J. Allan Hobson[1]
and Karl J. Friston[2]

Consciousness, Dreams, and Inference

The Cartesian Theatre Revisited

Abstract: *This paper considers the Cartesian theatre as a metaphor for the virtual reality models that the brain uses to make inferences about the world. This treatment derives from our attempts to understand dreaming and waking consciousness in terms of free energy minimization. The idea here is that the Cartesian theatre is not observed by an internal (homuncular) audience but furnishes a theatre in which fictive narratives and fantasies can be rehearsed and tested against sensory evidence. We suppose the brain is driven by the imperative to infer the causes of its sensory samples; in much the same way as scientists are compelled to test hypotheses about experimental data. This recapitulates Helmholtz's notion of unconscious inference and Gregory's treatment of perception as hypothesis testing. However, we take this further and consider the active sampling of the world as the gathering of confirmatory evidence for hypotheses based on our virtual reality. The ensuing picture of consciousness (or active inference) resolves a number of seemingly hard problems in consciousness research and is internally consistent with current thinking in systems neuroscience and theoretical neurobiology. In this formalism, there is a dualism that distinguishes between the (conscious) process of*

Correspondence:
Karl Friston, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, Queen Square, London, WC1N 3BG. *Email: k.friston@.ucl.ac.uk*

[1] Division of Sleep Medicine, Harvard Medical School, Boston, MA 02215, USA.

[2] The Wellcome Trust Centre for Neuroimaging, University College London, Queen Square, London, WC1N 3BG.

Journal of Consciousness Studies, **21**, No. 1–2, 2014, pp. 6–32

inference and the (material) process that entails inference. This separation is reflected by the distinction between beliefs (probability distributions over hidden world states or res cogitans) and the physical brain states (sufficient statistics or res extensa) that encode them. This formal approach allows us to appeal to simple but fundamental theorems in information theory and statistical thermodynamics that dissolve some of the mysterious aspects of consciousness.

Keywords: consciousness; prediction; free energy; neuronal coding; sleep; inference; neuromodulation.

Introduction

This paper calls on current theories about brain function — in wakefulness and sleep — to address questions about phenomenal consciousness. In brief, our understanding of the brain–mind as a theatre is not Cartesian — in that we renounce dualism (Dennett, 1991). We put in its place a dual aspect monism and explain how the two aspects depend on each other, differ from each other, and how they interact causally. Formally, we consider consciousness to be the process of perceptual inference about states of the world causing sensations (Helmholtz, 1866/1962; Gregory, 1980). Here, inference is taken to be the formation of *probabilistic beliefs* through optimizing the *sufficient statistics* of probability distributions. In other words, we consider consciousness as finding the best (in a Bayes optimal sense) probabilistic explanation for our sensorium.

The account on offer provides a monistic solution that bridges the Cartesian divide between the *res cogitans* and *res extensa* — the realms of thought and matter (Manuel, 2001, p. 97), where immaterial beliefs are probability distributions that are entailed by material sufficient statistics. This may sound like a rather obvious (and possibly facile) account of consciousness; however, it appears to have a degree of face validity and offers simple answers to seemingly hard questions. For example, both consciousness and inference are about something — in the sense that both have content. There is a unitary aspect — in the sense that inference produces a unique belief. Furthermore, both are inherently private and embodied — in the sense that your inference is about (sensory) evidence available to you, and only you. While we hope to transcend the homunculus concept inherent in Cartesian dualism, we recognize that postulating an innate self — arising early in development — raises the spectre of an internal observer, who watches, unifies, and responds to a show of cognitive machinations.

We will try to resolve this dialectic using the notion of hierarchical inference.

We start by reviewing some aspects of consciousness from the perspective of systems neuroscience. This review serves to contextualize some of the mind–brain issues addressed later. We then work through a series of questions about phenomenal consciousness, developing the arguments as we go — starting with the hard problem (Chalmers, 1995) and ending with questions about free will (Clark, 1999). Although we will not consider the mechanistic details of how a material object — the brain — can produce an immaterial process — consciousness — we present an argument for their intimate relationship and the potential research avenues that ensue. A discussion of the mechanisms that underlie the coupling between neuronal and inferential processes can be found in Friston (2012) — from a mathematical perspective — and Hobson and Friston (2012) — from a neurobiological perspective. This paper concludes with an epilogue (based on our correspondence) that highlights outstanding issues and potential ways forward.

Some Preliminaries

The neural correlates of consciousness (Mormann and Koch, 2007) can be defined and measured with a view to understanding how qualia are associated with underlying brain activity. When studied in the states of waking, sleeping, and dreaming, the neural correlates of consciousness demonstrate an encouraging consistency (Hobson and Wohl, 2005; Hobson, 2013), suggesting that the conscious experience is one aspect of a process that is embodied by the brain. Further consideration of this dual aspect model suggests that consciousness, as experienced in waking, has a fundamental relationship to the altered state of consciousness that we experience in dreaming. Ontogenetic and phylogenetic data further suggest that dream consciousness — and its neural substrates — precede and make possible the later and more elaborate form of consciousness in waking (Hobson, 2009a).

Dream consciousness and its physiological underpinnings have been considered as a virtual reality model of the world that prepares us for waking consciousness (*ibid.*). This virtual reality model is formally equivalent to the generative models implicit in unconscious inference as described by Helmholtz (1866/1962) and recent formulations of the Bayesian brain (Dayan, Hinton and Neal, 1995; Knill and Pouget, 2004). The idea here is that perceptual synthesis results — not from (bottom-up) sensory impressions forcing themselves on the brain —

but from an active process of (top-down) prediction and confirmation — where predictions (fantasies) are generated in a virtual model of the world and then tested against sensory reality. We will use these concepts to offer some answers to questions about the nature of consciousness; in particular, we consider consciousness in terms of inference based on the private theatres of virtual reality that are so manifest in dreaming.

Cognitive neuroscience recognizes many modular aspects of consciousness, where the usual approach is to investigate one module at a time. In contrast, more integrated approaches — like the conscious state paradigm (Stickgold and Hobson, 1995) — emphasize the global integration of modular neural processes and considers their integration and differentiation by specific brain mechanisms: *cf.* Baars (1997), Dehaene and Changeux (2011), Tononi (2000). The constancy and variation of these global aspects speak to the similarities and differences between phenomenal states at both the macroscopic level and the microscopic level of cellular and molecular neurophysiology. One physiologically informed framework (Hobson, 2009a; 2013) addresses the fundamental dimensions of phenomenal states and their neurobiological underpinnings that — unlike many formulations — accommodates fluctuations in the level and nature of consciousness.

The AIM model (Hobson, 2009a) uses the dimensions of *activation* (*A*), *input–output gating* (*I*), and *modulation* (*M*) to link phenomenal and neural or extensive levels of description: the term activation is used to express the level of energy consumption of the brain and its constituent circuits. Input–output gating facilitates or attenuates access to sensory information (input) from the outside world and the emission of motor commands from the brain (output) to the musculature. The modulatory microclimate of the brain is determined largely by neurons in the brainstem, which send axons to the forebrain, spinal cord, and cerebellum. Among the neurotransmitters released by these ascending modulatory systems are dopamine, noradrenalin, serotonin, histamine, and acetylcholine. Both waking and dreaming are characterized by high levels of activation, where input–output gating and modulation reliably differentiate these two states. Crucially, during rapid eye movement (REM) sleep, the brain is activated but sequestered from its sensory inputs by modulatory gating mechanisms. This is a rather remarkable state of affairs in its own right but there is something even more curious about this state of consciousness.

One of the most surprising and biologically significant findings — in sleep and dream research — is the relationship between thermoregulation and sleep. Only mammals and birds show thermoregulation and only mammals and birds show brain activation in sleep (Rechtschaffen *et al.*, 1989). Furthermore, only mammals and birds evidence high-level consciousness. Neurons that secrete norepinephrine, serotonin, and histamine are quiescent in REM sleep (Hilakivi, 1987) — without these neuromodulators, animals are deprived of precise sensory input and, in particular, cannot maintain homeothermy. This means that REM sleep is the only state of mammalian existence in which homeothermy is suspended (Parmeggiani, 2007). So what evolutionary imperatives mandate this risky physiological state?

We have previously considered the answer to this question in terms of how the brain optimizes its model of the world (Hobson and Friston, 2012). The answer that emerges is that sleep is a necessary process that requires the (nightly) suspension of sensory input — so that synaptic plasticity and homoeostasis can reduce the redundancy and complexity accrued during wakefulness (Gilestro, Tononi and Cirelli, 2009). In short, it is necessary to gate sensory input (and responses) to finesse the complexity of virtual reality models used to navigate the waking sensorium. This is an important theme that we will return to later.

From the perspective of consciousness research, these observations have something quite profound to say. First, percepts are not driven by sensory input — they can arise during dreaming in the absence of any sensations. In short, percepts are literally fantastic (from Greek *phantastikos*, able to create mental images, from *phantazesthai*). Second, the evolutionary pressure to maintain dream consciousness during REM sleep — despite predation and thermoregulatory costs — speaks to the importance of actively maintaining a generative model of the world. The importance of this maintenance is evident at a number of levels:

Phylogeny: differentiation of brain activation and sensory gating increase with evolution (Allison and Cicchetti, 1976). Thus, both neuroanatomical complexity and the differentiation of conscious states become increasingly prominent as the brain adds layer upon layer to its structure. The phylogenetic addition of layers to the brain's anatomy is important here, because it adds hierarchal depth to its putative models. Man is the apogee of this trend but other mammals and birds share many hierarchical features of brain organization. Models with a hierarchical form provide multiple levels of explanation for the causes

of (sensory) data. Implicit in much of our later discussion will be the distinction between explanations at low hierarchical levels — which we associate with *phenomenal* consciousness or qualia — and explanations at higher levels — which we associate with *access* consciousness (Block, 1998). This distinction is based purely upon the level of hierarchical inference and is not unrelated to the distinction between *primary* and *secondary* consciousness (Edelman, 2001). See also Clark (2000). Associating conscious processes with inference necessarily imbues consciousness with a hierarchical aspect. In other words, low-level inference of the sort associated with motor reflexes — or the unconscious inference implied by Helmholtz — does not in itself constitute conscious processing until contextualized by deep hierarchical inference at higher levels (see Figure 1).

Ontogeny: at all levels of mammalian development, the temporal precedence and predominance of activated sleep is striking (Roffwarg, Muzio and Dement, 1966). The marked preponderance of REM sleep in the last trimester of pregnancy and the first year of life decreases progressively as waking time increases. Despite its early decline, REM sleep continues to occupy an hour or so per day. This suggests a strong developmental contribution and that activated sleep subsequently plays an indispensable part in maintaining the adaptive and inferential capacity of the brain throughout life.

Phenomenology: both states (sleeping and dreaming) of brain activation fluctuate over the course of a day but neither is ever in complete abeyance or complete dominance (Aserinsky and Kleitman, 1953; Kripke *et al*., 2002). This is further evidence of their cooperative interaction. Waking and dreaming are not mutually exclusive: they both serve a common purpose — to optimize generative models of its world. The functional significance of this fact is that their interaction is continuous and that both states may be essential for normal consciousness — a consciousness predicated on a virtual reality, with one foot in the sensorium (and often no feet).

In summary, empirical and theoretical approaches to the brain as a conscious artefact — particularly fluctuations in consciousness during sleep and wakefulness — suggest the brain maintains a model or virtual reality that it uses to explain sensory inputs. The process of engaging that model during perception and action necessarily calls upon learning and inference — processes that can proceed in the absence of sensory exchange with the world. With this in mind, we now turn to some key questions about the nature of phenomenal consciousness.

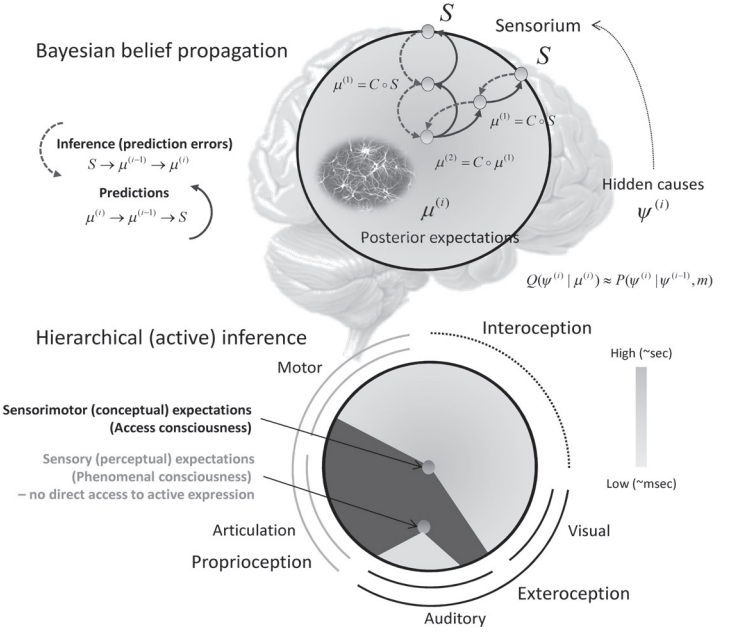

Figure 1. **Upper Panel**: this schematic illustrates hierarchical Bayesian inference in the brain using a centrifugal hierarchy (Mesulam, 1998) of cortico-cortical connections (Felleman and Van Essen, 1991) and Bayesian belief updating with predictive coding — a simple form of belief propagation. In these schemes, the sufficient statistics correspond to posterior expectations and are updated using prediction errors from the hierarchical level below. The prediction errors are formed by descending predictions based upon expectations. This recurrent signalling minimizes prediction error (and variational free energy), enabling the brain to perform approximate Bayesian inference in a neurally plausible fashion. **Lower Panel**: this schematic illustrates the qualitative difference between posterior expectations at high levels of the hierarchy — that predict proprioceptive input — and expectations at lower levels — that do not. This distinction may be important in terms of access and phenomenal consciousness: changes in high-level expectations change proprioceptive predictions and are in a position to elicit (self) report. Conversely, hierarchically lower expectations do not have access to (self) report, because they cannot change proprioceptive predictions engendering (inner) speech. This may provide a simple perspective on the difference between the products of access and phenomenal consciousness; namely, actionable concepts and qualia.

Some Hard Questions

The hard problem (Chalmers, 1995) is an interesting concept — and begs the question: is the problem hard to answer or hard to specify? We suppose, in line with deflationary arguments, the difficulties lie partly in formulating well posed questions — as opposed to supplying answers (Dennett, 1991). We will try to illustrate this by considering a series of increasingly hard questions and offering straightforward answers based on the idea that the brain is an inference machine (Dayan, Hinton and Neal, 1995; Hobson and Friston, 2012).

Can material (extensive) systems have (immaterial) attributes such as consciousness or qualia?

The answer to this is yes. Heuristically, any material system can have immaterial aspects — for example, the symmetry of an arrangement of marbles. More formally, the emergence of immaterial attributes from the collective behaviour of material ensembles underwrites most of the physical sciences. Key examples here include statistical thermodynamics, in which macroscopic quantities like temperature and pressure are induced by — or emerge from — the interaction of coupled physical systems (like atoms) that constitute an ensemble. Generally, the material attributes of a system can be cast in terms of fast *microscopic* variables, whose ensemble behaviour gives rise to slow *macroscopic* quantities. These macroscopic quantities are sometimes called *order parameters* or unstable (slow or dissipative) modes. This distinction — between the microscopic physical properties and the macroscopic attributes they entail — lies at the heart of synergetic formulations of complex systems and can be found in basic theorems in the physical sciences: such as the *centre manifold theorem* (Carr, 1981; Davis, 2006) and the *slaving principle* in physics (Haken, 1983). The dissociation into microscopic (material) and macroscopic (immaterial) dynamics is an inevitable consequence of the *separation of temporal scales* seen in all coupled dynamical systems (*ibid.*; Ginzburg and Landau, 1950). The nice thing about these formulations is that they appeal to a *circular causality* in which the microscopic properties cause the macroscopic behaviour while, at the same time, they are enslaved by macroscopic properties. While these observations do not specify how the activity of neural ensembles entails consciousness, they provide an encouraging framework for supposing that this is the case.

Perhaps one of the most important insights from understanding the nature of macroscopic properties is that many things we think of as

physical, such as temperature, are not. Temperature is, in fact, a sufficient statistic of a probability distribution (a Gibbs distribution) over the occupancy of microscopic states. This means that temperature is — in every respect — as immaterial and metaphysical as qualia. In other words, when we measure temperature, we are measuring an aspect of a probability distribution over physical (microscopic) states — we are not measuring the states *per se* (Landau and Lifshitz, 1976). The fact that these aspects are measurable speaks to the lawful dynamics of large ensembles of microscopic systems that show a collective — and sometimes self-organized — probabilistic behaviour.

A sufficient statistic is just a parameter of a probability distribution, such as the mean or variance (temperature is the variance of the Gibbs distribution). In Bayesian statistics (see glossary), probability distributions are known as *beliefs* and we will use the word belief in this sense. Notice that probability distributions are over alternative outcomes and therefore a belief entails some uncertainty that is generally resolved by observations. This resolution is known as Bayesian belief updating (Cox, 2001) based upon a generative model or Bayesian belief network (Pearl, 1988). Belief updating involves combining prior beliefs with (sensory) evidence to form posterior beliefs. Clearly, prior beliefs will be unique to any individual, which means different belief updates can be induced by the same observations — and are sensitive to our history or experience.

This Bayesian formulation of beliefs will become important later, when we consider the relationship between consciousness and beliefs. In the current context, the interesting thing here is that so-called physical (thermodynamic) properties are the sufficient statistics of probability distributions and it is the probability distributions that show lawful behaviour. In short, even at the level of classical thermodynamics, we quickly leave the microscopic material world and enter a world of probability distributions and how they interact (Frank, 2004).

In summary, there is an inevitable emergence of (macroscopic) probabilistic attributes that cannot be reduced to the (microscopic) material properties of dynamical systems. If one is prepared to consider consciousness as one such property, then there is an easy answer to the above hard question — yes.

Can we be conscious of being conscious?

This is a more interesting question (Schooler *et al.*, 2011). Can the symmetrical arrangement of marbles be itself symmetrical? Can one have a red red? More specifically, can we be conscious of qualia. If the

answer to this question is no then there is no hard problem because the question is ill posed (by any conscious entity). If the answer is yes, then one has to consider the constraints that being conscious of consciousness (having beliefs about beliefs) places on the nature of consciousness. One obvious constraint is that the operation of consciousness on itself must produce something that has the attribute of consciousness. There are many examples of this in mathematics; for example, functionals (functions of functions). But what are the functions of?

We have just seen above that real valued (material) quantities are sufficient statistics of probability distributions. By induction, this implies that the operation of consciousness on a sufficient statistic produces a sufficient statistic, which entails its own probability distribution. This means that if we associate consciousness with operations that produce the sufficient statistics of probabilistic beliefs, then it is perfectly possible to be conscious of being conscious (to have beliefs about beliefs or to have probability distributions over probability distributions).

This can be expressed formally in terms of a consciousness operator C that operates sensory data S to produce the sufficient statistics a probability distribution over their causes $\psi^{(1)}$. Similarly, applying the operator to the ensuing sufficient statistics produces beliefs about the causes of the causes, and so on (see also Figure 1):

$$
\begin{array}{ll}
\textit{res extensa}: & \textit{res cogitans}: \\
\mu^{(1)} = C \circ S & Q(\psi^{(1)} \mid \mu^{(1)}) \approx P(\psi^{(1)} \mid S,m) \\
\mu^{(2)} = C \circ C \circ S = C \circ \mu^{(1)} & Q(\psi^{(2)} \mid \mu^{(2)}) \approx P(\psi^{(2)} \mid \psi^{(1)},m) \\
\mu^{(3)} = C \circ C \circ C \circ S = C \circ \mu^{(2)} & Q(\psi^{(3)} \mid \mu^{(3)}) \approx P(\psi^{(3)} \mid \psi^{(2)},m) \\
\vdots & \vdots
\end{array}
$$

$$
\prod_i Q(\psi^{(i)} \mid \mu^{(i)}) \approx P(\psi^{(1)}, \psi^{(2)}, \psi^{(3)}, \ldots \mid S,m) \tag{1}
$$

This hierarchical composition of belief operators is at the heart of everyday statistical modelling with hierarchical models. For example, the summary statistic procedure for mixed effects analyses of within and between subject effects (Kass and Steffey, 1989), where the mean of each subject could correspond to $\mu^{(1)}$ and the group mean to $\mu^{(2)}$. Hierarchical aspects of statistical inference are also found in finance and economics theory; for example, the distinction between risk or known uncertainty and ambiguity or unknown uncertainty (Kahneman and Tversky, 1979).

To make the argument a bit less abstract, consider the following example. Suppose you were supplied with wavelength selective

sensory data from a small patch of cones in your retina. The relative amount of each wavelength can be explained by a finite number of colours entertained by your model of the colourful world. The belief operator above would select the sufficient statistics encoding a probability distribution over the competing hypotheses (colours) — encoding your subsequent belief that 'red' was the best explanation for these sensory data.

Note that 'red' is a fictive cause of the data, not a sufficient statistic — it does not exist other than as the support of a probability distribution. It is this belief we associate with qualia. Imagine now that you have access to the sufficient statistics inducing qualia from multiple patches of retinotopically mapped colours and hues. You then hierarchically optimize the next level of sufficient statistics to find the best hypothesis that explains the sufficient statistics at the retinotopically mapped level — and you select a belief that they are caused by a red rose. Again, the rose does not in itself exist other than to support a probability distribution associated with sufficient statistics — say neural activity. The key thing here is that the hypotheses underpinning (supporting) beliefs are specified by a generative model. This model furnishes a virtual reality that is used to explain sensory impressions through the act of inference.

Although we have taken some technical liberties above, the emergent distinction between conscious operations on material quantities — and the immaterial beliefs that are produced — highlights the nature of the dualism that underlies embodied beliefs. This dualism is resolved by associating conscious processing with probabilistic (Bayesian) updating, where material updates are specified by immaterial beliefs. The Bayesian aspect of this inference is highlighted by the last equality above, which shows that the cumulative product of beliefs is the posterior probability distribution over causes — at different hierarchical levels of description — given some sensations and a generative model.

Statistical inference is also an integral aspect of nearly every approach to biological self-organization (Ao, 2009; Kauffman, 1993) — ranging from the notion of unconscious inference (Helmholtz, 1866/1962) discussed above to the modelling perspective provided by Ashby and colleagues (Ashby, 1947; Conant and Ashby, 1970). In short, there is a nice consilience between inference, theoretical treatments of biological self-organization, and the constraint implied by an affirmative answer to the question: can one be conscious of being conscious?

The picture of consciousness that is emerging here is that consciousness is an operation that produces beliefs and is therefore quintessentially inferential in nature. For example, qualia are the product (beliefs) of inference on sensory data and access consciousness is the process of hierarchical inference that operates on qualia or the products of phenomenal consciousness.

Can we be conscious of being conscious to arbitrarily high order?

This question starts to address the hard aspects of access consciousness, in the sense that to understand what it is to be conscious one has to have a belief about consciousness (*cf.* theory of mind). In other words, one has to have probabilistic beliefs about probabilistic beliefs about probabilistic beliefs, and so on, *ad infinitum*. Clearly, the answer to this question is no. Heuristically, imagine that you wanted to paint a picture of yourself painting. In other words, the picture would be a picture of you painting a picture of you, painting a picture of you, and so on. It is clear that at some scale you will run out of canvas as the painting gets too big for the universe. More technically, there is an upper bound on the depth of models that embody hierarchical belief structures, because the number of sufficient statistics has to be finite.

In short, it is not possible to be conscious of being conscious to an arbitrarily high order; perhaps some people can attain second-, third-, or perhaps even fourth-order consciousness but probably not much beyond this. Indeed, an upper bound on the depth of recursion is a key factor in formal treatments of *sophistication* and optimal decision theory that underlies bounded rationality (Camerer, 2003; Yoshida, Dolan and Friston, 2008). Perhaps this is the hard part of the problem of consciousness, despite the fact that the answer is easy — no.

Are there imperatives for consciousness?

Here the answer is again a straightforward yes. If we associate conscious processes with the formation of hierarchically composed beliefs, then there are fundamental imperatives that govern these beliefs and the attending processes of consciousness. There are many schemes in artificial intelligence and theoretical neurobiology that have been proposed in this role. Our own work focuses on variational free energy minimization (Friston, 2009; Hobson, 2013). Heuristically, this means that probabilistic beliefs will minimize free energy or — more simply — try to provide a better account of the sensorium S under some virtual reality model m:

$$\mu = C \circ S = \arg\min_{\mu} F(S, Q(\psi \mid \mu')) \\ F \geq -\ln P(S \mid m) \tag{2}$$

This free energy is not abstract quantity — it can be measured precisely (given the mapping between biophysical states encoding hierarchical beliefs and the form of the probability distributions that constitute those beliefs). Furthermore, variational free energy is not a thermodynamic construct (although it borrows its name from thermodynamics): it is a statistical quantity whose minimization grandfathers all classical and Bayesian inference; at its simplest, it is the sum of squared prediction error. Free energy is essentially a measure of surprise about sensations (the inequality above), where conscious beliefs are unpacked hierarchically to predict sensory samples. Crucially, this means that conscious beliefs have a measure of theoretic underpinning. In other words, one can quantify the attributes of beliefs and make some clear statements about how those attributes will change over time.

In the context of variational free energy minimization, beliefs will try to find a lower energy state, very much like a physical force tries to reduce the potential energy of a massive object. However, variational free energy is not an attribute of physical states — it is an attribute of a probability distribution (belief) entailed by those states. In other words, conscious processing is equipped with a measure that is an attribute of beliefs. This can be contrasted with the analogous concept of thermodynamic free energy in statistical physics which — it could be argued — measures the physical state of systems. However, the arguments presented in response to the first question suggest that even thermodynamic free energy is an emergent property.

Is consciousness governed by the laws of physics?

Yes and no. If we allow ourselves to equate consciousness with inference — and, in particular, *approximate Bayesian inference* (Beal, 2003; Fox and Roberts, 2011) about the world; then consciousness is lawful and conforms to fairly straightforward mathematical principles (see Equation 2). This is because any approximate Bayesian inference can be cast in terms of minimizing variational free energy. This point is quite fundamental to our argument: Equation 2 shows that the material or extensive products of consciousness are determined by the immaterial beliefs that define variational free energy. There is nothing mysterious about this — the laws of thermodynamics describe how sufficient statistics — like temperature or energy — change as functionals of probability distributions. The only difference is that the

probability distributions in thermodynamics are about (unobservable) microscopic states, whereas the beliefs in inference are about (unobservable) macroscopic states of a virtual world.

However, variational free energy is not the thermodynamic free energy associated with statistical physics. In other words, the laws of statistical physics are not the laws governing variational free energy minimization. Variational free energy is an attribute of beliefs (access or phenomenal consciousness). Crucially, because free energy measures the surprise about some outcome under a model, it can be expressed as *accuracy* minus *complexity*. Intuitively, a belief that makes accurate predictions or explains sensory data accurately ameliorates the surprise associated with those sensations. However, this is not the whole story — the explanations have to be parsimonious to minimize free energy, because they also have to minimize complexity. This is nothing more than Occam's razor formalized as approximate Bayesian inference.

Complexity is important because it provides a deep link with the laws of physics and thermodynamic free energy. In brief, it is fairly easy to show that when complexity is minimized, the brain minimizes its thermodynamic free energy. This follows because when the brain is deprived of sensory perturbations, for a sufficiently long period of time, it will minimize thermodynamic free energy as it approaches equilibrium. In this state, complexity is also minimized because there are no sensations to explain with any accuracy. This means that the laws of (statistical thermodynamic) physics and the lawful process of inference (consciousness) are connected formally. The first is based upon macroscopic processes pertaining to probability distributions over the microstates of a canonical ensemble, while the second pertains to probability distributions over hidden causes of sensory exchange with the external milieu. In other words, thermodynamic free energy is a measure of the information *within* a physical system (neural states or sufficient statistics), while variational free energy is a measure of information *about* inferred causes. In short, inference (consciousness) has an imperative that goes beyond the laws of statistical thermodynamics but operates under these laws. We have introduced complexity here to address a key issue raised in the introduction:

Can we be conscious when asleep?

If consciousness is inference, then can conscious processes exist when there is nothing to infer? In other words, in the absence of an

active sampling or exposure to the sensorium, does inference make any sense? An obvious consideration here is sleep, where neuromodulatory gating of sensory input deprives the brain of sensations to explain. According to the above argument, the brain will therefore try to minimize complexity (and its thermodynamic free energy). Does this imply the suspension of consciousness? The answer is no, because the very process of minimizing complexity is part of inference (by virtue of minimizing variational free energy).

We have considered this complexity minimization in some detail elsewhere (Hobson and Friston, 2012). In brief, even in the absence of sensory information, inference can still proceed, because we can model data (acquired during wakefulness) to compare competing hypotheses (during sleep). This comparison rests on generating fictive sensations, using generative or virtual reality models — and then optimizing the model to minimize complexity or redundancy. This process has been considered — from both a phenomenological and neurobiological perspective — in terms of dreaming. This perspective on sleep suggests that it is an integral and possibly necessary part of the inferential processes that we equate with consciousness and suggests that sleep is just a special instance of conscious processing that is untethered from the sensorium.

Clearly, the suspension of sensory input during sleep may not be complete — indeed the eye movements in REM sleep depend on proprioceptive input from the oculomotor system. Furthermore, there is evidence that auditory input is processed to some level. For example, Hoelscher, Klinger and Barta (1981) show that spoken words influence subsequent REM dream content — if they are associated with the sleeper's goals. The finding is consistent with other effects of goal-related stimuli on subsequent thought content (Klinger, 2013). However, the imperative to minimize complexity still prevails, which might provide an interesting perspective on introspective brain states. For example, Smallwood (2011; 2013) has posited — and found evidence for — decoupling from external stimulation during waking mind-wandering. Interestingly, Baird *et al.* (2012) have shown that mind-wandering states also have a beneficial effect on incubation for subsequent task performance.

Can consciousness be experimentally altered in sleep?

Here again, the answer is an encouraging yes. Studies of lucid dreaming (Voss *et al.*, 2009) indicate that a wake-like state of consciousness can be introduced into REM sleep by pre-sleep autosuggestion. We

discuss the important implications of these experimental findings elsewhere (Hobson, 2009b) but here suggest that this paradigm exposes our theoretical considerations to empirical scrutiny. Beyond lucid dreaming, the modifiability of dream content has been demonstrated by Hoelscher, Klinger and Barta (1981) — with stimuli administered during sleep — and Nikles *et al.* (1998) — with instructions prior to sleep (provided that the suggested content is goal-related). Testing the Cartesian theatre in sleep is now a laboratory programme, not just a philosophical speculation.

Is consciousness causal?

The answer here bears on the issue of free will. The 'consciousness as inference' account provides a clear answer to questions of causality and free will; namely, consciousness is causal and will is free. The underlying argument appeals to non-reductive physicalism, which we take to mean that mental properties form a separate ontological class to physical properties. In other words, mental states (such as probabilistic beliefs or qualia) are not ontologically reducible to physical states (such as neural states or sufficient statistics). However, beliefs are instantaneously specified by neural states. This means that there can be no temporal dissociation between the neurophysiological states preceding willed or intended movements and the beliefs that underlie those intentions. This is not to deny that the reporting of these beliefs can occur after the beliefs are in place (Libet *et al.*, 1983) or the existence of illusory meta-beliefs (Wegner, 2002) or the deterministic neural dynamics that underlie inference (see below). However, the beliefs causing movement and choice are causally and instantaneously bound to movements and choices *per se*.

To see this clearly, we have to look a bit more closely at embodied or active inference (Friston, Mattout and Kilner, 2011). In active inference, the inferred states of the world include the trajectory of our bodies and their relationship to the environment. These beliefs are fulfilled through classical motor reflexes, thereby minimizing surprise (variational free energy) to produce the sensations predicted. In effect, this means that wilful movement is prescribed by prior beliefs about what will happen to our bodies next. There is a large body of anatomical and physiological evidence suggesting that motor commands are — in fact — descending predictions about the proprioceptive consequences of intended or willed behaviour (Adams, Shipp and Friston, 2013). From the point of view of phenomenal consciousness, this suggests that intentional qualia (in the motor domain) correspond to

beliefs about action that enslave peripheral reflex arcs to cause movement. In turn, this movement changes the sensations that are sampled from the environment — to minimize uncertainty about their causes. This provides a straightforward explanation for the way we sample the sensorium (e.g. visual search) that is consistent with empirical evidence (Friston *et al.*, 2012).

Having said this, the sufficient statistics (or neural states) encoding those beliefs conform to (deterministic) dynamics that should minimize surprise or variational free energy. In this sense, beliefs are caused by sensory samples and one could argue that there is a circular causality inherent in the ensuing action perception cycle (Fuster, 2001) that underlies active inference. In other words, willed or intentional sampling of the environment causes the sensations that induce the beliefs that cause the sampling.

The results of lucid dream experiments validate this account of 'action as inference'. Subjects can be trained, in waking, to increase their lucidity in sleep. Once lucid in sleep, they can execute motor commands (voluntary eye movements) and alter dream plots (Brylowski, Levitan and LaBerge, 1989). Interestingly — in contrast to oculomotor reflexes — peripheral (Hoffman) reflexes are attenuated (*ibid.*) — so the brain never knows that its descending motor predictions are unfulfilled. Thus, even dream consciousness is motoric and it can be manipulated by intentional means. Qualia may thus be nothing more or less than inferences about sensorimotor schemata that subserve the operation of the virtual reality model postulated by protoconsciousness theory (Hobson, 2009a). In this setting, conscious processing allows potential scenarios to be evaluated and approved or cancelled. An important aspect of this model is the evaluation of competing beliefs, in terms of their adequacy to explain — or bring about — the states of being we believe we should be in.

Conclusion

We have considered some hard questions and found that there are easy answers. Although this discussion is heuristic, some interesting conclusions emerge. First, consciousness is not a hard thing to understand, describe, or make hypotheses about — if one associates it with inference based on deeply structured hierarchical (probabilistic) beliefs about sensations. Crucially, these beliefs are based upon a model of the world that can generate virtual realities. Furthermore, by virtue of the circular causality implicit in the slaving principle, consciousness enslaves microscopic brain states and microscopic brain

states cause consciousness. The particular macroscopic quantities (measures) of interest here are probability distributions that have well-defined attributes such as entropy and (free) energy. These quantities conform to lawful dynamics that can be measured and understood in a pragmatic way — and indeed form the basis of most of cognitive neuroscience: e.g. studies of the neural code, perceptual synthesis, functional integration in the brain, repetition suppression, electrophysiology, learning, and so on (Friston, 2009).

This analysis suggests something quite interesting; namely, that consciousness is inference. This is an interesting perspective because it explains many aspects of consciousness in a fairly simple fashion. First, consciousness (inference) is a process that is entailed or embodied by changes in representational (material) states. However, the imperative for these changes can only be specified in terms of (immaterial) probability distributions or beliefs — and these probability distributions are defined in terms of a virtual reality model or private (Cartesian) theatre.

Altered states of consciousness and sleep

Casting consciousness as inference also provides an illuminating perspective on altered states of consciousness. This is easy to understand in terms of the relationship between inference and the (sensory) data on which inference is based. In a Bayesian setting, the relative influence of prior beliefs (relative to sensory evidence) depends heavily upon the *precision* of (or confidence in) data — changing data precision can lead to radically different sorts of inference (consciousness). Precision is another attribute of a probability distribution and is simply the inverse variance.

An interesting example here is sleep, in which the precision of sensory input is effectively abolished — through neuromodulatory gating or chemical mechanisms that induce sleep (Hobson, 2009a). This does not mean that inference or consciousness is abolished; more that the nature of inference is altered to focus on minimizing model complexity and simplifying our generative models of the world. This may sound fanciful, but scientists do this every day: they spend a short amount of time earnestly acquiring data from carefully designed experiments and then study those data using Bayesian model comparison to test different hypotheses — until they find one that provides the most accurate but parsimonious explanation. In one sense, this is precisely what we do: acquiring experiential data through designed interactions with the environment and then — in an altered state of

(sleeping) consciousness — we finesse the complexity of those models, until morning breaks.

The importance of precision in nuancing hierarchical inference (consciousness) is also interesting in relation to the action of psychotropic (and psychedelic) drugs that, universally, act on neuromodulator brain systems. These systems are exactly those thought to encode precision in the brain (Friston, 2009). Furthermore, the altered states of consciousness associated with psychopathology (e.g. schizophrenia) again implicate exactly the same systems; namely, classical neuromodulator transmitter systems (like the dopaminergic system) (Adams, Perrinet and Friston, 2012).

Lucid dreaming and conscious states

The science of lucid dreaming (Hobson, 2009b) raises serious questions about our proposed revisit to the Cartesian theatre. In non-lucid dreaming, brain activation, input–output gating, and neuromodulation conspire to produce a distinctive kind of consciousness, characterized by endogenous perceptions, the delusional belief that one is awake, bizarre incongruities and discontinuities, strong emotion (especially fear, elation, and anger), and practically total amnesia (Hobson, 2013). In normal dreaming all is unified and all is illusional.

By chance and enhanced by systematic pre-sleep autosuggestion, it is possible for subjects to become lucid and thus become aware that they are dreaming (Brylowski, Levitan and LaBerge, 1989). Once lucid, subjects can observe their dream as if they were at the theatre. Moreover, they can influence dream content and even voluntarily wake themselves — so as better to recall and control their subjective experience. In other words, they can choose the theatre and direct the show!

Dream lucidity is correlated with frontal brain activation (Voss *et al.*, 2009), which presumably allows the vaunted unity of self to be divided, such that there is a dreamer and an observer who coexist and interact dynamically. This, of course, does not mean that dreaming (or waking) is independent of the brain. Rather it clearly indicates that the mind–brain can be functionally split such that one part is awake while another is asleep — something that can be verified electrophysiologically, even in rats (Vyazovskiy *et al.*, 2011). The fact that the brain can be both a participant and an observer speaks again to hierarchical inference and is evidenced in its hierarchical neuroanatomy. As noted above, it is perfectly possible to make inferences about the products of inferences — even to infer that one is dealing with fictive or illusory

data. In fact, statisticians do this all the time when they use Monte Carlo simulations to compare models of real data with null hypotheses.

These facts have a powerful bearing upon our assumptions about how consciousness is engendered by the brain. We are forced to conclude that we live in something like a theatre and, while it is certainly not Cartesian, it does have properties that lend themselves to the sort of neurobiological and cognitive specification that we attempt to demonstrate in this paper.

Finally, associating consciousness with inference gets to the heart of the hard problem, in the sense that inferring that something is red is distinct from receiving selective visual sensations (visual data) with the appropriate wavelength composition. Furthermore, you can only see your own red that is an integral part of your virtual reality model. You cannot see someone else's red or another red because they are entailed by another model or hypothesis. In short, you cannot see my red — you can only infer that I can see red. In one sense, tying consciousness to active inference tells one immediately that consciousness is quintessentially private. Indeed, it is so private that other people are just hypotheses in your virtual reality model. In one sense, these ideas are also your ideas (however latent), because you have to know what you are going to see next before you can confirm it by reading these words — this is the essence of active inference and how we sample the world to minimize surprise.

Epilogue

Having completed this paper there was — between us — a lingering suspicion we had not addressed the hard problem of how the brain becomes conscious in a subjective sense. We thought it would be useful to acknowledge this by sharing our correspondence in the form of an epilogue.

AH to KF: 'At some point, we need to make it even clearer that we do not yet know how the brain becomes conscious. In my opinion this is an integral part of the hard problem and it persists…

…The really hard problem is to model subjectivity. I accept the functionalist approach but yearn for more, something like Eccles' 'psychons' — something more plausible and internal. I do not suppose that the brain–mind is influenced by anything like spiritual forces emanating from outer space — a Godhead — or the ghosts of dead people but I am at a loss to say exactly how a self arises or how that self constructs its model of the world. This, to me, is unfinished

business and, hence, an obdurate component of the hard problem. I hope that theoretical clarification will sensitize empirical investigation and bring protoconsciousness theory into register with cutting edge philosophical modelling of the self (Metzinger, 2003). Is energy also information? Are waves and particles a possibility? This seems to be what the quantum boys are betting on. How about you? What are qualia according to you?'

KF to AH: 'I think that there could be answers to these mechanistic questions — but much rests on deconstructing the sorts of answers people expect to hear. For example, if you asked me how gravity causes water to flow downhill, you might expect an appeal to Newtonian mechanics, which you would probably find quite satisfying. However, there are more universal accounts of the way things are [observed]. For example:

Classical (Newtonian) mechanics says that the path followed by a physical system minimizes action, where action is the path integral of a Lagrangian or energy. In other words, action satisfies a variational principle — the principle of stationary action — such that classical equations of motion can be derived from minimizing action (as opposed to solving differential equations). You will probably remember doing Hamilton's principle of least action at school? Crucially, the same principle applies in quantum mechanics and field theory. An important example here is Feynman's path integral formulation, where the probability of any path depends upon its action (note that the Schrödinger equation can be recovered from the path integral formulation). This is important because the 'consciousness as inference' argument is based upon exactly the same principle of stationary action, where the Lagrangian is variational free energy (used in approximate Bayesian inference).

This means that if you ask me "How does consciousness cause physical changes in the brain?", then I would answer: "consciousness causes perception through the variational principle of stationary action — that describes the physical states of a sentient system — because action is a function of sensations and qualia. Here, qualia are probability distributions over the hidden causes of sensations." This explanation is formally identical to an account of falling objects in terms of gravity. Here, consciousness is not an epiphenomenon — it is an integral part of how physical states change (as intimated by Equation 2). In other words, qualia are not just entailed (or induced) by physical states (sufficient statistics) — they determine the path of those states and so close the causal loop between the material and immaterial.

Mathematically, I think the deep challenge is to prove the existence of a duplet — comprising a generative model and variational density (whose marginals would correspond to qualia) — for any physical system that could be considered sentient. This may be easier than it sounds (I will send you something soon that speaks to this). Interestingly, I have just finished a compelling monograph by Sir Allan Cook (Cook, 1994), who argues that the very nature of (quantum and relativistic) physics derives from (the invariant properties of) observations and Bayesian inference.

In relation to "psychons", I would submit that "qualia" are quite sufficient and that it is only necessary to associate qualia with the variational densities entailed by physical states that — happily — are usually denoted by Q. In fact, I often joke — with a straight face — that this is why Q is used. "Q for qualia" would be a nice title for another paper?'

AH to KF: 'Philosophical: we are dual aspect monists, not Cartesian dualists. Why not say so?; "we do not believe in a Cartesian dualist theatre but we are forced to consider something like a theatre when we discuss consciousness, especially when we consider that presence of a self or agent as an integral part of the virtual reality model." This is a crucial point and we make it clear after we get into the paper but we renounce the homunculus a bit too cavalierly at the outset. I think we should say something like this in the intro: "Our understanding of the brain–mind as a theatre is not Cartesian in that we renounce dualism. We put in its place, a dual aspect monism and explain how the two aspects depend on each other, differ from each other and how they interact (in both directions!) causally."

Tone and Stance: I think that the tone of the paper could be more cautious and tentative without detracting from its boldness. For example, I do not believe that consciousness is only inference. It is also famously reflective; i.e. theatrical. I read your draft and consider its content reflectively. I see you sitting next to me in the dining room at the Hotel Russell. My inferences, in other words, are embedded in a richly nuanced set of perceptions and feelings. I am a subject trying to figure out how I could also be an object. How about: "We are surprised at the concordance of our approaches and, without diminishing our differences, attempt to communicate in this paper our shared vision of a science of consciousness."'

KF to AH: 'These suggestions are very nice — and I have adjusted the introduction accordingly. I fully concur with your points about reflection and introspection — I think this relates closely to the question about subjectivity above. It also touches on the growing focus on

prediction in consciousness research; for example, Andy Clark's compelling synthesis (Clark, 2013) and the nice work on predicting (subjective) interoceptive bodily states from Anil Seth's group (Seth, Suzuki & Critchley, 2011).

Reflection or rehearsal within a virtual reality model is a vital part of inference and brings along with it all sorts of mnemonic and prospective capacities — like imagination and planning. I suspect that the frustratingly impenetrable problem of subjectivity is a necessary price that we pay for models of a fictive future that can entertain alternative hypotheses. Just to sketch ideas that we can elaborate on elsewhere: if inference is about what will be and what can be, then one is in the insidious position of entertaining null hypotheses that can never be falsified. Many of these are existential in nature. For example, could I operate without access consciousness? Do philosophical zombies have conscious awareness? Would I be self-conscious without an internal narrative? And so on. The key point here is that none of these alternative states of affairs can ever be verified or falsified. For example, if I could operate without self-consciousness, how would I ever know? I suspect people have pursued this line of argument (with philosophical takes on Gödel's incompleteness theorems). I am not sure about this but I suspect the problem is not so much explaining subjectivity as a remarkable fact but explaining the fact that it is remarkable?'

Acknowledgments

This work was funded by the Wellcome Trust, the US National Institute of Mental Health, the National Science Foundation, and the MacArthur Foundation. We would like to thank an anonymous reviewer of this work for helpful guidance in presenting these ideas.

References

Adams, R.A., Perrinet, L.U. & Friston, K. (2012) Smooth pursuit and visual occlusion: Active inference and oculomotor control in schizophrenia, *PLoS One*, **7** (10), p. e47502.

Adams, R.A., Shipp, S. & Friston, K.J. (2013) Predictions not commands: Active inference in the motor system, *Brain Structure and Function*, **218** (3), pp. 611–643.

Allison, T. & Cicchetti, D.V. (1976) Sleep in mammals: Ecological and constitutional correlates, *Science*, **194** (4266), pp. 732–734.

Ao, P. (2009) Global view of bionetwork dynamics: Adaptive landscape, *Journal of Genetics and Genomics*, **36** (2), pp. 63–73.

Aserinsky, E. & Kleitman, N. (1953) Regularly occurring periods of ocular motility and concomitant phenomena during sleep, *Science*, **118**, pp. 361–375.

Ashby, W.R. (1947) Principles of the self-organizing dynamic system, *Journal of General Psychology*, **37**, pp. 125–128.
Baars, B.J. (1997) *In the Theater of Consciousness*, New York: Oxford University Press.
Baird, B., Smallwood, J., Mrazek, M.D., Kam, J.W.Y., Franklin, M.S. & Schooler, J.W. (2012) Inspired by distraction: Mindwandering facilitates creative incubation, *Psychological Science*, **23**, pp. 1117–1122.
Beal, M.J. (2003) *Variational Algorithms for Approximate Bayesian Inference*, PhD thesis, University College London.
Block, N. (1998) On a confusion about a function of consciousness, in Block, N., Flanagan, O. & Guzeldere, G. (eds.) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.
Brylowski, A., Levitan, L. & LaBerge, S. (1989) H-reflex suppression and autonomic activation during lucid REM sleep: A case study, *Sleep*, **12** (4), pp. 374–378.
Camerer, C.F. (2003) Behavioural studies of strategic thinking in games, *Trends in Cognitive Sciences*, **7** (5), pp. 225–231.
Carr, J. (1981) *Applications of Centre Manifold Theory*, Berlin: Springer-Verlag.
Chalmers, D. (1995) Facing up to the problem of consciousness, *Journal of Consciousness Studies*, **2** (3), pp. 200–219.
Clark, T.W. (1999) Fear of mechanism: A compatibilist critique of The Volitional Brain, *Journal of Consciousness Studies*, **6** (8–9), pp. 279–293.
Clark, A. (2000) A case where access implies qualia, *Analysis*, **60** (1), pp. 30–38.
Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science, *Behavioral and Brain Sciences*, **36** (3), pp. 181–204.
Conant, R.C. & Ashby, R.W. (1970) Every good regulator of a system must be a model of that system, *International Journal of Systems Science*, **1** (2), pp. 89–97.
Cook, A. (1994) *The Observational Foundations of Physics*, Cambridge: Cambridge University Press.
Cox, R.T. (2001) *Algebra of Probable Inference*, Baltimore, MD: Johns Hopkins University Press.
Davis, M.J. (2006) Low-dimensional manifolds in reaction-diffusion equations. 1. Fundamental aspects, *Journal of Physical Chemistry A*, **110** (16), pp. 5235–5256.
Dayan, P., Hinton, G.E. & Neal, R. (1995) The Helmholtz machine, *Neural Computation*, **7**, pp. 889–904.
Dehaene, S. & Changeux, J.-P. (2011) Experimental and theoretical approaches to conscious processing, *Neuron*, **70**, pp. 200–227.
Dennett, D. (1991) *Consciousness Explained*, London: Allen Lane/The Penguin Press.
Edelman, G. (2001) Consciousness: The remembered present, *Annals of the New York Academy of Sciences*, **929**, pp. 111–122.
Felleman, D. & Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex, *Cerebral Cortex*, **1**, pp. 1–47.
Fox, C. & Roberts, S. (2011) A tutorial on variational Bayes, in Robertson, D. (ed.) *Artificial Intelligence Review*, Berlin: Spinger.
Frank, T.D. (2004) *Nonlinear Fokker-Planck Equations: Fundamentals and Applications. Springer Series in Synergetics*, Berlin: Springer.
Friston, K. (2009) The free-energy principle: A rough guide to the brain?, *Trends in Cognitive Sciences*, **13** (7), pp. 293–301.
Friston, K. (2012) A free energy principle for biological systems, *Entropy*, **14**, pp. 2100–2121.

Friston, K., Mattout, J. & Kilner, J. (2011) Action understanding and active inference, *Biological Cybernetics*, **104**, pp. 137–160.
Friston, K., Adams, R.A., Perrinet, L. & Breakspear, M. (2012) Perceptions as hypotheses: Saccades as experiments, *Frontiers in Psychology*, **3**, p. 151.
Fuster, J.M. (2001) The prefrontal cortex — an update: Time is of the essence, *Neuron*, **30**, pp. 319–333.
Gilestro, G.F., Tononi, G. & Cirelli, C. (2009) Widespread changes in synaptic markers as a function of sleep and wakefulness in Drosophila, *Science*, **324** (5923), pp. 109–112.
Ginzburg, V.L. & Landau, L.D. (1950) On the theory of superconductivity, *Journal of Experimental and Theoretical Physics*, **20**, p. 1064.
Gregory, R.L. (1980) Perceptions as hypotheses, *Philosophical Transactions of the Royal Society of London B*, **290**, pp. 181–197.
Haken, H. (1983) *Synergetics: An Introduction. Non-equilibrium phase transition and self-selforganisation in physics, chemistry and biology*, 3rd ed., Berlin: Springer.
Helmholtz, H. (1866/1962) Concerning the perceptions in general, in *Treatise on Physiological Optics*, 3rd ed., New York: Dover.
Hilakivi, I. (1987) Biogenic amines in the regulation of wakefulness and sleep, *Medical Biology*, **65** (2–3), pp. 97–104.
Hobson, J.A. (2009a) REM sleep and dreaming: Towards a theory of protoconsciousness, *Nature Reviews Neuroscience*, **10** (11), pp. 803–813.
Hobson, J.A. (2009b) The neurobiology of consciousness: Lucid dreaming wakes up, *International Journal of Dream Research*, **2** (2), pp. 41–44.
Hobson, J.A. (2013) *Dream Consciousness*, Berlin: Springer.
Hobson, J.A. & Wohl, H. (2005) *From Angels to Neurons*, Parma: Mattioli 1885.
Hobson, J.A. & Friston, K.J. (2012) Waking and dreaming consciousness: Neurobiological and functional considerations, *Progress in Neurobiology*, **98** (1), pp. 82–98.
Hoelscher, T.J., Klinger, E. & Barta, S.G. (1981) Incorporation of concern and nonconcern related verbal stimuli into dream content, *Journal of Abnormal Psychology*, **49**, pp. 88–91.
Kahneman, D. & Tversky, A. (1979) Prospect theory: An analysis of decision under risk, *Econometrica*, **47** (2), pp. 263–291.
Kass, R.E. & Steffey, D. (1989) Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models), *Journal of the American Statistical Association*, **407**, pp. 717–726.
Kauffman, S. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford: Oxford University Press.
Klinger, E. (2013) Goal commitments and the content of thoughts and dreams: Basic principles, *Frontiers in Psychology*, **4**.
Knill, D.C. & Pouget, A. (2004) The Bayesian brain: The role of uncertainty in neural coding and computation, *Trends in Neurosciences*, **27** (12), pp. 712–719.
Kripke, D.F., Garfinkel, L., Wingard, D.L., Klauber, M.R. & Marler, M.R. (2002) Mortality associated with sleep duration and insomnia, *Archives of General Psychiatry*, **59**, pp. 131–136.
Landau, L.D. & Lifshitz, E.M. (1976) *Statistical Physics: Course of Theoretical Physics 5*, 3rd ed., Oxford: Pergamon Press.
Libet, B., Gleason, C.A., Wright, E.W. & Pearl, D.K. (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act, *Brain*, **106** (3), pp. 623–642.
Manuel, B.D.J. (2001) *Philosophy of Man: Selected Readings*, Manila: Goodwill Trading Co.

Mesulam, M.M. (1998) From sensation to cognition, *Brain*, **121**, pp. 1013–1052.
Metzinger, T. (2003) *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, MA: MIT Press.
Mormann, F. & Koch, C. (2007) Neural correlates of consciousness, *Scholarpedia*, **2** (12), p. 1740.
Nikles, C.D.I., Brecht, D.L., Klinger, E. & Bursell, A.L. (1998) The effects of current-concern- and nonconcern-related waking suggestions on nocturnal dream content, *Journal of Personality and Social Psychology*, **75**, pp. 242–255.
Parmeggiani, P.L. (2007) REM sleep related increase in brain temperature: A physiologic problem, *Archives Italiennes de Biologie*, **145** (1), pp. 13–21.
Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Fransisco, CA: Morgan Kaufmann.
Rechtschaffen, A., Bergmann, B.M., Everson, C., Kushida, C.A. & Gilliland, M.A. (1989) Sleep deprivation in the rat: X. Integration and discussion of the findings, *Sleep*, **12** (1), pp. 68–87.
Roffwarg, H.P., Muzio, J.N. & Dement, W.C. (1966) Ontogenetic development of the human sleep-dream cycle, *Science*, **152** (3722), pp. 604–619.
Schooler, J.W., Smallwood, J.K.C., Handy, T.C., Reichle, E.D. & Sayette, M.A. (2011) Meta-awareness, perceptual decoupling and the wandering mind, *Trends in Cognitive Sciences*, **15**, pp. 319–326.
Seth, A.K., Suzuki, K. & Critchley, H.D. (2011) An interoceptive predictive coding model of conscious presence, *Frontiers in Psychology*, **2**, p. 395.
Smallwood, J. (2011) Mind-wandering while reading: Attentional decoupling, mindless reading and the cascade model of inattention, *Language and Linguistics Compass*, **5** (2), pp. 63–77.
Smallwood, J. (2013) Distinguishing how from why the mind wanders: A process–occurrence framework for self-generated mental activity, *Psychological Bulletin*, **139** (3), pp. 519–535.
Stickgold, R. & Hobson, J.A. (1995) The conscious state paradigm: A neurocognitive approach to waking, sleeping, and dreaming, in Gazzaniga, M.S. (ed.) *The Cognitive Neurosciences*, Cambridge, MA: MIT press.
Tononi, G.E.G. (2000) *Consciousness: How Matter Becomes Imagination*, London: Allen Lane.
Voss, U., Holzmann, R., Tuin, I. & Hobson, J.A. (2009) Lucid dreaming: A state of consciousness with features of both waking and non-lucid dreaming, *Sleep*, **32** (9), pp. 1191–1200.
Vyazovskiy, V., Olcese, U., Hanlon, E.C., Nir, Y., Cirelli, C. & Tononi, G. (2011) Local sleep in awake rats, *Nature*, **472** (7344), pp. 443–447.
Wegner, D.M. (2002) *The Illusion of Conscious Will*, Cambridge, MA: MIT Press.
Yoshida, W., Dolan, R.J. & Friston, K.J. (2008) Game theory of mind, *PLoS Computational Biology*, **4** (12), p. e1000254.

Paper received June 2013; revised Ocotober 2013.

Glossary of (Bayesian) Terms

Bayesian belief updating: the combination of prior beliefs about the causes of an observation and the likelihood of that observation to produce a posterior belief about its hidden causes. This updating conforms to Bayes rule.

Likelihood: the probability of an observation under a generative model, given its causes.

Prior belief: a probability distribution over the hidden causes of observations, before they are observed.

Posterior beliefs: a probability distribution over the hidden causes of observed consequences, after they are observed.

Hidden causes: the unobserved (possibly fictive) causes of observed data.

Generative model: a probabilistic specification of the dependencies among causes and consequences; usually specified in terms of a prior belief and the likelihood of observations, given their causes.

Sufficient statistics: quantities or parameters that are sufficient to specify a probability distribution; for example, the mean (expectation) and precision (inverse variance) of a Gaussian distribution.

Approximate Bayesian inference: Bayesian belief updating in which (the sufficient statistics of) approximate posterior distributions are optimized by minimizing variational free energy. The approximate posterior convergences to the true posterior when free energy is minimized.

Variational free energy: a functional of a probability distribution (and observations) that upper bounds (is always greater than) the negative log evidence for a generative model. This negative log evidence is also known as surprise or self information in information theory.

Bayesian model evidence: this is the probability that some observations were generated by a model. It is also known as the marginal or integrated likelihood because it does not depend upon the hidden causes.

Surprise: also known as self information or surprisal, surprise is the negative log likelihood of some observations under a generative model.

Complexity: the difference or divergence between prior and posterior beliefs. The complexity of a model reflects the change in prior beliefs produced by Bayesian belief updating.