

Research Article

Cite this article: Wallace R (2019). Do aliens dream of offworld sheep? *International Journal of Astrobiology* **18**, 396–404. <https://doi.org/10.1017/S1473550418000289>

Received: 20 February 2018
Revised: 11 June 2018
Accepted: 21 June 2018
First published online: 30 July 2018

Key words:

consciousness; control theory; culture;
information theory; sleep cycle

Author for correspondence:

Rodrick Wallace, E-mail: rodrick.wallace@gmail.com

Do aliens dream of offworld sheep?

Rodrick Wallace

Division of Epidemiology, The New York State Psychiatric Institute, Box 47, 1051 Riverside Drive, New York, USA

Abstract

The Stanley Miller experiment suggests that amino acid-based life is ubiquitous in our universe, although its varieties will not have followed the particular, highly contingent and path-dependent, evolutionary trajectory found on Earth. Are many alien organisms likely to be individually conscious in ways we would recognize? Almost certainly. Will alien consciousness require a ‘sleep cycle’? A strong argument suggests it will. Can some species develop analogs to culture and high-order technology? Less likely, but still fairly probable. If so, will we be able to communicate with them? Only on a basic level, and only with profound difficulty. The reasoning is fairly direct and involves convolution of a learned heritage system with individual and collective consciousness.

Introduction

In spite of a popular social construction as such (e.g. Penrose, 1994; Tononi, 2012; Tegmark, 2015), individual consciousness is no great mystery, constituting a basic adaptation likely a half-billion years old representing a straightforward evolutionary exaptation of the crosstalk – signal leakage – that bedevils all information transmission, biological or not (Wallace and Wallace, 2009; Wallace, 2012). Bernard Baars’ global workspace/global broadcast model – the current front-runner in the Darwinian competition between consciousness theories (Dehaene and Naccache, 2001) – is itself nearly a generation old and accounts, in a qualitative manner, for much of what we know about individual consciousness (Baars, 1988, 2005):

1. The brain can be viewed as a collection of distributed specialized networks (processors).
2. Individual consciousness is associated with a global workspace in the brain – a fleeting memory capacity whose focal contents are widely distributed (broadcast) to many unconscious specialized networks.
3. Conversely, a global workspace can also serve to integrate many competing and cooperating input networks.
4. Some unconscious networks, called contexts, shape conscious contents, e.g. unconscious parietal maps modulate visual feature cells that underlie the perception of colour in the ventral stream.
5. Such contexts work together jointly to constrain conscious events.
6. Motives and emotions can be viewed as goal contexts.
7. Executive functions work as hierarchies of goal contexts.

Although this basic approach has been the focus of work by many researchers for two decades, scientific consciousness study has only recently, in the context of a deluge of empirical results from brain imaging experiments, begun digesting the perspective and preparing to move on (Baars, 2005).

A first point in developing a quantitative theory of individual consciousness based on Baars’ model is to recognize Dretske’s central argument (Dretske, 1994) that high level mental process of any nature inevitably involves the generation and transmission of information, both of which are constrained by the asymptotic limit theorems of information theory: Shannon Coding, Shannon–McMillan Source Coding, and the Rate Distortion theorems (Khinchin, 1957; Cover and Thomas, 2006).

The next stage is to recognize that cognitive process, in the sense of Atlan and Cohen (1998), involves comparison of a perceived signal with an internal picture of the world, and then, on that comparison, the choice of a response from a large repertoire of those possible, causing a reduction in a formal measure of uncertainty. It is then easy to show (Wallace, 2000, 2005, 2010, 2012) that a substantial class of such cognitive phenomena is necessarily associated with a well-behaved information source.

Extension to less well-behaved systems is via the formalism outlined in Wallace (2018a).

Cognition as an information source

Cognitive pattern recognition-and-selected response, following the model of Atlan and Cohen (1998), proceeds by convoluting an incoming external ‘sensory’ signal with an internal ‘ongoing activity’ – which includes, but is not limited to, a learned or inherited picture of

the world – and, at some point, triggering an appropriate choice of an action from a larger set of those available based on a decision that the pattern of sensory activity requires a response. It is not necessary to specify how the pattern recognition system is trained, and hence possible to adopt a weak model, regardless of learning paradigm, that can be more formally described by the Rate Distortion Theorem that dictates the minimum channel capacity necessary to keep the average distortion in a transmitted signal below some fixed limit (Wallace, 2012). Fulfilling Atlan and Cohen’s criterion of meaning-from-response, it is possible to define a language’s contextual meaning entirely in terms of system output.

Most directly, cognition involves choice, choice reduces uncertainty, and the reduction of uncertainty directly implies the existence of an information source ‘dual’ to that cognition. This argument is breathtakingly sufficient, and places cognition and consciousness within the statistical regularities imposed by the asymptotic limit theorems of both information and control theories, and certain constraints imposed by physical theory through the recognition of information as a form of free energy (e.g. Feynman, 2000). However, a more formal model, a conceptual simplification of the standard neural network, can be made as follows.

A pattern of ‘sensory’ input – incorporating feedback from the external world – is expressed as an ordered sequence y_0, y_1, \dots . This is mixed in a systematic (but unspecified) algorithmic manner with internal ‘ongoing’ activity, a sequence w_0, w_1, \dots , to create a path of composite signals $x = a_0, a_1, \dots, a_n, \dots$, where $a_j = f(y_j, w_j)$ for some function f . This path is then fed into a similarly unspecified decision function generating an output $h(x)$ that is an element of one of two sets B_0 and B_1 . We take $B_0 \equiv \{b_0, \dots, b_k\}$, $B_1 \equiv \{b_{k+1}, \dots, b_m\}$.

The model permits a graded response, supposing that if $h(x) \in B_0$ the pattern is not recognized, and if $h(x) \in B_1$ the pattern is recognized and some action b_j , $k + 1 \leq j \leq m$ takes place.

The principal focus is on those composite paths x triggering pattern recognition-and-response. Given a fixed initial state a_0 , such that $h(a_0) \in B_0$, we examine all possible subsequent paths x beginning with a_0 and leading to the event $h(x) \in B_1$. Thus $h(a_0, \dots, a_j) \in B_0$ for all $0 \leq j < m$, but $h(a_0, \dots, a_m) \in B_1$. The y_j , the ‘sensory’ input convoluted with the internal w_j , contains feedback from the external world, i.e. how well h matches intent with observation. Wallace (2012) uses this comparison to define a Rate Distortion Function and an associated average distortion characterizing the ‘goodness-of-fit’ between what a cognitive entity perceives and what exists.

For each positive integer n let $N(n)$ be the number of grammatical and syntactic high probability paths of length n that begin with some particular a_0 having $h(a_0) \in B_0$ and lead to the condition $h(x) \in B_1$. Call such paths ‘meaningful’ and assume $N(n)$ to be considerably less than the number of all possible paths of length n – pattern recognition-and-response is comparatively rare. The essential assumption is that the longitudinal finite limit

$$H = \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} \tag{1}$$

both exists and is independent of the path x . Call such a cognitive process *ergodic*. The essential idea surrounding an ergodic system is that time averages converge on ensemble averages.

Disjoint partition of state space is done according to sets of states which can be connected by meaningful paths from a particular base point, leading to a natural coset algebra of the system defining a groupoid (Wallace, 2012). A groupoid is an algebraic structure having a product that is not necessarily defined between all elements. The most simple example is a disjoint union of groups (e.g. Brown, 1987).

It is thus possible to define an ergodic information source X associated with stochastic variates X_j having joint and conditional probabilities $P(a_0, \dots, a_n)$ and $P(a_n|a_0, \dots, a_{n-1})$ such that appropriate joint and conditional Shannon uncertainties may be defined which satisfy the standard relations (Cover and Thomas, 2006).

$$\begin{aligned} H[X] &= \lim_{n \rightarrow \infty} \frac{\log[N(n)]}{n} \\ &= \lim_{n \rightarrow \infty} H(X_n|X_0, \dots, X_{n-1}) \\ &= \lim_{n \rightarrow \infty} \frac{H(X_0, \dots, X_n)}{n} \end{aligned} \tag{2}$$

where the Shannon uncertainties H are defined in terms of cross-sectional sums of the form $H = -\sum P_k \log[P_k]$. Again, the P_k constitute probability distributions, summing to 1 (Khinchin, 1957; Ash, 1990; Cover and Thomas, 2006).

This information source is taken as *dual* to the ergodic cognitive process.

Different quasi-languages will be defined by different divisions of the total universe of possible responses into various pairs of sets B_0 and B_1 . Like the use of different distortion measures in the Rate Distortion Theorem, however, the underlying dynamics will all be qualitatively similar. Again, the Rate Distortion Theorem mandates $R(D)$ as the minimum channel capacity needed to ensure that average distortion in a transmitted signal is below the fixed real positive value D , and will vary according to the nature of the channel studied.

Meaningful paths – creating an inherent grammar and syntax – have been defined entirely in terms of system response, as Atlan and Cohen (1998) propose.

In sum, if there are $N(n)$ possible behavioral – and hence temporal – output paths of that information source having length n , then there will be a path-independent limit H such that equation (1) holds.

This is the Shannon uncertainty of the (stationary, ergodic) information source dual to the cognitive process.

Extension to path-dependent ‘nonergodic’ systems, for which the time average is not the same as the ensemble average, can be made using the fact that, for individual sequences $x^n = \{a_0, a_1, \dots, a_n\}$, a path-wise uncertainty $\lim_{n \rightarrow \infty} \mathcal{H}(x^n) = \mathcal{H}(x)$ can always be defined, although it will not be given by a simple Shannon ‘entropy’ (Khinchin, 1957 p. 72). See Wallace (2018a) for details.

Cognitive phase transitions: constructing global workspaces

We envision the evolution of a broad set of unconscious cognitive modules within a reproducing organism that serves a number of independent purposes, ranging from the search for food and habitat, and the avoidance of predation, to modalities of reproduction. Thus even a simple organism will have a large network of

unconscious cognitive modules that, necessarily, interact, if only by crosstalk information leakage.

Evolution can then take this interaction and run with it, across multiple physiological systems at different temporal scales, processes including wound healing, immune response, tumour control, gene expression, sociocultural cognition and so on (Wallace, 2012). The argument is direct.

Given a network of interacting (unconscious) cognitive modules of any sort, we can construct an interacting network of dual information sources, linked by crosstalk – a kind of inevitable information leakage between them.

That is, the information sources dual to an organism's set of cognitive modules are not independent, but are correlated, so that a joint information source can be defined as having the properties

$$H(X_1, \dots, X_n) \leq \sum_{j=1}^n H(X_j) \quad (3)$$

This is the information chain rule (Cover and Thomas, 2006) and has implications for metabolic energy consumption and regulation in a biological process. Recall Feynman's description of how information and free energy have an inherent duality. Feynman, in fact, defines information precisely as the free energy needed to erase a message. The argument is again direct, and for very simple systems it is easy to design an (idealized) machine that turns the information within a message directly into usable work – free energy. Information is a form of free energy (Feynman, 2000) and the construction and transmission of information within living things – the physical instantiation of information – consume considerable free energy, with inevitable massive losses via the second law of thermodynamics.

Suppose an intensity of available metabolic free energy is associated with each defined joint and individual information source $H(X, Y)$, $H(X)$, $H(Y)$, e.g. rates $M_{X,Y}$, M_X , M_Y .

Again, although information is a form of free energy, there is a great entropic loss in its actual expression, so that the probability distribution of a source uncertainty H might be written in the standard Gibbs form (e.g. Landau and Lifshitz, 2007; Pettini, 2007) as

$$P[H] = \frac{\exp[-H/\kappa M]}{\int \exp[-H/\kappa M] dH} \quad (4)$$

assuming κ , the 'translation rate' of metabolic free energy into information, is very small.

To first order, then,

$$\hat{H} = \int H P[H] dH \approx \kappa M \quad (5)$$

and, using equation (3),

$$\hat{H}(X, Y) \leq \hat{H}(X) + \hat{H}(Y) \\ M_{X,Y} \leq M_X + M_Y \quad (6)$$

Thus, as a consequence of the information chain rule, allowing crosstalk consumes a lower rate of metabolic (or other) free energy than isolating information sources. That is, in general, it takes more free energy – higher total cost – to isolate a set of cognitive phenomena than it does allow them to engage in crosstalk,

a signal interaction that, under typical circumstances, grows as the inverse square of the separation between circuits. This is a well-known problem in electrical engineering that can consume considerable attention and other resources for proper address.

For a given underlying network topology of n interacting information sources representing cognitive submodules of an organism – necessarily without isolated subcomponents – the average crosstalk between them, say $\epsilon(n)$, will be a monotonic increasing function of n . For that given network topology, there will be some critical value ϵ_C at which a giant component emerges. That is, at ϵ_C , a very large number of cognitive processes become suddenly linked into a joint 'giant component' information source. For random networks, this phenomenon has been fully studied, and the conditions under which a giant component occurs are well understood (e.g. Spenser, 2010). While other topologies impose different detailed conditions, the sudden emergence of a giant component is virtually universal, covering a very large class of networks. Thus, for any ϵ_C , there will be a critical number of dual information sources n_C such that $\epsilon(n) > \epsilon_C$ if $n > n_C$.

This phase transition – from isolated individual components to a single 'giant component' – represents the onset of large-scale, dynamic, global broadcasts acting on a variety of scales of time as the organism becomes large enough to entertain a sufficient number of lower level cognitive modules.

The essential point is that this kind of phase transition to a 'global broadcast', necessarily involving changes in groupoid symmetry, is very general and will occur from the intra-organismal to societal level of organization in cognitive systems.

One suspects that evolutionary selection has, in fact, acted on cognitive network topology to favour forms that permit broadcasts – i.e., the establishment of a 'giant component' – at relatively low values of ϵ , albeit on varying timescales.

It was noted, after equation (1), that disjoint partition of state space is possible according to sets of states that can be connected by high probability paths to the same base point, leading to a groupoid structure characterizing the underlying network topology. Groupoids generalize ideas of both symmetry and equivalence class (e.g. Brown, 1987), and ϵ can be taken as a temperature analog in a spontaneous symmetry breaking (or making) argument in Landau's sense (Landau and Lifshitz, 2007). That is, increasing ϵ , a measure of crosstalk – information leakage – between cognitive submodules, leads to a punctuated transition to a higher groupoid symmetry representing a global broadcast.

Nothing in this realization of the Baars model seems restricted to terrestrial biological entities. Indeed, even on Earth, we know of widely different brain structures that instantiate conscious behaviours, ranging from familiar mammalian and avian forms to reptiles, cephalopods, and perhaps insect colonies (Cartmill, 2000; Griffin *et al.*, 2004; Edelman *et al.*, 2005).

The essential point, from an astrobiological perspective, is that any organism evolving a sufficiently large set of unconscious cognitive submodules is likely to undergo an evolutionary exaptation of the crosstalk (information leakage) between them into something having a shifting, tunable, global workspace/broadcast mechanism. Those broadcasts operating in the realm of a few hundred milliseconds would be analogous to consciousness-as-we-know-it.

The rate of cognition

We know that physiological global broadcasts, where underlying cognitive modules come together to address patterns of threat

or opportunity, operate across very different timescales. Wound healing can take 18 months, while high-order animal consciousness operates in the realm of a few hundred milliseconds. How can consciousness be so fast? Here, we adapt the arguments of (Wallace, 2016a,b).

Recall the argument above, that in the cognitive process, an internal picture of the world is compared with incoming sensory data. Such comparison, involving transmission of information along with a ‘sensory/cognitive’ channel, permits definition of an average distortion between what is happening and what is perceived, and, via the Rate Distortion Theorem (Cover and Thomas, 2006), permits estimation of the minimum channel capacity needed to keep that average distortion below some limit. That is, to survive ‘real time’ challenge, the organism must maintain a good ongoing estimate of the external real world so that cognitive choices can be made, also in real time.

What are we to do with such a biological Rate Distortion Function? The approach is via Arrhenius’ law (Laidler, 1987), which predicts an exponential rise in reaction rate with the rise in temperature. Here, however, the ‘temperature’ is the rate of supply of metabolic free energy to the cognitive system of interest.

The energetics of terrestrial biological reactions are remarkable. At 300 K, molecular energies represent something like 2.5 KJ/mol⁻¹ of available free energy. The basic terrestrial biological energy reaction – the hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate – at 300 K, produces some 50 KJ/mol⁻¹, implying a ‘reaction temperature’ of 6000 K. A high rate of ATP delivery thus provides enough energy for very rapid biocognition, given that neural tissues in higher animals consume an order of magnitude more metabolic free energy per unit time than other tissues.

A calculation

Given a simple chemical reaction of the form $aA + bB \rightarrow pP + qQ$, the rate of change in the concentration of chemical species P – written $[P]$ – is typically determined by an expression

$$d[P]/dt = k(T)[A]^n[B]^m \tag{7}$$

where n and m are constants depending on system structure. The reaction rate k is expressed by the Arrhenius relation as

$$k = \alpha \exp[-E_a/\kappa T] \tag{8}$$

where α is another characteristic constant, E_a is the reaction activation energy, T is the Kelvin temperature and κ a universal constant.

Note that $\exp[-E_a/\kappa T]$ is, using the Boltzmann distribution, the fraction of molecular interactions having energy greater than E_a .

The basic idea can be adapted to cognitive physiological processes. Above, we have associated cognition with a dual information source (Wallace, 2012). High level cognition is usually an all-or-nothing phenomenon (Sergeant and Dehaene, 2004) so that an incoming signal must exceed a threshold before becoming entrained into a characteristic general broadcast of large-scale developmental selection, very much akin to a slow form of consciousness (Wallace, 2012). Interest focuses on the Rate Distortion Function associated with the channel connecting the cognitive organism with an embedding and embodying ‘real time’ environment. Again, $R(D)$ defines the minimum rate of

information transmission needed to ensure that the average distortion between what is sent and what is received is less than or equal to $D \geq 0$, according to an appropriate distortion measure. Assuming a threshold R_0 for the perception of an incoming environmental signal, following Feynman (2000) identification of information as a form of free energy, it is possible to write a Boltzmann-like probability as

$$P[R \geq R_0] = \frac{\int_{R_0}^{\infty} \exp[-R/\omega M] dR}{\int_0^{\infty} \exp[-R/\omega M] dR} = \exp[-R_0/\omega M] \tag{9}$$

where M is the rate at which metabolic free energy is provided to the cognitive process. Figure 1, showing $\exp(-1/M)$, follows. Also shown is the energetic efficiency measure $\exp(-1/M)/M$. Peak efficiency – likely to be strongly selected for by evolutionary process – takes place at $R_0/\omega M = 1$.

In sum, cognition rate first increases exponentially sharply with the rate of free energy supply M , but rapidly tops out, so that highest efficiency is at an intermediate value of M .

For terrestrial higher animals, nervous system tissues, operating with time constants of about 100 ms, do so by requiring an order of magnitude greater metabolic energy support than other tissues. High-level cognition in other astrobiologies may, in fact, operate with time constants shorter than 100 ms if they can evolve to supply metabolic free energy at sufficiently high rates.

A very general statistical model of the punctuated accession-to-consciousness involving the sudden formation of the global workspace that takes place when an incoming ‘sensory’ signal reaches threshold is given in the Mathematical Appendix.

Are ‘sleep cycles’ inevitable in rapid biocognition?

We have now developed enough formal machinery to address the question raised in the title, adapting, somewhat, the arguments of Wallace (2016a).

As Maturana and Varela (1980) argue, the cognition is ubiquitous at all scales and levels of organization of the living state. Gene expression mechanisms at the cellular level must choose when and how to activate, and are closely regulated as to time, place, and manner. Failure of regulation results in developmental disorders of varying severity. Blood pressure must not only change to meet metabolic and other demands, but it must be closely regulated to prevent explosive organic damage. Similarly, in higher animals, the ‘stream of consciousness’ must be closely regulated by ‘riverbanks’ that constrain it to realms useful to the animal and its social groupings (Baars 1988; Wallace, 2005, 2012). We must, then, envision consciousness and its regulatory system as a paired dyad under a constraint on the maximum rate at which metabolic free energy is available.

Letting the subscript C represent consciousness and R its regulatory system, we thus seek to maximize an efficiency scalarization

$$\frac{\exp[-K_C/M_C]}{M_C} + \frac{\exp[-K_R/M_R]}{M_R} \tag{10}$$

subject to the constraint

$$M = M_C + M_R \tag{11}$$

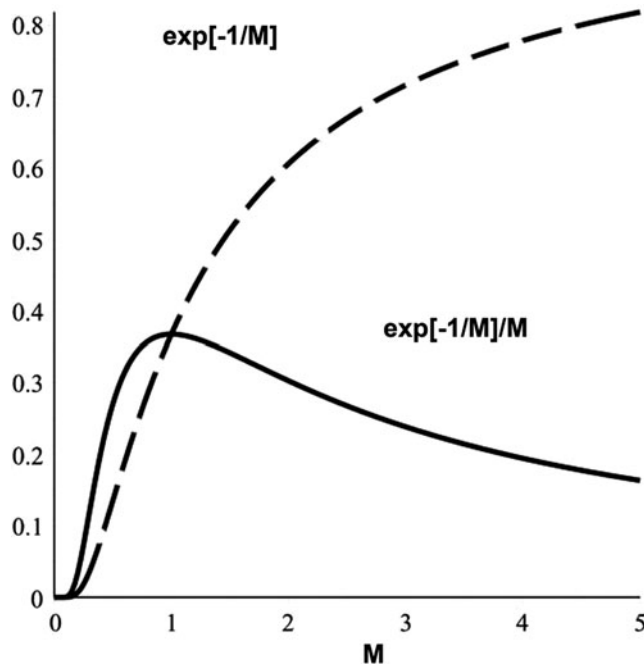


Fig. 1. ‘Arrhenius’ relation for the rate of a cognitive physiological process as a function of the rate of metabolic free energy consumption. The rate of cognition – dotted line – at first increases exponentially with supply of metabolic free energy but then tops out. The highest efficiency – solid line – occurs at the intermediate value $M = R_0/\omega$.

This leads, most simply, to a standard Lagrangian optimization on

$$L \equiv \frac{\exp[-K_C/M_C]}{M_C} + \frac{\exp[-K_R/M_R]}{M_R} - \lambda(M - M_C - M_R) \quad (12)$$

where λ represents a ‘cost’ variable dual to M . Under sufficiently pathological conditions, the ‘cost’ cannot be met (Jin *et al.*, 2008).

The Lagrange optimization condition is

$$\nabla_{M_C, M_R, \lambda} L = 0 \quad (13)$$

The resulting equation for solution pairs M_C, M_R implies the existence of several different possible optimization points for the system, strongly parameterized by the overall energy rate M . This expression represents an index theorem relating solutions of an analytic equation to underlying topological modes: increasing the number of systems to be optimized – adding terms of the form $\exp[-K_j/M_j]/M_j$ – increases the number of possible non-equilibrium quasi-stable states.

Making the first approximation to a sleep cycle, there will be only a small number of stages to the normal pattern. On Earth, these are NREM (non-rapid eye movement) sleep, which involves low rates of blood flow to the brain – low M – and REM sleep which can rival or exceed conscious state blood flows. Sleep states must, like all other neural processes, be highly regulated.

Taking the rate constants $K_C \approx K_R = K$ in equation (10), calculation shows a symmetric efficiency curve with one or two peaks, depending on the magnitude of $M = M_C + M_R$, as in Fig. 2. Taking $K = 1, M_C + M_R = M = 1 \rightarrow 3.6$, we obtain a maximum efficiency index, (about 0.74) at $M_C = M_R \approx 1$ for $M \approx 2$, or along symmetric ridge points for larger M . The maximum value of 3.6 for M is

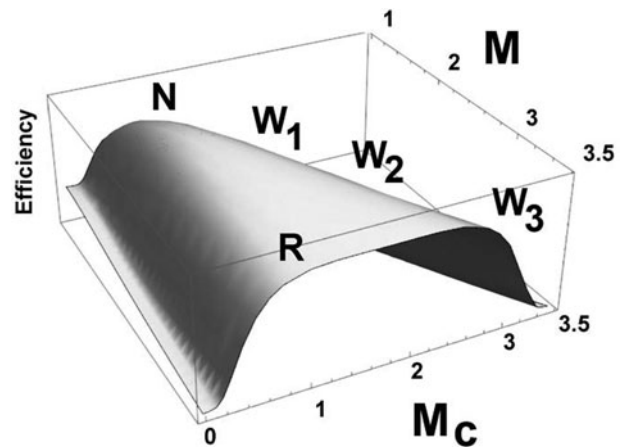


Fig. 2. Following Wallace (2016a), consciousness-and-regulator efficiency as a function of the metabolic free energy consumed by consciousness, M_C , for $K_x = 1, M = M_C + M_R = 1 \rightarrow 3.6$. A single maximum dominates the low energy mode, suggesting NREM sleep, denoted N , representing a 44% decline from the maximum M . The maxima at higher energies are ridges, suggesting REM sleep/awake modes, denoted R, W . Normal transitions are from W to N , followed by N - R oscillation. During sleep modes, ‘regulator’ systems for consciousness may perform routine maintenance, as does the immune system when not coping with infection, wound healing, or malignancy. NREM sleep then becomes a transition path between the two more active states. Changing the values of the K_x markedly shifts the relative heights and widths of the peaks, suggesting an added control mode. Some work produces the sorts of models described in Booth and Diniz Behn (2014). The progression $W_1 \rightarrow W_2 \rightarrow W_3$ would indicate increasing degrees of wakeful attention, requiring sequentially greater supplies of metabolic free energy. More convoluted evolutionary (or cultural) histories might impose other intermediate points for cycles on such energy/efficiency surfaces.

chosen from Madsen and Vorstrup (1991), who found a 44% decline in cerebral metabolic rate during deep, slow-wave, NREM sleep.

The highest efficiency peak is at the quiescent NREM-analog mode. At higher M there are two symmetric greatest efficiency points that suggest waking/active sleep modes in which REM ‘sleep’ seems to represent a parsimonious assumption of essential maintenance duties by systems otherwise dedicated to the regulation of awake consciousness. This appears to be similar to the immune system that, when not coping with infection, wound healing and malignancy, is committed to processes of routine cellular maintenance (Cohen, 2000).

Sleep deprivation may be very serious indeed.

As Harris *et al.* (2012) put it,

Increasingly, sleep is thought to play an energetically restorative role in the brain ... [D]uring sleep there is a transient increase in ATP level in cells of awake-active regions of the brain ... suggested to fuel restorative biosynthetic processes in cells that, during the day, must use all their energy on electrical and chemical signalling. This implies an energy consumption trade-off; a high use of ATP on synapses during awake periods is balanced by more ATP being allocated to other tasks during sleep.

Energy use in the awake state also increases due to synaptic potentiation ... compared to sleep...

These changes are reversed during sleep, presumably because of homeostatic plasticity ... Thus sleep is essential for adjusting synaptic energy use.

It is known that the normal route from waking, the W point in Fig. 2, to REM sleep, the point R , is from W to N , and then down the other ridge from N to R . Direct transition from R to W does

occur, and cycling between N and R is normal. In addition, occasional, culturally-dependent ‘trance states’ are known to exist that seem intermediate (e.g. Hubbard, 2003).

The progression $W_1 \rightarrow W_2 \rightarrow W_3$ in Fig. 2 could be taken as representing different degrees of attention within the waking state, each requiring a higher overall rate of metabolic free energy supply.

Note that higher values of $K_C = K_R$ generate much broader symmetric curves with far less well-defined peaks, while lower values generate surfaces that are much more sharply peaked. Unequal values raise or lower one or the other ridge. Thus ‘tuning’ these parameters individually would provide significant added system control, causing a transition from ‘waking’ to various forms of ‘sleep’ states.

This model is consistent with Hobson’s (Hobson (2009), figure 4) picture of the Western sleep cycle, where the normal progression is from waking to low-energy NREM sleep, followed by oscillation between NREM mode and the high-energy REM state. Note that, in other cultures, other patterns have been observed (e.g. Wallace, 2017 Section 5.10 and references therein).

Figure 2 is also consistent with the observations of Hudson *et al.* (2014) on the punctuated recovery to consciousness from anesthesia. See their figure 4C which identifies three hubs between which there is directed transition during recovery.

REM sleep is now understood as another state of consciousness that, compared to alert waking, is deficient in neuromuscular function, analytic ability, and transient episodic memory, while particularly rich in emotional cognition. Both waking and falling asleep involve complicated physiological processes to effect a transition between states. For some considerable time, researchers recognized the importance of the reticular activating mechanism in the change from sleep to waking (Evans, 2003). Recently, an arousal inhibitory mechanism – a thalamo-cortical process – has been recognized which transfers the body from waking to sleep (Evans, 2003). Other work (Saper *et al.*, 2005; Lu *et al.*, 2006) has identified a neural flip-flop switch structure which controls REM sleep.

The standard phenomenological model for sleep-as-we-know-it is the ‘two-process’ approach of Borbely (1982) and Daan *et al.* (1984). As Skeldon *et al.* (2014) put it, the homeostatic process takes the form of a relaxation oscillator that results in a monotonically increasing ‘sleep pressure’ during the wake that is dissipated during sleep. Switching from wake to sleep, and vice versa occurs at upper and lower threshold values of the sleep pressure respectively, where the thresholds are modulated by an approximately sinusoidal circadian oscillator. Skeldon *et al.* (2014) go on to show that the two-process model is essentially the same as the slow-time dynamic of the neurologically-based Phillips and Robinson (2007) model.

We argue, then, that under Darwinian selection pressure for high rates of waking-state consciousness, patterns of transition analogous to Fig. 2 must be seen. These might perhaps involve adding one or more efficiency terms to Eq. (10), or more intermediate transition points on the plateau, generalizing the $N - R - W$ and/or $W_1 \rightarrow W_2 \rightarrow W_3$ progressions.

Intelligent aliens likely dream, although perhaps not in the ways we do, or of what we would call sheep.

Indeed, as Worthman and Melby (2002) put the matter for humans,

...[P]hysical, social, and temporal factors generating variation in human sleep ecology ... may be paralleled by variation not only in sleep behaviour but also in its physiology...

...[A]ssociations of cultural ecologies of sleep to such ‘basic’ physiological regulatory systems as sleep biology, chronobiology, state regulation and emotion regulation would imply that these systems are partially influenced or organized through cultural ecologies operating developmentally and across the life-span.

In many respects, cultural differences already make us ‘aliens’ among ourselves. The intransigent difficulties of human inter-cultural communication should stand as a singular warning for scenarios involving ‘alien’ contacts. Recall that genetic differences between human populations are far less than those between chimpanzee populations.

We briefly explore this question further.

Consciousness and culture intertwined

The evolutionary anthropologist Robert Boyd has asserted that ‘Culture is as much a part of human biology as the enamel in our teeth,’ (Richerson and Boyd, 2004) and, while many other animals on Earth display some measure of culture as learned and transmitted behaviour (Avital and Jablonka, 2000), nothing defines humans quite like the interpenetration of mind and self with the embedding cultural milieu. Technology and its artefacts are, of course, one part of that milieu.

It is not difficult to extend the Baars model to include interaction with an embedding culture and with a hierarchical set of institutions within that culture seen as a generalized transmissible language associated with a nested set of information sources (Wallace, 2013, 2015). This includes both a form of niche construction (Wallace, 2010) and distributed cognitive institutions acting on various scales (Wallace and Fullilove 2008). This is most easily done by invoking the set of interacting information sources dual to cognitive process via network information theory (El Gamal and Kim, 2010, p. 2–26): Given a basic set of such dual information sources, say (X_1, \dots, X_k) , that can be partitioned into two ordered sets, say $X(\mathcal{J})$ and $X(\mathcal{J}')$, then the splitting criterion of the larger system becomes $H(X(\mathcal{J})|X(\mathcal{J}'))$. Generalization to three or more such ordered sets seems direct and leads to a Baars-like theory of collective consciousness in which different global workspaces act at different scales of size and time (Wallace and Fullilove, 2008; Wallace, 2015).

As the anthropologists will attest, an astounding variety of culturally-driven institutions, associated forms of mind and self, and dynamics of interaction, graces the world. Typically, humans, whose overall genetic structure is more uniform than that of chimpanzee populations, do not communicate well across the many different cultural modes. Wallace (2015) and Wallace and Fullilove (2008) suggest that stabilizing complex systems of interacting cognitive institutions is exceedingly difficult, given that canonical inability to communicate, and the planet seems to be facing a serious crisis of sustainability.

More relevant, perhaps, are recent studies of the role of culture in ‘basic biological’ aspects of human perception, matters that ‘should’ be independent of culture.

They are not. We loosely follow the arguments of Wallace (2018b).

Most particularly, Nisbett *et al.* (2001), continuing a long line of research (Markus and Kitayama, 1991; Heine, 2001), review an extensive literature on empirical studies of basic cognitive differences between individuals raised in East Asian and Western cultural heritages, which they characterize, respectively, as ‘holistic’ and ‘analytic’, finding that tools of thought embody a culture’s

intellectual history, that tools have theories built into them, and that users accept these theories, albeit unknowingly, when they use these tools.

Matsuda and Nisbett (2006) observed that, whereas East Asians view the world holistically, attending to the entire field and relations among objects, Westerners view the world analytically, focusing on the attributes of salient objects. Compared to Americans, East Asians were more sensitive to contextual changes than to focal object changes. These results strongly imply significant cultural variation in what may seem to be basic perceptual processes.

Similarly, Nisbett and Miyamoto (2005) argue that fundamental perceptual processes are influenced by culture, establishing a dynamic relationship between the cultural context and perceptual processes. The central implication is that for humans, who have relatively little genetic variability across populations, basic perception can no longer be regarded as consisting of processes that are universal across all people at all times.

Wallace (2007) explores analogous dynamics involving inattentional blindness and culture.

The inference, then, is that ‘alien’ species will have cognitive modalities also synergistic and intertwined with learned or inherited culture and that understanding and relating to such mutualistic constructs will be difficult indeed.

Quantum systems

The theory above is purely classical and produces Baars’ results directly. Wallace (Wallace, 2005, Chapter 5, Section 6), however, does examine how quantum versions of the asymptotic limit theorems of information theory (e.g. Bjelakovic *et al.*, 2003, 2004) might be used to generalize the model. Unfortunately, the quantum results are not well characterized, and an exact treatment is lacking. Nonetheless, it becomes quite clear that consciousness in quantum systems – at least those supporting relatively large coherence lengths – would be to consciousness-as-we-know-it much as a flask of superfluid helium is to a glass of water.

Tegmark (2000), in contrast to his later work, has convincingly argued the impossibility of macroscopic quantum treatments of consciousness at normal biological temperatures. Quantum effects at such temperatures are, of course, fully and strongly manifest at the molecular and neural levels as ‘chemistry’.

Since information is a form of free energy, even quantum systems having large coherence lengths will suffer second law heating through information transmission and transformation that will inherently limit the possible size of quantum-conscious structures. Typically 10^9 – 10^{11} interacting components are needed for high level mental function, involving large-scale information transfer. This scale of activity is likely to generate much heat and unlikely to be attained in quantum domains by an evolutionary process in the natural world, even if far smaller numbers of qubits are envisioned as sufficient for rapid lower level cognition.

The inference is, then, that the designs for ‘conscious quantum systems’ are likely to remain in the realm of perpetual motion machines of the second kind for some time. Indeed, proposals of this nature rapidly enter the notorious physics realm of ‘not even wrong’ (e.g. Penrose, 1994; Tegmark, 2015).

Conclusions

Life in the cosmos seems likely to be ubiquitous. Organisms that must react on timescales of a few hundred milliseconds should

host many ‘neural-like’ structures that, in the presence of sufficient crosstalk fuelled by a sufficient rate of metabolic (or other) free energy, provide an evolutionary process with the basic material to produce adaptive tunable/shifting global workspace/broadcast/phase transition phenomena that become fixed in reproduction and that we would likely recognize as conscious. Such systems, we argue, must also be highly regulated, leading to evolutionarily selected modes in which the regulatory machinery will have been given ‘off duty’ tasks of maintenance that will likely require some recognizable form of ‘sleep cycle’.

A significant number of alien organisms, or their social or mutualistic composites, will, over sufficient time, become synergistic with learned, transmissible, language-like patterns of adaptation analogous to a culture that include collective structures of various forms acting under distributed cognition and that are capable of large-scale cooperative activity. Assuming some few of these creatures able to stabilize the resulting systems of collective consciousness, the emergence of high technology seems likely, although communication with them would probably be limited to exchanges of Balmer series symbols and schematics for amino acids: recall the notable and continuing difficulties of human inter-cultural communication, beyond the most basic level of automatic language translation.

In summary, a cellular/molecular astrobiology and, with reservations, a limited kind of astropsychology, appear relatively straightforward. Cultural astroethology – understanding alien systems of learned heritage transmission and their role in astrobiological processes – by contrast, would be a profound intellectual challenge, far more daunting than the still-unmet difficulties of understanding different cultures and their shifting norms within human populations.

Mathematical appendix

Activation of the rate-of-cognition expression from Fig. 1, $F = \exp[-K/M]$, first requires recognition that an incoming ‘sensory’ signal – in a large sense – deserves response. For working coalitions of low-level cognitive modules, we have argued that there is a topologically-defined threshold for such detection, depending on the wiring of the crosstalk. Most generally, we can write a stochastic differential equation of the form

$$\begin{aligned} dF_t &= (dF/dM \times dM/dt)dt + F_t dB_t \\ &= F_t \left[\left(\frac{K}{M_t^2} \dot{M}_t \right) dt + dB_t \right] \\ &= F_t dY_t \end{aligned} \quad (14)$$

where B_t is a ‘noise’ that may not be simple Brownian, and Y_t is a very general stochastic process.

Equation (14) is in precisely the form required by the Doleans-Dade exponential (Protter, 2005), having an expectation in F that converges in probability as

$$\mathcal{E}(F) \rightarrow \infty \exp \left[Y_t - \frac{1}{2} [Y_t, Y_t] \right] \quad (15)$$

where $[Y_t, Y_t] \geq 0$ is the quadratic variation of the stochastic process Y_t (Protter, 2005). Heuristically, by the Mean Value Theorem, if $(1/2)d[Y_t, Y_t]/dt > dY_t/dt$, then $\mathcal{E}(F) \rightarrow 0$ in probability, and the ‘switch’ is not activated.

Thus $[Y_t, Y_t]$ can be taken as the noise-to-signal ratio of the incoming 'sensory' information, tuned according to the shifting topological configuration of the underlying network of 'low level' cognitive modules. For Brownian noise, $d[Y_t, Y_t]/dt \equiv \sigma^2$, and the noise-to-signal ratio has clear meaning. For 'colored' noises more characteristic of real neural processes, matters are more complicated but still, fall within the treatment.

The basic result can be extended to more general cognition rate expressions than $F = \exp[-K/M]$ as follows.

We have, from equations (9) and (10), leading to the expressions of equations (14) and (15), imposed a particular functional form for the effectiveness, efficiency, and the stochastic dynamics, of a tunable, cognitive 'global broadcast' network system responding to sensory challenge. More comprehensive results are available (Appleby *et al.*, 2008).

We suppose that, instead of the specific form leading to equations (14) and (15), we can express the stochastic dynamics of a generalized measure $\mathcal{F}(M)$ as

$$d\mathcal{F}_t = g(\mathcal{F}_t)dt + h(\mathcal{F}_t)dW_t \quad (16)$$

where dW_t is taken as Brownian noise, subject to the Lipschitz condition that, for a positive constant C , and for all $\mathcal{F}(M_2)$ sufficiently near to some fixed point $\mathcal{F}(M_1)$,

$$|g(\mathcal{F}(M_1)) - g(\mathcal{F}(M_2))| \leq C|\mathcal{F}(M_1) - \mathcal{F}(M_2)| \quad (17)$$

We further assume that $\mathcal{F}(0) = 0$, 'no excitation of the global broadcast' as above, is an 'unstable equilibrium state'.

Then, Appleby *et al.* (2008) show that, in one dimension, a function h can always be found that stabilizes an unstable equilibrium, i.e., keeps $\mathcal{F}(M, t)$ at zero. This generalizes the Doleans-Dade exponential of Eq. (15) to a much broader class of functions. On the one hand, sufficient 'noise-to-signal' ratio, measured by the function h , can keep a global broadcast from occurring, on the other, the function h itself is a tunable means of ensuring that, when incoming sensory signal is sufficiently large under the conditions defined by h , such a global broadcast can indeed take place.

References

- Appleby J, Mao X and Rodkina A (2008) Stabilization and destabilization of nonlinear differential equations by noise. *IEEE Transactions on Automatic Control* **53**, 126–132.
- Ash R (1990) *Information Theory*. New York: Dover Publications.
- Atlan H and Cohen I (1998) Immune information, self-organization, and meaning. *International Immunology* **10**, 711–717.
- Avital E and Jablonka E (2000) *Animal Traditions: Behavioral Inheritance in Evolution*. New York: Cambridge University Press.
- Baars B (1988) *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.
- Baars B (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in Brain Research* **150**, 45–53.
- Bjelakovic I, Kruger T, Siegmund-Schultz R and Szkola A (2003) Chained typical subspaces – a quantum version of Brieman's theorem, ArXiv, quant-ph/0301177.
- Bjelakovic I, Kruger T, Siegmund-Schultze R and Szkola A (2004) The Shannon-McMillan theorem for ergodic quantum lattice systems. *Inventiones Mathematicae* **155**, 203–222.
- Booth V and Diniz Behn C (2014) Physiologically-based modeling of sleep-wake regulatory networks. *Mathematical Biosciences* **250**, 54–68.
- Borbely A (1982) A two process model of sleep regulation. *Human Neurobiology* **1**, 195–204.
- Brown R (1987) From groups to groupoids: a brief survey. *Bulletin of the London Mathematical Society* **19**, 113–134.
- Cartmill M (2000) Animal consciousness: some philosophical, methodological, and evolutionary problems. *American Zoologist* **40**, 835–846.
- Cohen I (2000) *Tending Adam's Garden: Evolving the cognitive immune self*. New York: Academic Press.
- Cover T and Thomas J (2006) *Elements of Information Theory*, 2nd Edn. New York: John Wiley and Sons.
- Daan S, Beersma D and Borbely A (1984) Timing of human sleep: recovery process gated by a circadian pacemaker. *American Journal of Physiology* **246**, 161–178.
- Dehaene S and Naccache L (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* **79**, 1–37.
- Dretske F (1994) The explanatory role of information. *Philosophical Transactions of the Royal Society A* **349**, 59–70.
- Edelman D, Baars B and Seth A (2005) Identifying hallmarks of consciousness in non-mammalian species. *Consciousness and Cognition* **14**, 169–187.
- El Gamal A and Kim Y (2010) Lecture notes on network information theory. arXiv:1001.3404v4 [cs.IT].
- Evans B (2003) Sleep, consciousness and the spontaneous and evoked electrical activity in the brain. Is there a cortical integrating mechanism? *Neurophysiologie Clinique* **33**, 1–10.
- Feynman R (2000) *Lectures on Computation*. New York: Westview Press.
- Griffin D and Speck G (2004) New evidence of animal consciousness. *Animal Cognition* **7**, 5–18.
- Harris J, Jolivet R and Attwell D (2012) Synaptic energy use and supply. *Neuron* **75**, 762–777.
- Heine S (2001) Self as cultural product: an examination of East Asian and North American selves. *Journal of Personality* **69**, 881–906.
- Hobson JA (2009) REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews Neuroscience* **10**, 803–813.
- Hubbard T (2003) Some Correspondences and Similarities of Shamanism and Cognitive Science: Interconnectedness, Extension of Meaning, and Attribution of Mental States. *Anthropology of Consciousness* **14**, 26–45.
- Hudson A, Calderon D, Pfaff D and Proekt A (2014) Recovery of consciousness is mediated by a network of discrete metastable activity states. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9283–9288.
- Jin H, Hu Z and Zhou X (2008) A convex stochastic optimization problem arising from portfolio selection. *Mathematical Finance* **18**, 171–183.
- Khinchin A (1957) *Mathematical Foundations of Information Theory*. New York: Dover.
- Laidler K (1987) *Chemical Kinetics*, 3rd Edn. New York: Harper and Row.
- Landau L and Lifshitz E (2007) *Statistical Physics, Part I*. New York: Elsevier.
- Lu J, Sherman D and Saper C (2006) A putative flip-flop switch for control of REM sleep. *Nature* **441**, 589–594.
- Madsen P and Vorstrup S (1991) Cerebral blood flow and metabolism during sleep. *Cerebrovascular Brain Metabolism Review* **3**, 281–296.
- Markus H and Kitayama S (1991) Culture and self-implications for cognition, emotion, and motivation. *Psychological Reviews* **98**, 224–253.
- Matsuda T and Nisbett R (2006) Culture and change blindness. *Cognitive Science* **30**, 381–399.
- Maturana H and Varela F (1980) *Autopoiesis and Cognition*. Dordrecht: Riedel.
- Nisbett R, Peng K, Incheol C and Norenzayan A (2001) Culture and the system of thought: holistic versus analytic cognition. *Psychological Review* **108**, 291–310.
- Nisbett R and Miyamoto Y (2005) The influence of culture: holistic versus analytic perception. *Trends in Cognitive Science* **10**, 467–473.
- Penrose R (1994) *Shadows of the Mind: A Search for the Missing Science of Consciousness*. New York: Oxford University Press.
- Pettini M (2007) *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*. New York: Springer.
- Phillips A and Robinson P (2007) A quantitative model of sleep-wake dynamics based on the physiology of the brainstem ascending arousal system. *Journal of Biological Rhythms* **22**, 167–179.
- Protter P (2005) *Stochastic Integration and Differential Equations*, 2nd Edn. New York: Springer.

- Richerson P and Boyd R** (2004) *Not by Genes Alone; How Culture Transformed Human Evolution*. Chicago, IL: Chicago University Press.
- Saper C, Cano G and Scammell T** (2005) Homeostatic, circadian, and emotional regulation of sleep. *Journal of Comparative Neurology* **493**, 92–98.
- Sergeant C and Dehaene S** (2004) Is consciousness a gradual phenomenon? Evidence for an all-or-nothing bifurcation during the attentional blink. *Psychological Science* **15**, 720–728.
- Skeldon A, Dijk D and Derks G** (2014) Mathematical models for sleep-wake dynamics: comparison of the two-process and a mutual inhibition neuronal model *PlosOne* **9**, e103877.
- Spenser J** (2010) The giant component: the golden anniversary. *Notices of the American Mathematical Society* **57**, 720–724.
- Tegmark M** (2000) Importance of quantum decoherence in brain processes. *Physical Review E* **61**, 4194–4206.
- Tegmark M** (2015) Consciousness as a state of matter. *Chaos, Solitons, and Fractals* **76**, 238–270.
- Tononi G** (2012) *Phi: A voyage from the Brain to the Soul*. New York: Knopf.
- Wallace R** (2000) Language and coherent neural amplification in hierarchical systems: renormalization and the dual information source of a generalized spatiotemporal stochastic resonance. *International Journal of Bifurcation and Chaos* **10**, 493–502.
- Wallace R** (2005) *Consciousness: A Mathematical Treatment of the Global Neuronal Workspace Model*. New York: Springer.
- Wallace R** (2007) Culture and inattention blindness. *Journal of Theoretical Biology* **245**, 378–390.
- Wallace R and Fullilove M** (2008) *Collective Consciousness and its Discontents*. New York: Springer.
- Wallace RG and Wallace R** (2009) Evolutionary radiation and the spectrum of consciousness. *Consciousness and Cognition* **18**, 160–167.
- Wallace R** (2010) Expanding the modern synthesis. *Comptes Rendus Biologies* **333**, 701–709.
- Wallace R** (2012) Consciousness, crosstalk, and the mereological fallacy: an evolutionary perspective. *Physics of Life Reviews* **9**, 426–453.
- Wallace R** (2013) A new formal approach to evolutionary processes in socio-economic systems. *Journal of Evolutionary Economics* **23**, 1–15.
- Wallace R** (2015) Closed-system ‘economic’ models for psychiatric disorders: Western atomism and its culture-bound syndromes. *Cognitive Processing* **16**, 279–290.
- Wallace R** (2016a) High metabolic demand in neural tissues: information and control theory perspectives on the synergism between rate and stability. *Journal of Theoretical Biology* **409**, 86–96.
- Wallace R** (2016b) The metabolic economics of environmental adaptation. *Ecological Modelling* **322**, 48–53.
- Wallace R** (2017) *Computational Psychiatry: A systems approach to the epigenetics of mental disorders*. New York: Springer.
- Wallace R** (2018a) New statistical models of nonergodic cognitive systems and their pathologies. *Journal of Theoretical Biology* **436**, 72–78.
- Wallace R** (2018b) Culture and the trajectories of developmental pathology: insights from control and information theories. *Acta Biotheoretica* **66**, 79–112.
- Worthman C and Melby M** (2002) Toward a comparative developmental ecology of human sleep. In Carskadon M (ed.), *Adolescent Sleep Patterns: Biological, Social, and Psychological Influences*. New York: Cambridge University Press, pp. 69–117.