_{extrinsic\ or\ expected\ value}\\
&= \underbrace{E_{\tilde{Q}}[\ln Q(\mathbf{A}) - \ln Q(\mathbf{A}|s_\tau, o_\tau,\pi)]}_{ignorance} + \underbrace{D[Q(o_\tau|\pi)||P(o_\tau)]}_{risk}\\
&\quad + \underbrace{E_{\tilde{Q}}[H[P(o_\tau|s_\tau)]]}_{ambiguity}
\end{aligned}
\tag{2.2}
$$

where $\tilde{Q} = Q(o_\tau, s_\tau|\pi) = P(o_\tau|s_\tau)Q(s_\tau|\pi)$ is the posterior predictive distribution over hidden states and their outcomes under a particular policy. When comparing the penultimate expressions for expected free energy (see equation 2.2) with the free energy per se (see equation 2.1), one sees that the expected divergence becomes mutual information or information gain (see the appendix). Here, we have associated the information gain about the parameters with novelty and information gain about hidden states with intrinsic or epistemic value (i.e., salience). Similarly, expected log evidence becomes expected or extrinsic value provided we associate the prior preference (log probability) over future outcomes with value. The last equality provides a complementary interpretation in which the expected complexity of parameters and hidden states becomes ignorance and risk, while expected inaccuracy becomes ambiguity. We have chosen to label inverse novelty as ignorance in the sense that novelty affords the opportunity to resolve ignorance (i.e., nescience), namely, uncertainty about the contingencies that underwrite outcomes.

There are several special cases of expected free energy that appeal to (and contextualize) established constructs. For example, maximizing mutual information is equivalent to maximizing (expected) Bayesian surprise (Itti & Baldi, 2009), where Bayesian surprise is the divergence between posterior and prior beliefs. This can also be interpreted in terms of the principle of maximum mutual information or minimum redundancy (Barlow, 1961; Linsker, 1990; Olshausen & Field, 1996; Laughlin, 2001). Because mutual information cannot be less than zero, it disappears when the (predictive) posterior ceases to be informed by new observations. This means epistemic

behavior will search out observations that resolve uncertainty about the state of the world (e.g., foraging to resolve uncertainty about the hidden location of prey or fixating on an informative part of a face). However, when there is no posterior uncertainty, and the agent is confident about the state of the world, there can be no further information gain, and epistemic value will be the same for all policies, enabling extrinsic value to dominate (if it did not already). This resolution of uncertainty is closely related to satisfying artificial curiosity (Schmidhuber, 1991; Still & Precup, 2012) and speaks to the value of information (Howard, 1966), particularly in the context of evincing information necessary to realize rewards or payoffs (see Meder & Nelson, 2012). (See also Nelson et al., 2010, who compare different models of information gain in explaining perceptual decisions.) Expected complexity or risk is exactly the same quantity minimized in risk-sensitive or KL control (Klyubin, Polani, & Nehaniv, 2005; van den Broek, Wiegerinck, & Kappen, 2010), and underpins related (free energy) formulations of bounded rationality based on complexity costs (Braun, Ortega, Theodorou, & Schaal, 2011; Ortega & Braun, 2013). In other words, minimizing expected complexity renders behavior risk sensitive, while maximizing expected accuracy induces ambiguity-resolving behavior.

The new term introduced by this article is the information gain pertaining to the likelihood mapping between hidden states and outcomes. This term means that policies will be more likely if they resolve uncertainty—not about hidden states – but about how hidden states generate outcomes. Put simply, this means policies that expose the agent to novel combinations of hidden states and outcomes become attractive because they provide evidence for the way that outcomes are generated. In other words, policies that afford a high degree of novelty resolve ignorance about the relationship between causes and consequences. The subsequent resolution of this ignorance or uncertainty lends meaning to outcomes (consequences) in terms of hidden states (causes). This epistemic affordance will be important in what follows.

**2.3 Belief Updating.** Having defined our objective function, the sufficient statistics encoding posterior beliefs can be updated by minimizing variational free energy. Figure 2 provides these updates. Although the updates look complicated, they are remarkably plausible in terms of neurobiological schemes, as discussed in Friston et al. (2014) and Friston, FitzGerald et al., 2016). The update rules for expected policies (policy selection) and learning are the solutions that minimize free energy, while the updates for expectations over hidden states (for each policy and time) are formulated as a gradient descent. This is important because it provides a dynamical process theory that can be tested against empirical measures of neuronal dynamics. We will see examples of simulated neuronal responses later. Note that the solution for expected policies is a classical softmax function of expected free energy, while learning entails accumulation of concentration

parameters based on the co-occurrence of outcomes and combinations of expected hidden states. Here, the expected hidden states constitute Bayesian model averages over policy-specific expectations (See Friston, FitzGerald et al., 2016) for a more thorough discussion of the neurophysiological implementation of these updates.)

In novel environments, the heavy lifting rests on learning the parameters (and form) of the likelihood mapping. The interesting aspect of these parameters is that they mediate interactions among different hidden states. In other words, they play the role of connections from hidden states to predicted outcomes. From a neurobiological perspective, this means that the connections generating predicted outcomes from expected states (or updating hidden states based on outcomes) are necessarily activity dependent and context sensitive. For example, the first term in the expression for state estimation or perception in Figure 2 is a linear mixture of outcomes formed by a connectivity matrix that itself depends on expectations of over hidden states. In other words, hidden states interact or conspire to generate predictions—or select mixtures of outcomes for Bayesian belief updating. We will return to the importance of these interactions when we consider structure learning.

**2.4 Summary.** By assuming a generic (Markovian) form for the generative model, it is straightforward to derive Bayesian updates that clarify the interrelationships among perception, planning (i.e., policy selection), and learning. In brief, the agent first infers the hidden states under each policy that it entertains. It then evaluates the evidence for each policy based on observed outcomes and beliefs about future states. Posterior beliefs about each policy are then used to select the next action. The ensuing outcomes are used in conjunction with combinations of expected hidden states to accumulate experience and learn contingencies or model parameters. Figure 2 (right) shows the functional anatomy implied by the belief updating and mean-field approximation in Figure 1. Table 1 lists the sources of uncertainty encoded by (expected) variational free energy and the behaviors entailed by its minimization. As noted in section 1, this formalism provides a nice ontology for perception, planning, and learning where planning or policy selection has distinct novelty, information, and goal-seeking components (driven by novelty, salience, and extrinsic value, respectively). We will use this formalism in the next section to illustrate the behavioral (and electrophysiological) responses that attend rule learning.

## 3 Curiosity and Learning

> The first question is why human beings devote so much time and effort to the acquisition of knowledge (Berlyne, 1954).

The (rule-learning) paradigm considered in this section is sufficiently difficult to challenge audiences but simple enough for us to unpack formally. Its agenda is to illustrate curiosity in terms of pursuing policies that afford novelty and the epistemic learning that ensues. The paradigm involves three input modalities (*what*, *where*, and *feedback*) and four sets of hidden states that generate these outcomes—two encoding contextual factors (*rule* and *color*) and two hidden states that can be controlled (*where* and *choice*; see Figure 3).

In brief, artificial subjects could fixate or attend to a fixation point or one of three cue locations. They were told that the color of the central cue specified a rule that would enable them to report the correct color (*red*, *green*, or *blue*) with a button press (*red*, *green*, *blue*, or *undecided*). They were told to report the correct color as accurately as possible after looking at three cues or fewer. The rule the subjects had to discover was as follows: the color of the central cue specifies the location of the correct color. For example, if the subject sees red in the center, the correct color is on the left. When demonstrating this task to audiences we usually say something like:

> On each trial, we will present three colored dots, arranged around a central fixation point. Your task is to choose the correct color. The dots will be red, blue, or green, and dots of the same color can appear together. All we will tell you is that there is a rule that enables you to identify the correct color on every trial—and that this rule is indicated by the color of the central dot. To make things interesting, you can see only one dot at a time, and we expect a decision after you have looked at three dots. Here is the first trial, which color do you think is correct?

Clearly, we did not instruct our synthetic subject verbally. These instructions were conveyed via prior beliefs about the likelihood of outcomes and prior preferences over a feedback modality. These prior preferences ensured that the subject believed that she was unlikely to be wrong and that she was highly likely to decide after the third epoch (i.e., she was likely to comply with task instructions, even if this entailed choosing the wrong color). These prior beliefs were coded in terms of negative value (i.e., Cost) in a feedback modality; $m = 3$, with three levels (undecided, right, and wrong):

$$\mathbf{C}_\tau^m = -\ln P(o_\tau^m) = \begin{cases} 4 & o_\tau^m = wrong : \forall \tau > 0, m = 3 \\ 8 & o_\tau^m = undecided : \forall \tau > 3, m = 3 \\ 0 & otherwise \end{cases} \tag{3.1}$$

In addition to the (visual) *color* and (auditory) *feedback* modalities, subjects also received a (proprioceptive) feedback signaling *where* they were currently looking. Here, $\tau$ indicates the number of saccades or sampled cues in each trial.

The hidden state space induced by the above instructions has four factors: the subject knew that there were three rules; three correct colors (red, green, or blue); where they were looking (left, center, right, or fixation); and their choice (red, green, blue, or undecided). We equipped subjects with six actions: they could look at (or attend to) any of the cue locations without making a choice, or they could return to the fixation point and report their chosen color. In these simulations, policies were very simple and comprised the past sequence of actions plus one of the six actions above. (See Figure 3 for a schematic depiction of the implicit hidden state space.)

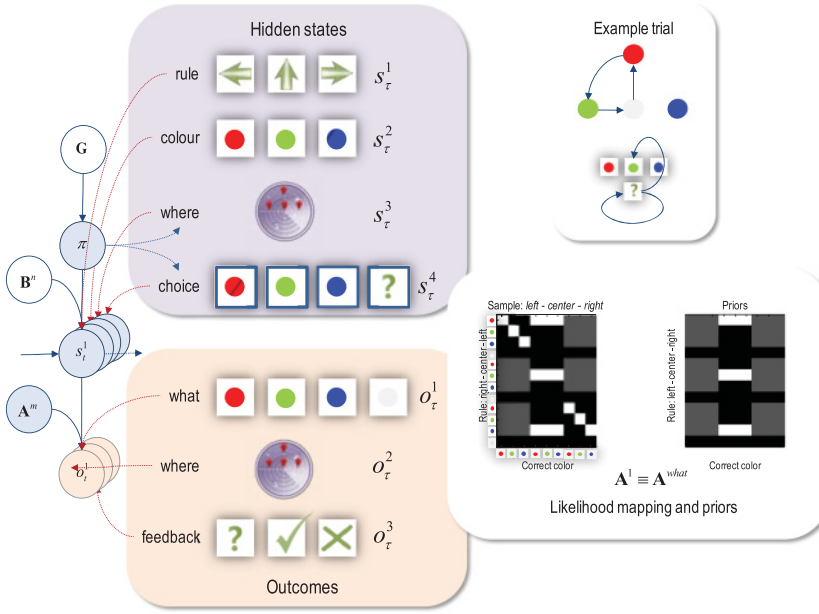
The transition matrices were also simple. The first two are identity matrices, because the context (rule and color) states do not change within a trial. In what follows, each trial begins with a new set of cues and comprises a sequence of epochs, where an epoch corresponds to the belief updating following each new observation (e.g., saccadic eye movement). The remaining (where and choice) probability transition matrices depend on action, where the action invariably changes the hidden state to where the subject looks or the choice she makes:

$$\mathbf{B}^{1,2}_{ij}(u) = \begin{cases} 1 & i = j, \forall u \\ 0 & i \neq j, \forall u. \end{cases}$$

$$\mathbf{B}^{3,4}_{ij}(u) = \begin{cases} 1 & i = u, \forall j \\ 0 & i \neq u, \forall j. \end{cases} \tag{3.2}$$

Finally, prior beliefs about the initial states were uniform distributions apart from the sampled location and choice, which was always looking at the fixation point prior to making a choice $\mathbf{D}^n = [0, \ldots, 0, 1]$.

The only outstanding parameters of the generative model are the concentration parameters of the likelihood $\mathbf{A}^m$ that link hidden states and outcomes. The agent effectively knew the mapping to *where* and *feedback* outcomes, in the sense that we made the concentration parameters high for the correct contingencies (with a value of 128) and zero elsewhere. In other words, the agents knew that feedback depended on choosing the correct color. Furthermore, we used informative concentration parameters to tell the subject that each of the three rules determined the color of the central cue. However, the agents did not know how the rule determined outcomes. This ignorance corresponds to uniform concentration parameters (of one) in the mapping between the correct color and the color seen at each location, under all three levels of the rule. The important parts of the resulting likelihood array are shown in Figure 3 (right inset panel). Here, we have tiled matrices mapping from the correct color (red, green, blue) to the visual outcome (red, green, blue, gray) for each location sampled (columns) and rule (rows). This arrangement reveals the contingencies generating outcomes. For example, if the agent is looking at the central location, she will see a unique color under each rule (middle column: red, green, and blue for




each of the three rules). Although the color sampled at the central location signifies the rule for this trial, the subject has no concept of what this rule means (see the uniform priors on either side of the central fixation in Figure 3). This means that the subject believes, a priori, there is no relationship between the correct color and the color observed.

This completes our specification of the subject's generative model. An important aspect of this formulation is that we were able to transcribe task instructions or intentional set into prior beliefs. This suggests that one can regard task instructions as a way of inducing prior beliefs in an experimental setting. After instilling these prior beliefs, the synthetic subject knows quite a lot about the structure of the problem but nothing about its solution. In other words, she knows the number of hidden states and their levels and mappings between some hidden states and others. However, this knowledge is not sufficient to avoid surprising outcomes: making mistakes. Notice that we have been able to specify a fairly complete model of a paradigm, including where subjects look, when they expect to respond, and the sensory modalities entailed. This may appear to be overkill; however, it allows us to make specific predictions about behavior that can be tested empirically. Furthermore, it shows how purposeful, epistemic behavior can emerge under minimal assumptions. In what follows, we will see abstract problem solving and rule learning emerge from the minimization

of expected free energy (i.e., expected surprise or uncertainty) under prior beliefs that make indecisive or erroneous choices surprising.

**3.1 The Rule.** Hitherto, we have just specified the generative model used by an agent. Clearly, to generate outcomes, we have to specify the true generative process. This is identical to the generative model with one exception: the mapping from states to outcomes contains the causal structure or rule that the subject will learn. As noted, the actual rule used to generate outcomes is as follows: the rule (left, center, right) specifies the location of the correct color. For example, if the subject sees red in the center, the correct color is on the left. However, if she sees green in the center, the correct color is in the center, which is always green. The corresponding part

Figure 3: Graphical representation of the generative model (Left) The Bayesian network shows the conditional dependencies implied by the generative model in Figure 1. The variables in open circles constitute (hyper) priors, while the blue circles contain random variables. This format shows how outcomes are generated from hidden states that evolve according to probabilistic transitions, which depend on policies. The probability of a particular policy being selected depends on its expected free energy. The left panels show the particular hidden states and outcome modalities used to model rule learning. Here, there are three output modalities comprising colored visual cues (*what*), proprioceptive cues signaling the direction of gaze (*where*), and (auditory) cues providing feedback (*feedback*). These three sorts of outcomes are generated by the interactions among four hidden states or factors: an abstract rule indicating the location of an informative color cue (*rule*); the correct color (*colour*), where attention or saccadic eye movements are directed (*where*); and a (manual) response (*choice*). Hidden states interact to specify outcomes in each modality. In other words, each combination of hidden states has an associated column in the likelihood array that specifies the relative likelihood of outcomes in each modality. For example, if the rule is *left*, the correct *color* is red, and the subject is looking at the left cue, the *what* outcome will be *red* and the *where* outcome will be *left*. (Right) The panel on the upper right shows an example of a trial, where a subject looks from the starting position to the central location, sees a red cue, and subsequently looks to the left. After she has seen a green cue, she knows the correct color and returns to the start position, while indicating her choice (*green*). The "?" denotes an undecided choice state (and feedback). The matrices (lower left panel) show the likelihood mapping between hidden states and (*color*) outcomes assumed, a priori, (right) and used to generate actual outcomes (left). These matrices show the likelihood mapping from hidden states to *what* outcomes—the **A** array for the first or *what* modality. This is a five-dimensional array, of which four dimensions are shown under the *undecided* level of the *choice* factor. In other words, these are the contingencies in play until a decision is made. These parameters are shown as a block matrix with $3 \times 3$ blocks (*rule* times *where*). Each block shows the $4 \times 3$ matrix mapping the correct color to the outcome. (See Figure 7 for further details.)

of the likelihood mapping (under the undecided state) is shown in Figure 3 (left inset panel). In contrast to the prior beliefs, the true likelihoods mean that if one is looking to the left, the observed color maps to the correct color under the left rule, and similarly when looking to the right. This true likelihood also contains some redundancy. For example, if the subject looks at the central cue when the correct color is red, she will still see a green cue. However, this combination of hidden states never occurs because, a posteriori, a central green cue means the correct color is green (and this is encoded in the likelihood mapping under decided states). It should be noted that the (synthetic) subject does not "know" about these contingencies in an explicit or even subpersonal sense. These contingencies are encoded in model (concentration) parameters that can be associated with synaptic efficacy or connection strengths in the brain.

The rule above may sound simple, but it introduces interesting context sensitivity or interactions among the hidden factors causing outcomes. For example, the outcome depends on a two-way interaction between the correct color and where the subject is looking, but only when the rule is right or left. The subject has to learn these contingencies by accumulating coincidences of inferred states and outcomes. However, this is not a simple problem because agents do not have complete knowledge about hidden states and therefore do not know which states are responsible for generating outcomes. Imagine you had to identify these interactions by designing a multifactorial experiment. You would then manipulate hidden states and record the outcomes. However, there is a problem: you do not know the hidden states (because they are hidden from direct observation). In other words, our synthetic subject has to learn the parameters, while inferring the hidden states. So how does she fare using active inference?

Figure 4 shows the results of simulating 32 trials, where each trial comprises six epochs in which the subject can sample a new cue or make a choice. The upper two panels summarize performance in terms of posterior expectations over policies (top panel) and the final outcomes (second panel). The colored dots indicate the rule for each trial (upper panel) and the final outcomes (second panel). A key point to observe is that the final outcome is usually correct. This is because the subject is allowed to change her mind after making a mistake. These choices are based on posterior expectations about policies that are initially ambiguous and become more precise with learning. By the end of each trial, only the last three policies are entertained (choosing red, green, or blue). In the first trial, two options are entertained with equal probability, but by the 10th trial, any ambiguity appears to be resolved. After the 14th trial, performance becomes perfect. Although there is no definitive phase transition or aha moment, these results suggest that the rule is dawning on the agent.

The lower panels illustrate the implicit transition from ignorance to understanding after trial 14 (highlighted in blue). The second and third panels

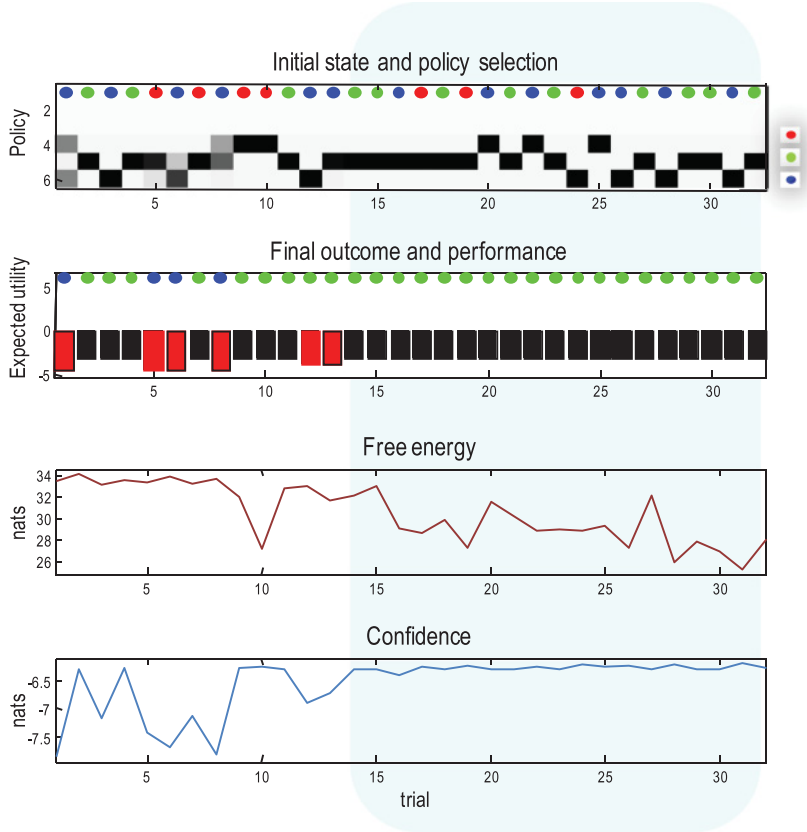



Figure 4: Simulated responses during learning. This figure reports the behavioral responses during 32 successive trials. The first panel shows the first (*rule*) hidden state as colored circles and subsequent policy selection (in image format) over the policies considered. Darker means more probable. There are six policies corresponding to a saccade to each of the three locations (without making a choice) and three choices (while moving back to the starting location). These policy expectations reflect the fact that at the end of the trial, a choice is always made with greater or lesser confidence, as reflected in the relative probability of the final three policies. The second panel reports the final outcomes (encoded by colored circles) and performance measures in terms of expected cost (see equation 3.1 and Table 1), summed over time (black bars). The red bars indicate mistakes (i.e., the incorrect color is chosen at some point).The lower two panels report the free energy at the end of each trial and fluctuations in confidence as learning proceeds. The cyan region indicates the onset of confident and correct responses.

show the free energy over trials and associated confidence in behavior: $\boldsymbol{\pi} \cdot \widehat{\boldsymbol{\pi}}$ (i.e., the negative entropy of beliefs about policies, where entropy scores uncertainty). Note that free energy is the difference between accuracy and complexity. The increase in confidence therefore reflects a dialectic between complexity and accuracy. Here, the increase in confidence (decrease in entropy over policies) is more than offset by the increase in accuracy afforded by confident behavior. This is reflected by the progressive decrease in free energy that, unlike confidence, shows trial-to-trial fluctuations. The persistent increase in confidence is underwritten by epistemic behavior that resolves both uncertainty and ambiguity.

Figure 5 shows the expectations over states and action as a function of epochs within the first trial (see Figure 5A) and the penultimate trial, after rule learning (see Figure 5B). The four panels on the left show the expectations over the four marginal hidden states, while the two panels on the right show the equivalent expectations over action. Note that there are two actions that control transitions among hidden states: the *where* factor and the *choice* factor. The cyan dots show the true hidden states and action selected. In the first trial, the agent first looks to the center, then looks to the left, stays there for one epoch, and then makes two incorrect choices. Conversely, in the later trial, the agent looks at the center, and the right and then chooses correctly.

The key point these results illustrate is the apparent attractiveness of right and left locations after the first saccade, which disappears in later trials. It is this behavior that is driven by novelty (see equation 2.2). To understand the importance of this behavior, it is useful to realize that the right and left locations are inherently aversive because they deliver ambiguous outcomes. Normally, an agent would avoid these locations in the same way that you and I might avoid a noisy restaurant or ambiguous invitation. However, the naive agent does not know these locations are ambiguous—and this is a known unknown that affords an opportunity for the agent to fill in her knowledge gaps. Crucially, after rule learning, the subject knows that the left location is ambiguous and avoids it (compare the probabilities in the upper right panels in Figures 5A and 5B). She therefore looks immediately to the informative location, enabling her to retrospectively infer that the correct color is blue (compare the upper left panels in Figures 5A and 5B). This inference is based on Bayesian belief updating, which we now consider more closely.

**3.2 The Neural Correlates of Cognizance.** A close inspection of the (synthetic) neuronal updating—underwriting the behavior above—shows a profound difference in the temporal structure of evoked responses. Figure 6 shows the activity of units encoding the expectation of the hidden color state over six epochs as a function of time. The results are shown for the same (before and after learning) trials of the previous figure. The upper

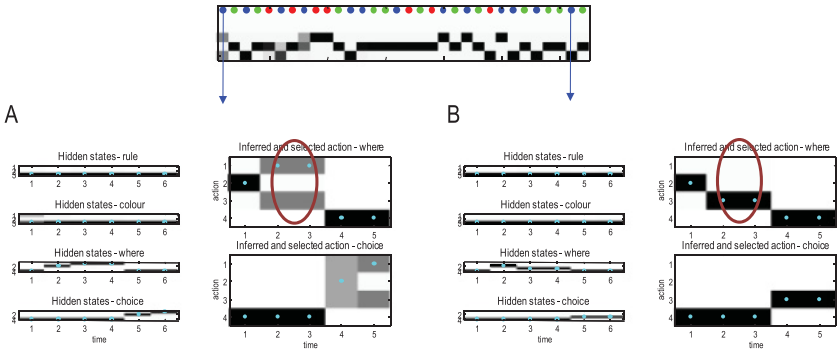



Figure 5: Simulated exploration. This figure reports the belief updating behind the behavior shown in Figure 4 for an early trial (A) and after the rule has been learned (B). (A) Each panel shows expectations in image format, with black representing 100% probability. For the hidden states (left panels), each of the four factors is shown separately, with the true states indicated by cyan dots. Here, there are five saccades, and expectations are shown after completion of the last saccadic, which means that, retrospectively, the agent believes it started in a right rule context (hidden states—rule) and, despite making two mistakes, is able to infer the correct color by elimination (hidden states—*color*). The two panels on the right report the equivalent expectations about action for the two controllable hidden states (*where* and *choice*). The sequence of sampling (inferred and selected action—where) indicates that the subject first interrogated the center (observing a blue cue), looked to the left, and then made two incorrect choices (see Figure 6). (B) However, after learning, the subject is much more confident about where to look because she now knows that the color of the right cue will reduce uncertainty about the correct color. The important aspect of these results is that prior to learning, the right and left locations are equally attractive—and more attractive than the (initially sampled) central location (highlighted with red circles). This is despite the fact these locations do not reduce risk or uncertainty (because the agent does not know the meaning of the cues in these locations). However, the subject does know that she is ignorant and can resolve this ignorance by exposing herself to novel outcomes.

panels show belief updates as a raster image (left) and as functions of peristimulus time (right). The raster highlights the fact that there are explicit representations of the six epochs at each point in time and that these expectations are updated over time. This means that the blocks of the raster above the leading diagonal encode the past, while the blocks below the leading diagonal encode the future.

The key observation here is that the onset of discriminatory responses is much earlier after learning than before. This is due to—and only to—learning the mapping between hidden states and consequences, enabling

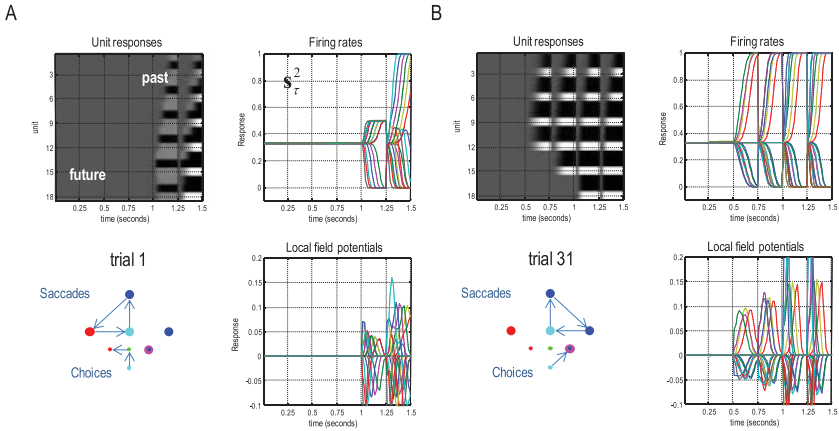



Figure 6: Simulated electrophysiological responses. This figure shows expectations about hidden states for the same two trials illustrated in the previous figure—before learning (A) and after learning (B). (A) The upper left panel shows the activity (firing rate) of units encoding the correct color in image (raster) format over the six intervals between five saccades. These responses are organized such that the upper rows encode the probability of alternative states in the first epoch, with subsequent epochs in lower rows. In other words, the top row shows the expectations about the three hidden (*color*) states at the beginning of the trial and how these expectations evolve over time. Conversely, the first column shows expectations about the three colors at successive time points in the future. The plot to the right of the image presents the same information to illustrate the evidence accumulation and the resulting disambiguation of context. These values are expectations about hidden states described by the first equations in Figure 2 and can be interpreted as neuronal firing rates of units encoding expectations. The associated local field potentials for these units (i.e., the rate of change of neuronal firing) are shown in the lower plot. The insert (lower left) shows the sequence of moves and decisions. Here, the subject makes a saccade to the center location and then looks to the left and finally back to the center, at which point she makes a choice (green), which elicits the wrong feedback. She then changes her mind and (incorrectly) tries red. However, the correct color is blue (circled in magenta). This behavior can be contrasted with the responses on the right (after learning). (B) Here, the correct color is identified on the first choice. Crucially, this is based on precise expectations about the correct color that have been accumulating since the second saccade, as reflected in the simulated neuronal responses.

the agent to infer the correct color after the second saccade to the (informative) location. The lower panels show the corresponding behaviors (left) and simulated local field potentials or event-related potentials (right). These are simply the first derivatives of the responses in the upper panels. In

summary, after the rule has dawned on the agent, she knows exactly where to find unambiguous information to make precise inferences about the underlying context and veridical choices. In terms of simulated electrophysiology, this means representations of latent states of the world are activated much earlier during evidence accumulation, after the meaning of cues has been disambiguated through epistemic learning. These simulations are not inconsistent with event-related potential and fMRI studies of insight, reviewed in the discussion (e.g., Jung-Beeman et al., 2004; Mai, Luo, Wu, & Luo, 2004; Bowden et al., 2005).

## 4 Structure Learning and Bayesian Model Reduction

> The second question is why, out of the infinite range of knowable items in the universe, certain pieces of knowledge are more ardently sought and more readily retained than others (Berlyne, 1954).

The previous section illustrated the role of novelty in driving curious behavior and the epistemic learning it elicits. In this section, we turn to a different sort of learning: learning the structure of a likelihood model after evidence has been accumulated. The previous simulations suggest that there are behavioral and electrophysiological correlates of curiosity in epistemic learning. However, there was no clear homologue of an aha moment or an instance of revelation associated with insight. To move closer to "the perception of what passes in a man's own mind," this section considers Bayesian model selection and structure learning as a metaphor of understanding and self-knowledge, in the sense of Locke (Nimbalkar, 2011). It considers the fact that subjects not only have prior beliefs about the parameters of their models but also prior beliefs about models per se; for example, they know there are rules. In what follows, we will see that this prior knowledge about models naturally induces abductive reasoning, when the models themselves minimize variational free energy.

Loosely speaking, one can associate awareness of the world with inference about its hidden states based on a generative model and knowledge with learning model parameters. Here, we turn to a third level of optimization that minimizes free energy with respect to the model per se. Selecting models that have the greatest evidence (least free energy) is known as Bayesian model selection. This procedure furnishes models that, on average, provide the best explanation for the data at hand. As such, it can be thought of as inference to the best explanation (Harman, 1965), or abductive reasoning.

We focus on a particular but ubiquitous form of Bayesian model selection; Bayesian model reduction. Essentially, Bayesian model reduction evaluates the evidence of reduced forms of a parent or full model by eliminating redundant parameters. Crucially, Bayesian model reduction can be applied to the posterior beliefs after the data have been assimilated. In other words,

Bayesian model reduction is a post hoc optimization that refines current beliefs based on alternative models that may provide potentially simpler explanations (Friston & Penny, 2011). The alternative (reduced) models are defined in terms of their priors—for example, a precise prior belief that some parameters are zero. Heuristically, equation 2.1 shows that free energy is a complexity minus accuracy, where complexity is the divergence between posterior and prior beliefs. Previously, we have focused on optimizing free energy with respect to the (approximate) posterior as encoded by its expectations. However, we can also minimize free energy with respect to the priors, thereby eliminating redundant parameters to reduce model complexity.

Neurobiologically, this model optimization resembles mechanisms that have been proposed during sleep. While awake, the brain learns causal associations, through associative plasticity, that are embodied in an exuberance of synaptic connections. During sleep, redundant connections are subsequently removed (Tononi & Cirelli, 2006) to minimize complexity and free energy in the absence of any further sensory input (Hobson & Friston, 2012). In this setting, sleeping is literally a way of clearing one's mind.

Technically, Bayesian model reduction is a generalization of ubiquitous procedures in statistics, ranging from the Savage-Dickey ratio (Savage, 1954), through to classical $F$-tests in parametric statistics. In our context, it reduces to something remarkably simple: by applying Bayes' rules to full and reduced models, it is straightforward to show that the change in free energy can be expressed in terms of posterior concentration parameters $\mathbf{a}$, prior concentration parameters $a$, and the prior concentration parameters that define a reduced or simpler model $a'$. Using $\mathrm{B}(\cdot)$ to denote the beta function, we get (see the appendix)

$$\Delta F = \ln \mathrm{B}(\mathbf{a}) + \ln \mathrm{B}(a') - \ln \mathrm{B}(a) - \ln \mathrm{B}(\mathbf{a} + a' - a). \tag{4.1}$$

This equation returns the difference in free energy we would have observed had we started observing outcomes with simpler prior beliefs. Clearly, to engage with this form of free energy minimization, one has to have a space of models or reduced priors to evaluate. This is the key feature of Bayesian model reduction that lends it an abductive aspect: in other words, model selection is ampliative, meaning that the conclusion goes beyond what could otherwise be induced or inferred. This abductive characteristic rests on the selection of competing hypotheses that are plausible. In other words, if I know that my data were produced like this or like that, I can appeal to a relatively small number of plausible explanations and implicitly exclude a universe of alternative explanations (models), including the full model used to acquire my knowledge. This ampliative aspect of Bayesian model selection appeals to a hierarchical structuring of plausible explanations such that each successive level provides broad (or abstract) constraints on

plausible explanations for the level below; for example, at the highest level, we may know there are a small number of plausible hypotheses and a large number of implausible hypotheses. (See Navarro & Perfors, 2011, for a discussion of this key issue in the context of sparse hypothesis or model spaces.)

To make this process clear, consider the following example. If our subject knows she is being asked to discover a rule, she knows that some combinations of hidden states for each factor will be informative and others will not. Therefore, under any particular combination of hidden states, there are only two plausible contingencies: either all allowable outcomes are equally probable, or there is a definitive outcome that constitutes part of the rule. One can take this explanatory reduction (of model or hypothesis space) even further based simply on the assumption that rules entail some form of symmetry or invariance. For example, if the correct color red always generates a red outcome, other levels of the same (*color*) hidden state will be informative under a particular combination of the other hidden states.

This heuristic can be absorbed gracefully into the imperative to minimize expected free energy. This is because expected free energy scores the ambiguity of a generative model. Therefore, prior beliefs about models based on their expected free energy will necessarily favor unambiguous mappings between (latent) causes and consequences. In other words, in exactly the same way that action selection minimizes ambiguity, when agents are equipped with the latitude to optimize their models, model selection is restricted to models whose plausibility is determined by their ability to disambiguate the causes of observed outcomes.

Figure 7 shows the results of applying Bayesian model reduction before the 12th trial in the simulations above. In this example, we compared the evidence for a full model (in which the correct color generated outcome colors of equal probability) with reduced models in which the correct color generated its own color (under each combination of the remaining hidden states). These reduced models were specified by changing the prior counts from 1 to 8 to specify a prior belief that the hidden color specified the outcome more precisely. A comparison was performed under every combination of the remaining (three) hidden states. If there was positive evidence for the reduced, simpler, or unambiguous model, the concentration parameters mediating uninformative outcomes were removed (by setting them to zero) and the posterior concentration parameters (or counts) were assigned to the remaining parameters.

The upper left panel of Figure 7 reproduces the true likelihood from Figure 3. The upper right panel shows the full model that corresponds to the agent's prior beliefs. The equivalent posterior beliefs after 12 trials are shown on the lower left. It is clear that the leading diagonal blocks of the array may be better explained by an unambiguous one-to-one mapping between the hidden and outcome colors. Indeed, when we performed Bayesian model comparison, the reduced model had more evidence than

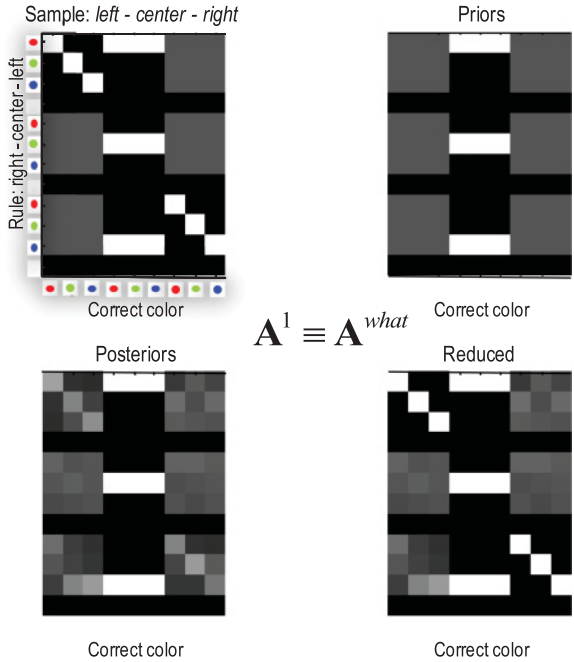



Figure 7: Structure learning. This figure shows the parameters (connection strengths) that constitute the likelihood mapping from hidden states to *what* outcomes—the **A** array for the first modality. This is a five-dimensional array, of which four dimensions are shown under the undecided level of the *choice* factor. These parameters are shown as a block matrix with $3 \times 3$ blocks (*rule* times *where*). Each block shows the $4 \times 3$ matrix mapping the correct color to the outcome. (Upper left) These represent the true parameters or contingencies. When the agent is looking at the center cue (middle column), the rule is uniquely specified by the color of the outcome. However, when the agent is looking to the left or right (left and right columns), the outcome depends on only the correct color if the agent is looking toward the left (when the rule is *left*) or to the right (when the rule is *right*). This context sensitivity is modeled by the diagonal matrices on the upper left and lower right. In all other situations, any color could be seen. (Upper right) These are the corresponding expectations of a naive agent. Note that the diagonal matrices have disappeared and there are no beliefs about the relationship between the correct and observed colors. (Lower left) After 12 trials, the agent has accumulated sufficient experience to acquire knowledge about the (context-sensitive) interactions and knows that observed and correct colors predict each other when, and only when, looking in the appropriate location. (Lower right) This knowledge is sufficient to recover the correct contingencies following Bayesian model reduction (a period of sleep or reflection). Note that the implicit model optimization removes redundant parameters (connections) between the correct and observed colors, enabling more confident behavior.

the full model, leading to the posterior parameterization on the lower right. This has correctly removed redundant (off-diagonal) parameters from the top left and bottom right blocks, thereby equipping the agent with the correct prior belief that when looking at the right or left locations (prior to making a choice), the correct color unambiguously specifies the color that will be seen.

Note that Bayesian model reduction or top-down structure learning is not just a question of finding simple models with unambiguous contingencies; it rests on finding the best balance between accuracy and complexity. This means model parameters will be pruned until further model reduction compromises the model's ability to explain accumulated evidence. This is evident in the current simulations, which eschew very simple models that are not fit for purpose in relation to the contingencies generating outcomes (e.g., models with unambiguous outcomes in all contexts). This reflects Einstein's famous assertion: "Everything should be made as simple as possible, but not simpler."

Figure 8A shows the ensuing performance after Bayesian model reduction. This uses the same format as the lower panels in Figure 4, with the free energy in the upper panel and confidence in the lower panel. The dotted lines reproduce the results from Figure 4, while the solid and broken lines show the improvement following Bayesian model reduction. Indeed, performance becomes perfect (and confident) after trial 12. The difference between the solid and broken lines rests on an additional optimization that speaks to the difference between REM (rapid eye movement) and non-REM sleep, which we now consider.

**4.1 Bayesian Model Reduction and Sleep.** In the example above, we eliminated redundant parameters when there was positive evidence for the reduced model. Quantitatively, this corresponds to $\Delta F \leq -3$. This corresponds to an odds ratio or Bayes factor of $\exp(-3) \approx 0.05$. In other words, strong evidence for the reduced model relative to a full model means that the reduced model is about 20 times more likely (Penny, 2012). Neurobiologically, removing redundant parameters corresponds to a synaptic regression of the sort implicated in synaptic homoeostasis (Tononi & Cirelli, 2006). For example, if we consider the concentration parameters as encoding synaptic efficacy, connections mapping from expectations of hidden states to predicted outcomes are eliminated with a nuanced winner-take-all-like mechanism. Heuristically, if one connection is sufficiently larger than all others, then the other connections are lost. Otherwise, all connections remain in play. One could associate this synaptic regression with the homoeostatic mechanisms thought to occur during non-REM sleep (Tononi & Cirelli, 2006). What about REM sleep?

Because we have eliminated model parameters, it is necessary to reevaluate the posteriors under the new (reduced) priors. Usually this would be done analytically (see the appendix). However, here, we have not simply

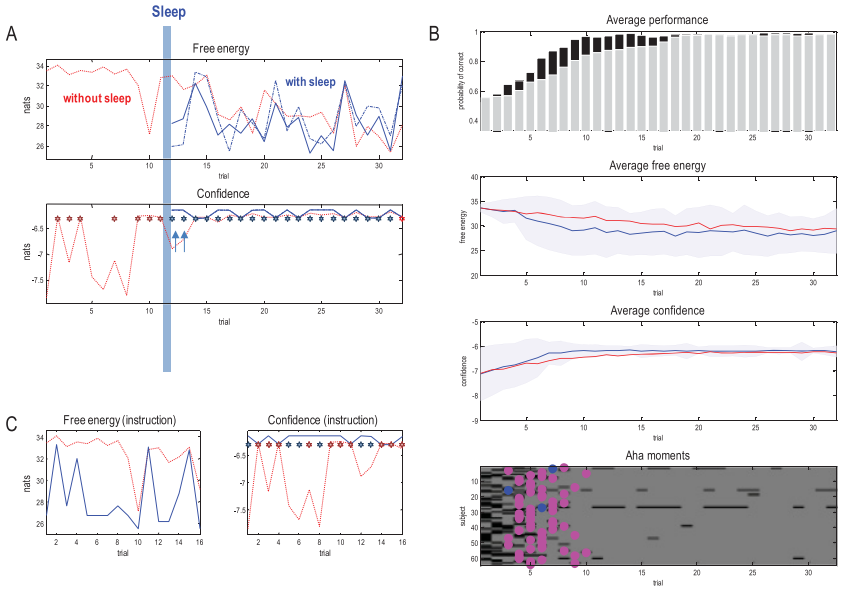



Figure 8: Performance and sleep. This figure illustrates the performance over 32 trials with and without Bayesian model reduction (or sleep), and instruction. The results of a single simulation are shown in panel A, while panel B summarizes the results over 64 agents. (A) The upper panels show performance using the same format as in Figure 4 in terms of free energy (upper panel) and confidence (second panel). The blue lines report the simulations with sleep—REM (solid) and non-REM (broken)—while the red lines are without sleep (and recapitulate the results in Figure 4). The key thing to note here is that a period of sleep (before the 12th trial) markedly improves performance over subsequent trials. This is indicated by the arrows highlighting mistakes made in the absence of sleep (the pink stars are correct responses prior to sleep, while the blue stars indicate correct responses after sleep). (B) These results illustrate performance averaged over 64 subjects, each performing 32 trial sessions. The upper right panel shows the probability of a correct response (pooled over successive trials), where the gray bars correspond to normal performance and the black bars show the improvement when subjects engage in Bayesian model reduction (reflection) after each trial. The middle panels show the average free energy and confidence over subjects with (blue) and without (red) reflection. The shaded blue area corresponds to the 90% confidence interval. The lower panel shows the corresponding results for each subject by scoring their mistakes (black bars) as a function of trials with Bayesian model reduction. The dots record points of abduction or Bayesian model reduction (veridical, magenta; superstitious or incorrect, blue). (C) These panels show the free energy and confidence with (blue) and without (red) instruction (i.e., with and without informative priors over models). The right panel shows that performance is perfect following instruction with a high level of confidence (pink stars). The blue stars indicate when mistakes were made without instruction.

increased the prior of one connection over others; we have actually removed connections. This means it is necessary to reevaluate the posterior concentration parameters under the new model. One obvious way of doing this is to rehearse the same sequence of outputs under the new model. This means that the process of Bayesian belief updating its repeated but using outcomes that have already been sampled or generated de novo with the same statistics (Louie & Wilson, 2001; Dragoi & Tonegawa, 2014; Pezzulo, van der Meer, Lansink, & Pennartaz, 2014). Neurobiologically, this implies several processes. First, we have to update expectations about hidden states (e.g., encoded in hippocampal and parietal systems) in the absence of new sensory information. Furthermore, this optimization rests on active inference and, in this example, saccadic eye movements. In short, we have a mathematical metaphor for hippocampal dependent learning (Walker & Stickgold, 2004) during sleep (Rasch, Buechel, Gais, & Born, 2007) with an emphasis on procedural memory in REM sleep (Marshall & Born, 2007).

The solid lines in Figure 8A correspond to performance using posterior parameter estimates (synaptic connection strengths) after sleep-like reevaluation. The broken lines show a (similar) performance when we approximate this process by assigning posterior concentration parameters to the surviving connection (as above). It can be seen that there is no difference in terms of performance or confidence, with no systematic difference in the free energy or expected free energy. We therefore use the (computationally more expedient) approximation in what follows.

**4.2 Simulating the Aha Moment.** The previous simulations explored the notion that Bayesian model reduction and the minimization of complexity might occur during sleep. This "clears the mind" such that after a brief nap, our subject has a clear and simple understanding of the contingencies and can choose more efficiently and confidently. Crucially, this Bayesian model reduction introduces a qualitative change before and after sleep that speaks to a qualitative transition to a state of knowing (and using what you know) associated with an aha moment. In what follows, we use Bayesian model reduction to redistribute posterior concentration parameters as above; however, here we associate this post hoc model optimization not with sleep but with reflection. In brief, we allowed the subject to reflect on each trial, thereby performing online Bayesian model reduction. The aim of these simulations was to show that model selection can be implemented continuously (as opposed to sporadically during sleep) and leads to qualitative transitions in the generative model and subsequent inference. It is these transitions we associate with aha moments. Indeed, in the rodent hippocampus, one sees (sharp wave-ripple associated) sequences not only during sleep but whenever a rat is disengaged from exploratory or foraging behavior (and associated theta rhythms), such as during short rests or grooming after consuming a reward (Pezzulo et al., 2014).

Figure 8B shows the results of repeating the above simulations with 64 artificial subjects, each exposed to a different sequence of rules and cues. The upper panel shows the results with and without abductive Bayesian model reduction following each trial, in black and gray, respectively (the performance has been slightly smoothed to reveal the underlying trends). One can see a marked improvement in performance, particularly after the fifth trial. The benefits of abductive reduction mean that nearly all subjects attain 100% performance at around the 10th trial (see below). Chance performance here is about 30%. The associated changes in free energy and confidence are shown in the middle panels in terms of the average over subjects (blue line) and 90% confidence intervals based on the standard deviation over subjects. The red lines show the averages without reflection.

The lower panel in Figure 8B shows individual performances, with each subject along one row and mistakes shown in black. The colored dots correspond to qualitative changes in the model following abductive Bayesian model reduction. The magenta dots record a veridical model reduction using a model space that allowed unambiguous (one-to-one) mappings between the correct color and outcome that were deployed unambiguously over combinations of remaining hidden states. In detail, we considered models in which the correct color specified a unique outcome. Similarly, this mapping was specified uniquely by each rule over locations. This created 36 models where each of six different (unique) mappings between the three correct colors and outcomes was arranged according to six different (unique) combinations of rule and location. Because all of these models provide an unambiguous explanation for causal structure in the paradigm, they enable ampliative inference, enabling rules to be "recognized" after a handful of trials. This exploration of model space is slightly more exhaustive and sophisticated than the sleep example above. Here, we compared 36 ($6 \times 6$) models with the full (ambiguous) model, as opposed to comparing ambiguous and unambiguous (*what* $\times$ *color*) models for each of the 36 ($3 \times 3 \times 4$) combinations of hidden states (*rule* $\times$*where* $\times$ *choice*).

The blue dots (in three subjects) correspond to decreases in the correlation between the priors and true model structure that occur when an informative color mapping is abducted in the wrong context, for example, when looking to the left under the center rule (in which case the outcome is uninformative). These *superstitious* models illustrate the trade-off between the ampliative benefits of abduction (going judiciously beyond the evidence at hand) and a susceptibility to superstitious beliefs that arise from chance occurrences that are consistent with prior beliefs. These false insights are associated with persistently poor behavior (see Figure 8B).

These results also show profound intersubject variability. Many subjects attain perfect performance after a handful of trials and early aha moments; however, some agents experience aha moment only after 10 trials. Strictly speaking, this is not an attribute of agents; it is an attribute of the outcomes sampled. This follows from the fact that all the agents are identical, starting

with the same priors and responding deterministically in their active inference. This highlights a key aspect of active inference: there are no random or stochastic aspects. Everything that changes does so to minimize variational free energy, in accord with Hamilton's principle of least action. This should be contrasted with sampling schemes and reinforcement learning that would take a very long time to learn this rule (see section 5).

A key prediction arises from associating Bayesian model selection with abductive processes: an aha moment is necessarily subpersonal (i.e., pertaining to a biophysical level below the personal or conscious level). In other words, one can never remember or articulate abductive reasoning at the level of the model in question. This is because optimizing a model is fundamentally different from modeling an optimization. This subpersonal aspect is consistent with the physiological mediation of the abductive reasoning afforded by Bayesian model selection during sleep (and reflection or mind wandering) (Bar, Aminoff, Mason, & Fenske, 2007), which is mediated by synaptic regression and competition (e.g., (Holzel et al., 2011). In other words, when things click into place, we have no explicit (generative) model of the underlying process. Having said this, it is possible that agents have "in mind" an explicit model space that they will test from time to time. This introduces a further hierarchical level to generative models, where the higher level constitutes a space of alternative models or hypotheses. In the last example, we used 36 models. An interesting corollary of possessing a portfolio or lexicon of potential models is that the results of abductive reasoning can be communicated to other agents, in terms of prior beliefs about models, provided all agents possess the same the lexicon or model space (Friston & Frith, 2015). We illustrate this briefly in the final simulation.

**4.3 Model Reduction and Communication.** We have emphasized the autodidactic nature of structure learning with Bayesian model reduction. However, exactly the same (variational) optimization principles apply when using reduced models as priors in other naive agents. In other words, an important aspect of inference over models is that the conclusions can be transcribed or communicated to update the prior beliefs of others, such as children or other naıve conspecifics (Frith, 2010). Note that this is not instructed or supervised learning in the sense that transcribing reduced priors does not tell a naıve agent what has been learned—just what is learnable. In other words, communicating posterior beliefs about models is distinct from communicating posterior beliefs about model parameters.

To illustrate the potency of good prior knowledge in a multi-agent or prosocial setting, we created a new agent and equipped it with the reduced priors of an experienced agent (from Figure 4). Unsurprisingly, the naive subject performed perfectly, with maximum confidence from the first trial (see Figure 8C). This perfect performance does not mean the agent has consolidated its received wisdom (accumulated concentration parameters), but her inferences are sufficiently precise to enable prior beliefs (about

responding correctly) to be fulfilled. This simple illustration does not address some more fundamental issues of neuronal hermeneutics (Frith & Wentzer, 2013). This is because we have assumed that the priors from the experienced agent can be received by the naıve agent. This presupposes that there is a shared space of hypotheses and model spaces that enable the naive agent to properly implement received priors. This raises the interesting issue of what we communicate to each other: Is this knowledge or meta-knowledge conveyed in the form of prior beliefs (Shea et al., 2014).

**4.4 Summary.** In summary, we have created a synthetic subject that has all the hallmarks of a "good scientist." She starts off with prior beliefs that she will, after a period of sampling data, report her conclusions—and not be wrong. These are the only beliefs necessary to specify behavior; everything else follows from minimizing expected free energy. First, (sensory) data are sampled to resolve ambiguity about the current state of the world while at the same time reducing the risk of making a mistake. Furthermore, experiments are performed carefully to resolve ignorance about how outcomes are generated by latent or hidden causes. Having acquired data, the "good scientist" reflects on what she knows (and perhaps sleeps on it), implicitly testing plausible hypotheses of a progressively simpler (less complex and less ambiguous) nature that could provide an accurate account of the data at hand. Equipped with a more generalizable (generative) model, which conforms to Occam's principle, the active inference process starts again, providing more evidence that the agent is what she thinks she is: a good scientist (cf. the self-evidencing brain; Hohwy, 2016).

## 5 Discussion

We have presented an active inference formulation of curiosity (epistemic) and insightful (structural) learning in the setting of abstract rule learning. The resulting behavior rests on two generalizations of free energy minimization: the first is the inclusion of beliefs about model parameters in the expected free energy that induces novelty. This leads to a form of epistemic behavior (curiosity) that goes beyond resolving uncertainty about the context in which an agent is operating, to resolving ignorance about whether uncertainty can be resolved. The second advance pursues a hierarchical theme by noting that free energy can also be reduced online (reflection) or offline (sleep) through Bayesian model reduction. Together, these processes produce inferential behavior that acquires knowledge of lawful structure in the sensorium and points to the role of self-improvement and (subpersonal) introspection in making the most of that knowledge.

**5.1 The Empirical Correlates of Insight.** From a neuroscience perspective, our simulations suggest the following correlates of insight:

- A sustained increase in performance accuracy after epistemic learning
- A concomitant increase in posterior confidence about action or choices
- A predilection for sampling novel cues, which abates with experience
- A profound reduction in the latency of evoked neuronal responses, when subjects know the meaning of cues (or have learned a rule); equivalently, an increase in the amplitude of event-related responses to initial cues in a sequence, when their implications for subsequent outcomes can be inferred (see Figure 6)
- An increase in the above markers of awareness when agents are allowed to reflect on their choices after each trial

Perhaps the empirically most relevant prediction is the reduction in the latency of evoked responses following the acquisition of insight. Several studies speak to this prediction. For example, event-related potential studies of Chinese riddles (with and without insight) implicate the anterior cingulate cortex: "Dipole analysis localized the generator of the N380 in the ACC. N380 therefore probably reflects an 'Aha!' effect, and the ACC generator may be involved in the breaking of mental set" (Mai et al., 2004). It is tempting to associate the notion of "breaking of mental set" with the process of Bayesian model reduction in the sense that both entail a dissembling of prior constructs in the search of better explanations. Complementary fMRI studies of verbal insight have "revealed increased activity in the right hemisphere anterior superior temporal gyrus for insight relative to noninsight solutions" (Jung-Beeman et al., 2004). These findings speak to a hemispheric lateralization of the neuronal correlates of insight; see Bowden et al., 2005, for discussion. Interestingly, "The same region was active during initial solving efforts. Scalp electroencephalogram recordings revealed a sudden burst of high-frequency (gamma-band) neural activity in the same area beginning 0.3 s prior to insight" (Jung-Beeman et al., 2004). Findings of this sort suggest that there may be neuronal correlates of (subpersonal) reflection that rest on short-term plasticity in regions whose intrinsic connectivity encodes likelihood mappings.

The short-term plasticity implicit in Bayesian model reduction fits nicely with the role of restructuring in insight. Insight entails "a mental restructuring that leads to a sudden gain of explicit knowledge allowing qualitatively changed behavior. Anecdotal reports on scientific discovery suggest that pivotal insights can be gained through sleep. Sleep consolidates recent memories and, concomitantly, could allow insight by changing their representational structure" (Wagner, Gais, Haider, Verleger, & Born, 2004). Using a cognitive task that required the learning of stimulus-response sequences, Wagner et al. demonstrated an abrupt improvement in performance (reaction times) following insight into an abstract rule underlying the sequences. Crucially, the prevalence of insight-dependent performance changes

doubled following nocturnal sleep. The authors concluded "that sleep, by restructuring new memory representations, facilitates extraction of explicit knowledge and insightful behavior" (Wagner et al., 2004). These findings suggest that there may be an interesting interplay between Bayesian model reduction in sleep (Tononi & Cirelli, 2006; Hobson & Friston, 2012) and the waking (online) processes that we have associated with aha moments. The facilitation of insight may not be limited to sleep. For example, spectral analyses of resting EEG, prior to solving anagrams, again support right-lateralized hemispheric asymmetry and "reveal a relationship between resting-state brain activity and [subsequent] problem-solving" (Kounios et al., 2008).

**5.2 Relationship to Formal Theories of Insight.** The formulation of insight offered by active inference inherits much from previous work. A common theme here is the minimization of complexity implicit in finding simpler explanations for the evidence we encounter—for example, "People may be surprised to notice certain regularities that hold in existing knowledge they have had for some time. That is, they may learn without getting new factual information. We argue that this can be partly explained by computational complexity" (Aragones et al., 2005). This approach to fact-free learning is closely related to the notion that an understanding of the world entails active learning that affords compressibility, regularity, and predictability (Schmidhuber, 2006). Technically, this compressibility corresponds to reducing complexity and speaks to the intimate relationship between variational free energy and minimum description or message lengths (Hinton & van Camp, 1993; Hinton & Zemel, 1993; Wallace & Dowe, 1999; Yedidia, Freeman, & Weiss, 2005). In brief, if (expected) free energy comprises expected accuracy and expected complexity, then the minimizing expected free energy entails the minimization of expected complexity (i.e., maximization of epistemic value), thereby furnishing more parsimonious, nicely compressed explanations for the nature of the sampled world (Schmidhuber, 2010).

A key insight here is that structure or fact-free learning proceeds, by definition, in the absence of new facts or evidence. This is important from the perspective of free energy minimization because it means the only term in play is *complexity*. In other words, in the absence of new data, the only way that we can optimize our generative models is by making them simpler.[3] This emphasizes the key role of complexity in Bayesian model selection and structure learning. We have associated insight with Bayesian model

[3] Although, as noted by our reviewers, agents could anticipate forthcoming data (or experience) and forearm themselves with more complicated (e.g., full) models. Indeed, one could argue that evolution has endowed infants with over-parameterized (e.g., full) models that are subsequently pruned through epigenetics and experience-dependent plasticity (i.e., learning).

reduction of a full model. Bayesian model reduction is a particular form of Bayesian model selection that represents a top-down approach, where alternative models are distilled from a full model, much like a sculpture is revealed by the artful removal of stone. This particular form of structure learning eludes challenging questions about how to develop models in a bottom-up fashion. In other words, we have avoided many important questions about the construction and exploration of model spaces in the absence of a full model (Gershman & Niv, 2010; Navarro & Perfors, 2011; Collins & Frank, 2013; Tervo et al., 2016). This calls on things like nonparametric Bayesian methods that have been used to model cognitive control over learning; e.g., (Collins & Koechlin, 2012; Collins & Frank, 2013) and the emergence of goal codes (Stoianov, Genovesio, & Pezzulo, 2016). Indeed, this theoretical line of thinking has enabled neuroimaging studies to identify the functional (prefrontal cortical) anatomy of structure learning in terms of "hypothesis testing for accepting versus rejecting newly created strategies" (Donoso et al., 2014).

The artificial insight in this article is exemplified within a particular experimental task. This begs the question of whether human participants would demonstrate the same sort of insight predicted by the model. For example, category learning work suggests that participants tend to use simple regularities initially but then turn to more complex constructs after training (Nosofsky & Palmeri, 1998). This speaks to interesting questions about the implicit nature of model selection—in particular, the model or hypothesis spaces from which the most apt models are selected. Technically this would be treated as a greedy search problem—for example, selecting the best among simple models and then considering simple models plus systematic (symmetric) exceptions. In our example, the simple models that conformed to putative rules contained the true model, obviating the need for a greedy search. Informal results (when asking audiences to perform the above task) suggest people acquire insight after seven to eight trials. This suggests that people are using prior knowledge about the nature of rules to perform some sort of Bayesian model selection because just accumulating experience in a Bayes-optimal fashion (without model reduction) would require about 14 trials (see Figure 4). We are currently running laboratory experiments with eye tracking on real subjects using the rule learning paradigm described above and hope to supplement these experiments with large cohort studies of reaction times and subjective reports of insight (e.g., crowd-sourcing neuroscience; Mohammadi, 2015).

**5.3 Active Inference and Artificial Intelligence.** From an artificial intelligence perspective, the rule-learning problem above would confound most conventional approaches. For example, reinforcement learning and optimal control theories are not applicable because the problem is epistemic (belief based) in nature. This means that the optimal action depends on beliefs or uncertainty. This precludes solutions based on the Bellman optimality

principle (Bellman, 1952). Although, in principle, a belief-state (partially observed) Markov decision process could be considered (Bonet & Geffner, 2014), the combinatorics of formulating beliefs states over $3 \times 3 \times 4 \times 4 = 144$ (rule, color, *where*, and *choice*) hidden states, with $4 \times 4 \times 3 = 36$ (*what*, *where*, and *feedback*) outcomes and a deep decision tree of five moves or $(4 \times 4)^5 = 1{,}048{,}576$ policies, is daunting. Furthermore, simply optimizing behavior does not address the problem of learning model parameters or structure.

If we made the problem simpler and presented all the cues instantaneously (i.e., ignored the problem of what should be sampled), it is possible that reinforcement or deep learning schemes (LeCun et al., 2015; Mnih et al., 2015) could learn the model parameters. However, there is a principled reason that such solutions may be uninteresting. If the objective is to minimize the path or time integral of free energy, then one needs to minimize the time (number of computations and samples) required to reduce free energy by a given amount. This is the motivation for trying to solve the above problems with a small number of samples. There is an interesting corollary of this variational principle (of least action); one can use the thermodynamic free energy as a proxy for variational free energy. This means that any solution that requires a large computer or extensive processing time does not conform to the variational principle of least free energy and is probably not a candidate for artificial intelligence capable of insight. In other words, if the total amount of thermodynamic energy expended during convergence to an (approximately) optimal solution is large, then the path integral of variational free energy will also be large and the path taken will therefore violate Hamilton's principle of least action (and the variational free energy principle). The relationship between thermodynamic and variational free energy is relatively straightforward to demonstrate using the Jarzynski equality (Jarzynski, 1997), which allows one to show that variational and thermodynamic free energy share the same minimum, in the limit of no sensory samples (Sengupta, Stemmler, & Friston, 2013). The upshot of this analysis is that a measure of the quality of intelligence is the simplicity and (thermodynamic) efficiency with which it can be simulated. If this argument is right, it suggests that if we want to simulate intelligence or create artificial consciousness, we should focus on the objective function and underlying variational principles as opposed to large corpuses of training data.

We started this article with an oblique reference to artificial consciousness, in the sense of shareable knowledge. In light of the treatment of insight above, in terms of Bayesian model reduction, and the utility of prior beliefs in facilitating the inference of other agents, it should be evident that the sort of consciousness we are talking about is very elementary. We have previously associated conscious processing with the process of inference (Hobson & Friston, 2014). In pursuing that theme, Bayesian model reduction and implicit structure learning represent a hierarchically deep aspect of inference; in the sense that it contextualizes inference at lower levels

of generative models (i.e., learning the parameters of a generative model and inference about hidden states conditioned on those parameters). One could argue that evolution performs Bayesian model (i.e., natural) selection (Frank, 2012; Campbell, 2016; Hobson & Friston, 2016). However, we do not generally consider evolution as a conscious process: evolution is not curious and does not profess insights. So what is special about the form of inference considered in this article that entitles us to talk about consciousness? Perhaps the simplest answer is the ability to select among competing hypotheses or models that are entertained within the same inference engine (i.e., brain or mind). In other words, the hallmark of mindful inference may be the ability to represent or entertain counterfactual hypotheses (Palmer, Seth, & Hohwy, 2015; Seth, 2015). This is not only a prerequisite for Bayesian model selection—of the sort we have associated with insight—but is also necessary for planning as inference (Attias, 2003; Botvinick & Toussaint, 2012): the selection of actions under beliefs about their consequences. Indeed, we have previously argued that the hard problem of consciousness itself (Chalmers, 1995) emerges from being able to entertain the counterfactual hypothesis that we might not be conscious (Hobson & Friston, 2014).

## Software Note

Although the generative model, specified by the $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ matrices, changes from application to application, the belief updates in Figure 2 are generic and can be implemented using standard routines (here, **spm_MDP_VB_X.m**). These routines are available as Matlab code in the SPM academic software: http://www.fil.ion.ucl.ac.uk/spm/. The simulations in this article can be reproduced (and customized) via a graphical user interface: by typing >> DEM and selecting the **rule learning** demo.

## Appendix

This appendix describes the generative model and how associated variational free energy is minimized during active inference. It is included for readers who are versed in the formalism of Markov decision processes and variational Bayes and those who want to understand the Matlab code that reproduces the simulations in the main text (see the software note). For simplicity, we deal with a single output modality and hidden factor. The generalizations for multimodal outputs and factorial states are provided in the figures.

**A.1 The Generative Model.** The generative model underlying nearly all formulations of (discrete) choice behavior can be parameterized as a Markov decision process as follows (see Table 1 for a definition of the terms in these and subsequent equations):

$$P(\tilde{o}, \tilde{s}, \pi, \mathrm{A}) = P(\pi)P(\mathrm{A})\prod_{\tau=1}^{T} P(o_\tau|s_\tau)P(s_\tau|s_{\tau-1}, \pi),$$

$$\begin{aligned} P(o_\tau|s_\tau) &= Cat(\mathrm{A}), \\ P(s_{\tau+1}|s_\tau, \pi) &= Cat(\mathbf{B}_\tau^\pi), \\ P(s_1|s_0) &= Cat(\mathbf{D}), \\ P(o_\tau) &= \sigma(-\mathbf{C}_\tau), \\ P(\pi) &= \sigma(-\mathbf{G}), \\ P(\mathrm{A}) &= Dir(a). \end{aligned} \tag{A.1}$$

The approximate posterior over unknown or hidden states and parameters $x = (\tilde{s}, \pi, \mathrm{A})$ can be expressed in terms of its sufficient statistics or expectations $\mathbf{x} = (\mathbf{s}_0^\pi, \ldots, \mathbf{s}_T^\pi, \boldsymbol{\pi}, \mathbf{a})$

$$Q(x) = Q(s_1|\pi)\ldots Q(s_T|\pi)Q(\pi)Q(\mathrm{A}),$$

$$\begin{aligned} Q(s_t|\pi) &= Cat(\mathbf{s}_t^\pi), \\ Q(\pi) &= Cat(\boldsymbol{\pi}), \\ Q(\mathrm{A}) &= Dir(\mathbf{a}). \end{aligned} \tag{A.2}$$

In this model, observations depend on only the current state, while state transitions depend on a policy or sequence of actions. This (sequential) policy is sampled from a Gibbs distribution or softmax function of expected free energy $\mathbf{G}(\pi)$ that depends on a prior cost $\mathbf{C}$ over future outcomes (see below). In more general formulations, the expected free energy would be scaled by a precision or inverse temperature parameter that we have previously associated with dopaminergic signals (Friston et al., 2014).

**A.2 Belief Updating.** Bayesian inference corresponds to minimizing variational free energy with respect to the expectations that constitute posterior beliefs. Free energy can be expressed as the (time-dependent) free energy under each policy plus the complexity incurred by posterior beliefs about (time-invariant) policies and parameters, where (ignoring constants)

$$\begin{aligned} F &= D[Q(x)||P(x)] - E_Q[\ln P(o_t|x)] \\ &= \sum_\tau E_Q[F(\pi, \tau)] + D[Q(\pi)||P(\pi)] + D[Q(\mathrm{A})||P(\mathrm{A})] + \ldots \\ &= \boldsymbol{\pi} \cdot (\widehat{\boldsymbol{\pi}} + \mathbf{F} + \mathbf{G}) + (\mathbf{a}_i - a_i) \cdot \widehat{\mathbf{A}}_i - \ln \mathrm{B}(\mathbf{a}_i) + \ldots \end{aligned}$$

The free energy of hidden states is given by:

$$\begin{aligned}
\mathbf{F}_\pi &= F(\pi) \\
F(\pi) &= \sum_\tau F(\pi, \tau) \\
F(\pi, \tau) &= E_Q[\ln Q(s_\tau|\pi) - \ln P(s_\tau, o_\tau|s_{\tau-1}, \pi)] \\
&= E_Q[\ln Q(s_\tau|\pi) - \ln P(s_\tau|s_{\tau-1}, o_\tau, \pi) - \ln P(o_\tau)] \\
&= E_Q[\ln Q(s_\tau|\pi) - \ln P(s_\tau|s_{\tau-1}, \pi) - \ln P(o_\tau|s_\tau)] \\
&= \underbrace{E_Q[D[Q(s_\tau|\pi)||P(s_\tau|s_{\tau-1}, o_\tau, \pi)]]}_{\textit{relative entropy}} - \underbrace{\ln P(o_\tau)}_{\textit{log evidence}} \\
&= \underbrace{E_Q[D[Q(s_\tau|\pi)||P(s_\tau|s_{\tau-1}, \pi)]]}_{\textit{complexity}} - \underbrace{E_Q[\ln P(o_\tau|s_\tau)]}_{\textit{accuracy}} \\
&= \mathbf{s}_\tau^\pi \cdot (\widehat{\mathbf{s}}_\tau^\pi - \widehat{\mathbf{B}}_{\tau-1}^\pi \mathbf{s}_{\tau-1}^\pi - \widehat{\mathbf{A}} \cdot o_\tau).
\end{aligned} \tag{A.3}$$

The expected free energy of a policy has the same form:

$$\begin{aligned}
\mathbf{G}_\pi &= G(\pi) \\
G(\pi) &= \sum_\tau G(\pi, \tau)
\end{aligned}$$

$$\begin{aligned}
G(\pi, \tau) &= E_{\tilde{Q}}[\ln Q(\mathrm{A}, s_\tau|\pi) - \ln P(\mathrm{A}, s_\tau, o_\tau|\tilde{o}, \pi)] \\
&= E_{\tilde{Q}}[\ln Q(\mathrm{A}) + \ln Q(s_\tau|\pi) - \ln P(\mathrm{A}|s_\tau, o_\tau, \tilde{o}, \pi) \\
&\quad - \ln P(s_\tau|o_\tau, \tilde{o}, \pi) - \ln P(o_\tau)] \\
&\approx E_{\tilde{Q}}[\ln Q(\mathrm{A}) + \ln Q(s_\tau|\pi) - \ln Q(\mathrm{A}|s_\tau, o_\tau, \pi) \\
&\quad - \ln Q(s_\tau|o_\tau, \pi) - \ln P(o_\tau)] \\
&= \underbrace{E_{\tilde{Q}}[\ln Q(\mathrm{A}) - \ln Q(\mathrm{A}|s_\tau, o_\tau, \pi)]}_{(\textit{negative parametric})\ \textit{mutual information}} \\
&\quad + \underbrace{E_{\tilde{Q}}[\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)]}_{(\textit{negative state})\ \textit{mutual information}} - \underbrace{E_{\tilde{Q}}[\ln P(o_\tau)]}_{\textit{expected value}}
\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{E_{\tilde{Q}}[\ln Q(\mathrm{A}) - \ln Q(\mathrm{A}|s_\tau, o_\tau, \pi)]}_{(negative)\ novelty} \\
&\quad + \underbrace{E_{\tilde{Q}}[\ln Q(o_\tau|\pi) - \ln P(o_\tau|s_\tau)]}_{(negative) intrinsic\ value} - \underbrace{E_{\tilde{Q}}[\ln P(o_\tau)]}_{extrinsic\ value} \\
&= \underbrace{E_{\tilde{Q}}[\ln Q(\mathrm{A}) - \ln Q(\mathrm{A}|s_\tau, o_\tau, \pi)]}_{ignorance} + \underbrace{D[Q(o_\tau|\pi)||P(o_\tau)]}_{risk} \\
&\quad + \underbrace{E_{\tilde{Q}}[H[P(o_\tau|s_\tau)]]}_{ambiguity} \\
&= \mathbf{o}_\tau^\pi \cdot \mathbf{W} \cdot \mathbf{s}_\tau^\pi + \mathbf{o}_\tau^\pi \cdot (\widehat{\mathbf{o}}_\tau^\pi + \mathbf{C}_\tau) + \mathbf{H} \cdot \mathbf{s}_\tau^\pi,
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{H} &= -diag(\mathbf{A} \cdot \widehat{\mathbf{A}}), \\
\mathbf{W} &= (\psi(\mathbf{a}) - \psi(\mathbf{a}_0)) - (\psi(\mathbf{a}+1) - \psi(\mathbf{a}_0+1)) = \mathbf{a}_0^{-1} - \mathbf{a}^{-1}, \quad \text{(A.4)}
\end{aligned}
$$

where $\tilde{Q} = Q(o_\tau, s_\tau|\pi) = P(o_\tau|s_\tau)Q(s_\tau|\pi)$ is the posterior predictive distribution over future outcomes and hidden states. Figure 3 provides the update rules based on minimizing variational free energy via a gradient descent:

$$
\begin{aligned}
\dot{\widehat{\mathbf{s}}}_\tau^\pi &= -\partial_{\mathbf{s}} F, \\
\partial_{\mathbf{s}} F &= \widehat{\mathbf{s}}_\tau^\pi - (\widehat{\mathbf{A}} \cdot o_\tau + \widehat{\mathbf{B}}_{\tau-1}^\pi \mathbf{s}_{\tau-1}^\pi + \widehat{\mathbf{B}}_\tau^\pi \cdot \mathbf{s}_{\tau+1}^\pi).
\end{aligned}
$$

The auxiliary variables $\widehat{\mathbf{s}}_\tau^\pi = \ln \mathbf{s}_\tau^{n,\pi}$ can be regarded as a postsynaptic depolarization in a neuronal setting, while the resulting firing rate is a sigmoid (softmax) function of depolarization $\mathbf{s}_\tau^\pi = \sigma(\widehat{\mathbf{s}}_\tau^\pi)$. The remaining update rules are derived in a straightforward way as the expectations (of policies and concentration parameters) at which their free energy gradients are zero—and free energy is therefore minimized.

**A.3 Bayesian Model Reduction.** The relative evidence for a full model and a reduced model with priors $a'$ can be derived from the application of Bayes' rule to both models (assuming $A$ is a column vector for simplicity):

$$
\frac{P(\mathrm{A}|\tilde{o}, m_R)}{P(\mathrm{A}|\tilde{o}, m_F)} = \frac{P(\mathrm{A}|m_R)}{P(\mathrm{A}|m_F)} \frac{P(\tilde{o}|m_F)}{P(\tilde{o}|m_R)} \Rightarrow
$$

$$
\frac{P(\tilde{o}|m_R)}{P(\tilde{o}|m_F)} = \int d\,\mathrm{A}\, P(\mathrm{A}|\tilde{o}, m_F) \frac{P(\mathrm{A}|m_R)}{P(\mathrm{A}|m_F)} \approx \int d\,\mathrm{A}\, Q(\mathrm{A}) \frac{P(\mathrm{A}|a')}{P(\mathrm{A}|a)}
$$

$$= \frac{\mathrm{B}(a)\mathrm{B}(\mathbf{a}+a'-a)}{\mathrm{B}(\mathbf{a})\mathrm{B}(a')}$$

$$P(\mathrm{A}|a) = Dir(a) = \mathrm{B}(a)\prod_i \mathrm{A}_i^{a-1}.$$

Here, $\mathrm{B}(\cdot)$ denotes the beta function. The evidence ratio in the second equality can now be expressed as a change in free energy:

$$\begin{aligned}\Delta F &= \ln P(\tilde{o}|m_F) - \ln P(\tilde{o}|m_R)\\ &= \ln \mathrm{B}(\mathbf{a}) + \ln \mathrm{B}(a') - \ln \mathrm{B}(a) - \ln \mathrm{B}(\mathbf{a}+a'-a).\end{aligned}$$

This provides a criterion to accept or reject an alternative hypothesis or reduced model structure that is encoded by the hyperparameters $a'$. Finally, the reduced posteriors follow from the above equalities, should we accept the reduced model:

$$Q(\mathrm{A}|m_R) = \mathrm{B}(\mathbf{a}+a'-a)^{-1}\prod_i \mathrm{A}_i^{\mathbf{a}+a'-a-1} = Dir(\mathbf{a}+a'-a).$$

Table 2: Glossary of Variables and Expressions.

| Expression | Description |
|---|---|
| $o_\tau = (o_\tau^1, \ldots, o_\tau^M) : o_\tau^m \in \{1, \ldots, D_m\}$<br>$\mathbf{o}_\tau = (\mathbf{o}_\tau^1, \ldots, \mathbf{o}_\tau^M) \in [0,1]$<br>$\widehat{\mathbf{o}}_\tau^m = \ln \mathbf{o}_\tau^m \in \mathbb{R}^{D_m}$ | Outcomes in $M$ modalities (the $m$th modality has $D_m$ outcomes), their (future) posterior expectations and logarithms |
| $\tilde{o} = (o_1, \ldots, o_t)$ | Sequences of outcomes until the current time point. |
| $s_\tau = (s_\tau^1, \ldots, s_\tau^N) : s_\tau^n \in \{1, \ldots, D_n\}$<br>$\mathbf{s}_\tau = (\mathbf{s}_\tau^1, \ldots, \mathbf{s}_\tau^N) \in [0,1]$<br>$\widehat{\mathbf{s}}_\tau^n = \ln \mathbf{s}_\tau^n \in \mathbb{R}^{D_n}$ | Hidden states over $N$ factors (the $n$th modality has $D_n$ states), their posterior expectations and logarithms |
| $\tilde{s} = (s_1, \ldots, s_T)$ | Sequences of hidden states until the end of the current trial |
| $\pi \in \{1, \ldots, K\}$<br>$\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_K) : \boldsymbol{\pi} \in [0,1]$<br>$\widehat{\pi} = \ln \boldsymbol{\pi} \in \mathbb{R}^{T\times K\times N}$ | $K$ policies specifying action sequences, their posterior expectations and logarithms |
| $u^n \in \{1, \ldots, L\}$<br>$U^n \in \mathbb{R}^{K\times T}$<br>$U_\tau^{n,\pi} \in \{1, \ldots, L\}$ | Action or control variables for each factor of hidden states and sequences of actions under the $\pi$th policy comprising allowable actions |

Table 2: Continued.

| Expression | Description |
|---|---|
| $\mathbf{A}^m = E_Q[A^m] = \frac{\mathbf{a}^m}{\mathbf{a}_0^m} \in \mathbb{R}^{D_m \times D_1 \times \ldots \times D_N}$<br>$\widehat{\mathbf{A}}^m = E_Q[\ln A^m] = \psi(\mathbf{a}^m) - \psi(\mathbf{a}_0^m)$<br>$\mathbf{a}_{0ij}^m = \sum_i \mathbf{a}_{ij}^m$ | Expected outcome probabilities (likelihood) for the $m$th modality under each combination of hidden states and their expected logarithms, which depend on the sum of concentration parameters for outcome |
| $a^m \in \mathbb{R}^{D_m \times D_1 \times \ldots \times D_N}$<br>$\mathbf{a}^m \in \mathbb{R}^{D_m \times D_1 \times \ldots \times D_N}$ | Prior and posterior concentration parameters of the likelihood |
| $\mathbf{B}_\tau^{n,\pi} \in \{\mathbf{B}_1, \ldots \mathbf{B}_U\} \in \mathbb{R}^{D_n \times D_n}$<br>$\mathbf{B}_\tau^{n,\pi} \mathbf{B}_{U_\tau^{n,\pi}}$<br>$\widehat{\mathbf{B}}_\tau^{n,\pi} = \ln \mathbf{B}_\tau^{n,\pi}$ | Transition probabilities for the $n$th hidden state under each action prescribed by a policy at a particular time and its logarithm |
| $\mathbf{C}_\tau^m = -\ln P(o_\tau^m)$ | Surprise associated with the $m$th outcome—i.e., prior cost or negative preference |
| $\mathbf{D}^n = P(s_1^n \mid s_0^n) \in \mathbb{R}^{D_n}$ | Prior expectation of the $n$th hidden factor at the beginning of each trial |
| $\mathbf{F} : \mathbf{F}_\pi = F(\pi) = \sum_\tau F(\pi, \tau) \in \mathbb{R}^K$ | Variational free energy for each policy |
| $\mathbf{G} : \mathbf{G}_\pi = G(\pi) = \sum_\tau G(\pi, \tau) \in \mathbb{R}^K$ | Expected free energy for each policy |
| $\mathbf{H}^m \in \mathbb{R}^{D_1 \times \ldots \times D_N}$<br>$\mathbf{H}_{jk\ldots}^m = -\sum_i \mathbf{A}_{ijk\ldots}^m \cdot \widehat{\mathbf{A}}_{ijk\ldots}^m$ | An array encoding the entropy or ambiguity over outcomes for each combination of hidden states |
| $\mathbf{W}^m = 1/\mathbf{a}_0^m - 1/\mathbf{a}^m$ | An array encoding uncertainty about the likelihood for each combination of outcomes and hidden states |
| $\mathbf{A} \circ \mathbf{s} = \mathbf{A} \circ (\mathbf{s}^1, \ldots, \mathbf{s}^N)$<br>$(\mathbf{A} \circ \mathbf{s})_i = \sum_{j,k\ldots} \mathbf{A}_{i,j,k\ldots} \mathbf{s}_j^1 \mathbf{s}_k^2 \ldots$ | Generalized dot product (or sum of products), returning a vector |
| $\mathbf{A} \circ \mathbf{s}^{/n}$ | Generalized dot product over all but the $n$th vector, returning a matrix |
| $\mathbf{s}_\tau^{n,\pi} = \sigma(\widehat{\mathbf{s}}_\tau^{n,\pi})$ | Expected hidden states $n$ under a particular policy |
| $\mathbf{o}_\tau^{m,\pi} = \mathbf{A}^m \circ \mathbf{s}_\tau^\pi$ | Expected outcomes $m$ under a particular policy |
| $\mathbf{s}_\tau^n = \sum_\pi \boldsymbol{\pi}_\pi \cdot \mathbf{s}_\tau^{n,\pi}$ | Bayesian model average of hidden states over policies |
| $Cat(A)$<br>$Dir(a)$ | Categorical and Dirichlet distributions, defined in terms of their sufficient statistics (probabilities and concentration parameters) |

Table 2: Continued.

| Expression | Description |
|---|---|
| $\sigma(-\mathbf{G})_\pi = \frac{\exp(-\mathbf{G}_\pi)}{\sum_\pi \exp(-\mathbf{G}_\pi)}$ | Softmax function, returning a vector that can be treated as a proper probability distribution |
| $\psi(\mathbf{a}) = \frac{d}{d\mathbf{a}} \ln \Gamma(\mathbf{a})$ | Digamma function, defined as the derivative of the log gamma function |
| $B(\mathbf{a}) = \frac{\Gamma(\mathbf{a}_1)\Gamma(\mathbf{a}_2)\cdots}{\Gamma(\mathbf{a}_1+\mathbf{a}_2\cdots)}$ | Beta function, used in the definition of the Dirichlet distribution |

## Acknowledgments

K.J.F. is funded by the Wellcome Trust (Ref: 088130/Z/09/Z). We thank our reviewers for detailed and helpful guidance in reporting this work.

## Disclosure Statement

We have no disclosures or conflict of interest.

## References

Aragones, E., Gilboa, I., Postlewaite, A., & Schmeidler, D. (2005). Fact-free learning. *American Economic Review*, *95*, 1355–1368.

Attias, H. (2003). Planning by probabilistic inference. In *Proc. of the 9th Int. Workshop on Artificial Intelligence and Statistics*.

Auble, P. M., Franks, J. J., & Soraci, S. A. (1979). Effort toward comprehension: Elaboration or "aha"? *Memory and Cognition*, *7*, 426–434.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*, 329–349.

Bar, M., Aminoff, E., Mason, M., & Fenske, M. (2007). The units of thought. *Hippocampus*, *17*, 420–428.

Baranes, A., & Oudeyer, P. Y. (2009). R-IAC: Robust intrinsically motivated exploration and active learning. *IEEE Transactions on Autonomous Mental Development*, *1*, 155–169.

Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. In W. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

Beal, M. J. (2003). *Variational algorithms for approximate bayesian inference*. Ph.D. thesis, University College London.

Bellman, R. (1952). On the theory of dynamic programming. *Proc. Natl. Acad. Sci. USA*, *38*, 716–719.

Berlyne, D. E. (1950). Novelty and curiosity as determinants of exploratory behaviour. *British Journal of Psychology—General Section*, *41*, 68–80.

Berlyne, D. E. (1954). A Theory of human curiosity. *British Journal of Psychology*, *45*, 180–191.

Bonet, B., & Geffner, H. (2014). Belief tracking for planning with sensing: Width, complexity and approximations. *Journal of Artificial Intelligence Research*, *50*, 923–970.

Botvinick, M., & An, J. (2008). Goal-directed decision making in prefrontal cortex: A computational framework. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems*, *21*. Cambridge, MA: MIT Press.

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends Cogn. Sci.*, *16*, 485–488.

Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends Cogn. Sci. 9*, 322–328.

Braun, D. A., Ortega, P. A., Theodorou, E., & Schaal, S. (2011). Path integral control and bounded rationality. In *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (pp. 202–209). Piscataway, NJ: IEEE.

Campbell, J. O. (2016). Universal Darwinism as a process of Bayesian inference. *Frontiers in Systems Neuroscience*, *10*.

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*, 200–219.

Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering and generalizing task-set structure. *Psychological Review*, *120*, 190–229.

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLoS Biology 10*, e1001293.

Coltheart, M., Menzies, P., & Sutton, J. (2010). Abductive inference and delusional belief. *Cogn. Neuropsychiatry*, *15*, 261–287.

Donoso, M., Collins, A. G., & Koechlin, E. (2014). Human cognition: Foundations of human reasoning in the prefrontal cortex. *Science*, *344*, 1481–1486.

Dragoi, G., & Tonegawa, S. (2014). Selection of preconfigured cell assemblies for representation of novel spatial experiences. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *369*, 20120522.

Dresler, M., Wehrle, R., Spoormaker, V. I., Steiger, A., Holsboer, F., Czisch, M., & Hobson, J. A. (2015). Neural correlates of insight in dreaming and psychosis. *Sleep Medicine Reviews*, *20*, 92–99.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132.

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

Frank, S. A. (2012). Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *Journal of Evolutionary Biology*, *25*, 2377–2396.

Friston, K. (2013). Life as we know it. *J. R. Soc. Interface*, *10*, 20130475.

Friston, K., Adams, R., & Montague, R. (2012). What is value—accumulated reward or evidence? *Frontiers in Neurorobotics*, *6*, 11.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience and Biobehavioral Reviews*, *68*, 862–879.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Comput.*, *29*, 1–49.

Friston, K., & Frith, C. (2015). A duet for one. *Conscious Cogn. 36*, 390–405.

Friston, K. J., Litvak, V., Oswal, A., Razi, A., Stephan, K. E., van Wijk, B. C., . . . Zeidman, P. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *NeuroImage 128*, 413–431.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern.*, *104*, 137–160.

Friston, K., & Penny, W. (2011). Post hoc Bayesian model selection. *NeuroImage*, *56*, 2089–2099.

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

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: Active inference and agency. *Front. Hum. Neurosci.*, *7*, 598.

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

Frith, C. (2010). What is consciousness for? *Pragmatics and Cognition*, *18*, 497–551.

Frith, C., & Wentzer, T. (2013). Neural hermeneutics. In B. Kaldis (Ed.), *Encyclopedia of philosophy and the social sciences*. Thousand Oaks, CA: Sage.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Curr. Opin. Neurobiol.*, *20*, 251–256.

Harman, G. H. (1965). The inference to the best explanation. *Philosophical Review*, *74*, 88–95.

Hinton, G. E., & van Camp, D. (1993). Keeping neural networks simple by minimizing the description length of weights. In *Proceedings of COLT-93* (pp. 5–13).

Hinton, G. E., & Zemel, R. S. (1993). Autoencoders, minimum description length and Helmholtz free energy. In *Proceedings of the 6th International Conference on Neural Information Processing Systems* (pp. 3–10). San Mateo, CA: Morgan Kaufmann.

Hobson, J. A. (2009). REM sleep and dreaming: Towards a theory of protoconsciousness. Nat. *Rev. Neurosci.*, *10*, 803–813.

Hobson, J. A., & Friston, K. J. (2012). Waking and dreaming consciousness: Neurobiological and functional considerations. *Prog. Neurobiol.*, *98*, 82–98.

Hobson, J. A., & Friston, K. J. (2014). Consciousness, dreams, and inference: the Cartesian theatre revisited. *Journal of Consciousness Studies*, *21*, 6–32.

Hobson, J. A., & Friston, K. J. (2016). A response to our theatre critics. *Journal of Consciousness Studies*, *23*, 245–254.

Hohwy, J. (2016). The self-evidencing brain. *Noûs 50*, 259–285.

Holzel, B. K., Carmody, J., Vangel, M., Congleton, C., Yerramsetti, S. M., Gard, T., & Lazar, S. W. (2011). Mindfulness practice leads to increases in regional brain gray matter density. *Psychiatry Research—Neuroimaging*, *191*, 36–43.

Howard, R. (1966). Information value theory. *IEEE Transactions on Systems, Science and Cybernetics*, *SSC-2*, 22–26.

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

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, *78*, 2690–2693.

Jones, G. (2003). Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1017–1027.

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., . . . Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biology*, *2*, E97.

Klyubin, A. S., Polani, D., & Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In Proc. *IEEE Congress on Evolutionary Computation* (pp. 128–135). Piscataway, NJ: IEEE.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Mem. Cognit.*, *29*, 1000–1009.

Koechlin, E. (2015). Prefrontal executive function and adaptive behavior in complex environments. *Curr. Opin. Neurobiol.*, *37*, 1–6.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science 302*, 1181–1185.

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., & Jung-Beeman, M. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia*, *46*, 281–291.

Laughlin, S. B. (2001). Efficiency and complexity in neural coding. *Novartis. Found. Symp., 239*, 177–187.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

Linsker, R. (1990). Perceptual neural organization: Some approaches based on network models and information theory. *Annu. Rev. Neurosci.*, *13*, 257–281.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*, 75–98.

Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron 29*, 145–156.

Lu, H., Rojas, R. R., Beckers, T., & Yuille, A. L. (2016). A Bayesian theory of sequential causal learning and abstract transfer. *Cogn. Sci.*, *40*, 404–439.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 176–201.

Mai, X. Q., Luo, J., Wu, J. H., & Luo, Y. J. (2004). "Aha!" effects in a guessing riddle task: An event-related potential study. *Hum. Brain Mapp.*, *22*, 261–270.

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*, 94–122.

Markant, D. B., Settles, B., & Gureckis, T. M. (2016). Self-directed learning favors local, rather than global, uncertainty. *Cogn. Sci.*, *40*, 100–120.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences*, *11*, 442–450.

Meder, B., & Nelson, J. D. (2012). Information search with situation-specific reward functions. *Judgment and Decision Making*, *7*, 119–148.

Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. Cambridge, MA: MIT Press.

Mirza, M. B., Adams, R. A., Mathys, C. D., & Friston, K. J. (2016). Scene construction, visual foraging, and active inference. *Frontiers in Computational Neuroscience*, *10*, 56.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., . . . Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*, 529–533.

Mohammadi, D. (2015). ENIGMA: Crowdsourcing meets neuroscience. *Lancet Neurology*, *14*, 462–463.

Moutoussis, M., Trujillo-Barreto, N. J., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). A formal model of interpersonal inference. *Front Hum. Neurosci.*, *8*, 160.

Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychol. Rev.*, *118*, 120–134.

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

Nelson, J. D., McKenzie, C. R., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychol. Sci.*, *21*, 960–969.

Nimbalkar, N. (2011). John Locke on personal identity. *Mens Sana Monographs*, *9*, 268–275.

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin and Review*, *5*, 345–369.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*, 349–357.

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychon. Bull. Rev.*, *10*, 289–318.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*, 883–899.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Ortega, P. A., & Braun, D. A. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Proc. R. Soc. A.*, 4692153.

Palmer, C. J., Seth, A. K., & Hohwy, J. (2015). The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism. *Conscious Cogn. 36*, 376–389.

Penny, W. D. (2012). Comparing dynamic causal models using AIC, BIC and free energy. *NeuroImage*, *59*, 319–330.

Pezzulo, G., van der Meer, M. A., Lansink, C. S., & Pennartz, C. M. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.*, *18*, 647–657.

Rasch, B., Buechel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, *315*, 1426–1429.

Saegusa, R., Metta, G., Sandini, G., & Sakka, S. (2009). Active motor babbling for sensorimotor learning. In *2008 IEEE International Conference on Robotics and Biomimetics* (vols. 1–4). Piscataway, NJ: IEEE.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schmidhuber, J. (1991). Curious model-building control systems. In *Proc. International Joint Conference on Neural Networks* (vol. 2, pp. 1458–1463). Piscataway, NJ: IEEE.

Schmidhuber, J. (2006). Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science*, *18*, 173–187.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, *2*, 230–247.

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., & Friston, K. (2015). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cereb. Cortex*, *25*, 3434–3445.

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., & Friston, K. (2015). Optimal inference with suboptimal models: Addiction and active Bayesian inference. *Med. Hypotheses*, *84*, 109–117.

Sengupta, B., Stemmler, M. B., & Friston, K. J. (2013). Information and efficiency in the nervous system—a synthesis. *PLoS Comput. Biol.*, *9*, e1003157.

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cogn. Neurosci.*, *5*, 97–118.

Seth, A. K. (2015). Inference to the best prediction. In T. K. Metzinger & J. M. Windt (Eds.), Open *MIND* Frankfurt: MIND Group.

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends Cogn. Sci.*, *18*, 186–193.

Shen, W., Yuan, Y., Liu, C., & Luo, J. (2016). In search of the "Aha!" experience: Elucidating the emotionality of insight problem-solving. *British Journal of Psychology*, *107*, 281–298.

Solway, A., & Botvinick, M. (2012). Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychological Review*, *119*, 120–154.

Still, S., & Precup, D. (2012). An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Biosciences* (Theorie in den Biowissenschaften), *131*, 139–148.

Stoianov, I., Genovesio, A., & Pezzulo, G. (2016). Prefrontal goal codes emerge as latent states in probabilistic value learning. *J. Cogn. Neurosci.*, *28*, 140–157.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285.

Tervo, D. G., Tenenbaum, J. B., & Gershman, S. J. (2016). Toward the neural implementation of structure learning. *Curr. Opin. Neurobiol.*, *37*, 99–105.

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Med. Rev.*, *10*, 49–62.

van den Broek, J. L., Wiegerinck, W. A. J. J., & Kappen, H. J. (2010). Risk-sensitive path integral control. *Uncertainty in Artificial Intelligence*, *6*, 1–8.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*, 352–355.

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, *44*, 121–133.

Wallace, C. S., & Dowe, D. L. (1999). Minimum message length and Kolmogorov complexity. *Computer Journal*, *42*, 270–283.

Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Inform. Theory*, *51*, 2282–2312.

Zhang, H., & Maloney, L. T. (2012). Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience*, *6*, 1.

---

Received January 15, 2017; accepted April 27, 2017.