



**HAL**  
open science

## On 'Machine Consciousness'

Rodrick Wallace

► **To cite this version:**

| Rodrick Wallace. On 'Machine Consciousness'. 2023. hal-03946881

**HAL Id: hal-03946881**

**<https://hal.science/hal-03946881>**

Preprint submitted on 19 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Draft  
Draft

## On ‘Machine Consciousness’

Rodrick Wallace

*Division of Epidemiology*

*The New York State Psychiatric Institute, Box 47*

*1051 Riverside Drive*

*New York, NY 10032 USA*

*rodrick.wallace@gmail.com, rodrick.wallace@nyspi.columbia.edu*

Consciousness in higher animals, by virtue of its 100 millisecond time constant, is a necessarily greatly simplified and stripped-down version of more complex multiple tunable workspace cognition/regulation dyads like wound healing, immune function, gene expression, institutional function and the like. These more complex dynamic entities emerged through evolutionary exaptation of the inevitable information crosstalk between coexistent cognitive modules. In consequence of the debrided nature of consciousness, it should not be difficult to construct a fast, single workspace ‘conscious machine’ that mimics the human tunable neuronal global workspace system. Tied to a ‘backbrain’ AI that has learned hyperrapid stereotypic pattern responses to some particular set of likely challenges, the result is an elementary ‘emotional’ conscious machine. A clever designer, however, may want to use available high-speed electronics to mimic the more capable multiple-workspace/workforce systems inherently less susceptible to inattentive blindness and related failings of overfocus and thrashing. Contrary to current social constructions, however, the ultimate utility of such machines remains obscure. Here, we explore these matters in formal detail, restricting argument to the asymptotic limit theorems of information and control theories.

*Keywords:* consciousness; control theory; crosstalk; groupoid; information theory; phase transition; symmetry-breaking

***Horses for courses.***

— British aphorism.

***Sure, hogs are smart, but I still like bacon and eggs.***

— American aphorism.

### 1. Introduction

The engineering discipline of ‘machine consciousness’, in a large sense, remains mired in longstanding – and essentially theological – debates about the ‘nature’ of human consciousness, socially constructed as a deep and intractable scientific and philosophical problem (e.g., Chella et al. 2019). Indeed, rapid empirical progress in the study of consciousness over the late 19th and early 20th centuries was short circuited for many decades by ideological imperatives, until the intellectual logjam

2 *Rodrick Wallace*

was recently broken by Bernard Baars and collaborators (Baars 1988, 2005; Barrs and Franklin 2003; Baars and Gage 2010) leading to the contemporary flowering of debate, to employ something of a euphemism.

Engineering applications seem tantalizing because, over their 500myr course, conscious organisms, as sculpted by evolutionary process, appear – at first glance – to dominate the biological hierarchy, although most conscious species have, in fact, gone extinct over the history of the planet. Conscious machines, it is nonetheless posited, will be of both intellectual and practical interest.

This, of course, remains to be seen.

We begin by focusing on what Adams (2003) has called ‘the informational turn in philosophy’ – explicit application of communication theory formalism and concepts to “purposive behavior, learning, pattern recognition, and... the naturalization of mind and meaning”. We will, at a later stage, retread similar ground more explicitly from the ‘tunable global workspace’ perspective of Baars.

One of the first comprehensive attempts at an ‘informational turn’ was that of Dretske (1981, 1988, 1994), whose work Adams describes as follows:

It is not uncommon to think that information is a commodity generated by things with minds. Let’s say that a naturalized account puts matters the other way around, viz. it says that minds are things that come into being by purely natural causal means of exploiting the information in their environments. This is the approach of Dretske as he tried consciously to unite the cognitive sciences around the well-understood mathematical theory of communication...

Dretske himself (1994) writes:

Communication theory can be interpreted as telling one something important about the conditions that are needed for the transmission of information as ordinarily understood, about what it takes for the transmission of semantic information. This has tempted people... to exploit [information theory] in semantic and cognitive studies, and thus in the philosophy of mind.

...Unless there is a statistically reliable channel of communication between [a source and a receiver]... no signal can carry semantic information... [thus] the channel over which the [semantic] signal arrives [must satisfy] the appropriate statistical constraints of communication theory.

The asymptotic limit theorems of information theory provide, then, necessary conditions for understanding all cognitive process, including consciousness. These are the Coding, Source Coding, and Rate Distortion Theorems (Khinchin 1957; Cover and Thomas 2006).

Embodiment – an essential matter for robotics – from this perspective, is then instantiated by another – recently-uncovered – asymptotic limit theorem, the Data Rate Theorem, that links control and information theories (e.g., Nair et al. 2007).

- The Coding Theorem posits that, for information transmission along a noisy channel, there is a ‘channel capacity’ such that, if a properly coded message is sent at a rate less than or equal to that value, a sufficiently long signal will be received with an arbitrarily small error (Cover and Thomas 2006).

- The Source Coding Theorem (Khinchin 1957) states that messages sent by an information source – having an inherent grammar and syntax – can be broken into two sets. The first, having a large size, but with vanishingly small probability, is not consonant with the grammar and syntax. The second, much smaller, is consonant, and has – in sum – probability approaching certainty. If the source is stationary, so that probabilities do not change in time, and ergodic, so that cross-sectional measures converge on time series measures, then a ‘Source Uncertainty’ can be assigned that follows a Shannon Entropy form. We will be interested in nonergodic stationary systems for which source uncertainties vary with messages.

Although nonergodic information sources can be approximated as a kind of ‘sum’ of ergodic ones, for our purposes this is like the Ptolemaic treatment of planetary motion. We are seeking Keplerian laws.

- The Rate Distortion Theorem, a kind of inverse of the Coding Theorem, states that, under noisy circumstance, for some scalar measure of distortion  $D$  between what is sent and what is received, there is a minimum channel capacity  $R(D)$  such that, if the capacity is greater or equal to minimum, the distortion will be less than or equal to  $D$ .

- The Data Rate Theorem links control and information theories. It states that, for an inherently unstable control system, if control information is sent at a rate greater than the rate at which the system itself generates ‘topological information’, then control will not fail. Think of driving a vehicle at some specific speed on a twisting, pot-holed roadway. The driver must provide control signals – steering, braking, etc. – at a rate greater than the road twists, turns, and bottoms out the vehicle. See Nair et al. (2007) for the standard derivation.

Under a ‘simple’ assumption that information is a form of free energy (Feynman 2000; Wallace 2022 Sec. 1.4), subject to an Onsager-like nonequilibrium thermodynamics formalism (de Groot and Mazur 1984), the Data Rate Theorem can be derived from the Rate Distortion Theorem, as outlined in the Mathematical Appendix.

The key element in the ‘dynamic’ argument follows Atlan and Cohen’s (1998) treatment of the immune system as a cognitive agent, and not simply a ‘reflex’. Cognition implies choice, and choice reduces uncertainty, implying existence of an information source ‘dual’ to the cognitive process under study. The argument is direct and unambiguous, if not entirely elementary.

4 *Rodrick Wallace*

Indeed, suppose we have  $n$  possible choices, equally probable. Then, for a stationary ergodic system, the Shannon uncertainty is

$$H = \sum_{i=1}^n -\frac{1}{n} \log(1/n) = \log(n) \quad (1.1)$$

If a single element is subsequently chosen, then  $n = 1$  and the (classic) Shannon uncertainty declines from  $\log(n)$  to  $\log(1) = 0$ .

Cognitive systems must constantly make choices under dynamic ‘selection pressures’.

How is all this to be formalized?

## 2. Toward a model

The first step is to recognize that cognitive systems – biological or mechanical – are embedded in an environment that includes themselves. There will be (at least) three resource streams necessary for successful function. The first is the rate at which elements of the system can communicate with each other, instantiated by some channel capacity  $\mathcal{C}$ . The second is the rate at which ‘sensory information’ is available from the embedding circumstance, say according to a channel capacity  $\mathcal{H}$ . Finally, there is the rate at which ‘material/materiel’ resources can be provided, a rate  $\mathcal{M}$ . For an organism, this might be measured by the rate at which metabolic free energy is provided. Machine/institutional systems will have a different measure.

In general, these three rates will ‘crosscorrelate’ in a sense, so that one is actually confronting a 3 by 3 (or greater) matrix  $\mathbf{Z}$ . An  $n \times n$  matrix will have  $n$  scalar invariants determined by the polynomial relation

$$p(\gamma) = \det[\mathbf{Z} - \gamma\mathbf{I}] = (-1)^n \gamma^n + (-1)^{n-1} r_1 \gamma^{n-1} + \dots - r_{n-1} \gamma + r_n \quad (2.1)$$

$\mathbf{I}$  is the  $n \times n$  identity matrix,  $\det$  the determinant, and  $\gamma$  a real-valued parameter. The first invariant is the matrix trace, and the last the determinant.

The most direct index might be  $Z = \mathcal{C} \times \mathcal{H} \times \mathcal{M}$ , but the real world is not likely to be so simple. In any event, we now postulate a scalar index  $Z$  representing information and material resource rates. Wallace (2021) outlines generalization to more than one such index.

We must move beyond the ergodic restriction on information sources (Cover and Thomas 2006). Only when cross-sectional and longitudinal means are the same can information source uncertainty be expressed as a conventional Shannon ‘entropy’ (Khinchin 1957). We specify that source uncertainties do converge for sufficiently long paths, not that they fit a particular functional form, i.e., that of ‘Shannon entropy’. It is the values themselves of those uncertainties that will be of concern, not their particular functional expressions in terms of probabilities. Thus we study what might be called Adiabatically Piecewise Stationary (APS) systems, in the sense of

the famous Born-Oppenheimer approximation from molecular physics that assumes nuclear motions are so slow in comparison with electron dynamics that they can be effectively separated, at least on appropriately chosen trajectory ‘pieces’ that may characterize the various phase transitions available to such systems. Extending theory to nonstationary dynamics remains to be done. In sum, between phase transitions, we assume the system changes slowly, and that the asymptotic limit theorems of information and control theories are sufficiently accurate.

We carry out this approximation via a fairly standard Morse Function methodology (Pettini 2007).

Our system of interest is made up of a set of cognitive submodules that engage in crosstalk among themselves. At every scale and level of organization all such submodules are constrained by both their own internals and by the developmental paths and by the persistent regularities of the embedding environment, including the actions of adversaries and the regularities of ‘grammar’ and ‘syntax’ imposed by embedding selection pressures.

There are, in addition structured uncertainties imposed by the large deviations possible within that environment, again including the behaviors of adversaries/partners who may be constrained by quite different developmental trajectories and ‘punctuated equilibrium’ evolutionary transitions.

The Morse Function construction assumes a number of interacting components:

- As Atlan and Cohen (1998) argue, cognition requires choice that reduces uncertainty. Such reduction in uncertainty directly implies the existence of an information source ‘dual’ to that cognition at each scale and level of organization. The argument is unambiguous and sufficient.
- Cognitive processes – under the Data Rate Theorem – are highly regulated, in the same sense that the human ‘stream of consciousness’ flows between cultural and social ‘riverbanks’. That is, a cognitive information source  $X_i$  is generally paired with a regulatory information source  $X^i$ .
- Environments (in a large sense) impose temporal event sequences of very high probability: night follows day, hot seasons follow cold, wet season follows dry, and so on. Thus environments impose their own ‘meaningful statements’ onto entities and interactions embedded within them via an information source  $V$ .
- ‘Large deviations’, following Champagnat et al. (2006) and Dembo and Zeitouni (1998), also involve sets of high probability developmental pathways, often governed by ‘entropy’-like laws that imply the existence of yet one more information source  $L_D$ .
- Consequently, full system dynamics are characterized by a joint and path dependent nonergodic information source uncertainty

$$H(\{X_i, X^i\}, V, L_D) \leq \sum_i H(X_i, X^i) + H(V) + H(L_D) \quad (2.2)$$

by the information theory chain rule (Cover and Thomas 2006).

The joint information source  $H$  is fully-characterized by the information sources  $X_i, X^i, V$  and  $L_D$ . The chain rule, however, with the identification of information as a form of free energy (Feynman 2000), ensures that preventing crosstalk between information sources, as the electrical engineers will attest, is far more difficult than embracing it as an essential underlying mechanism, subject to evolutionary exaptation for a great variety of uses. This is a fundamental point.

Individual dynamic paths can be assigned a value of that joint source uncertainty, denoted by  $H(x)$  for a path  $x$ .

This ‘fundamental representation’ is now defined by individual dynamic path values of source uncertainty and not represented as an ‘entropy’ function defined for all high-probability paths by an underlying probability distribution (Khinchin 1957). Each path has a possibly unique  $H$ -value, but the functional form of that value is not known in terms of some probability distribution.

In summary, the set  $\{X_i, X^i\}$  includes the internal interactive cognitive dual information sources of the system of interest and their associated regulators,  $V$  is taken as the information source of the embedding environment that may include the actions and intents of adversaries/symbionts, as well as ‘weather’.  $L_D$  is the information source of the associated large deviations possible to the system, possibly including ‘punctuated equilibrium’ evolutionary transitions.

Again, the full matrix of essential resources and their interactions has been projected onto a scalar rate index  $Z$ , according to the argument following Eq.(1). This may not always be possible, leading to the multidimensional complexities described in Wallace (2021a).

The underlying equivalence classes of developmental-behavioral-dynamic system paths used to define groupoid symmetries can now be defined fully in terms of the magnitude of individual path source uncertainties of individual dynamic paths  $H(x_j)$  such that  $x_j = \{x_j^0, x_j^1, \dots, x_j^n, \dots\}$  at times  $m = 0, 1, 2, \dots, n \rightarrow \infty$  alone. Again, see Khinchin (1957) for details of the nonergodic limit argument. The essential point is that individual paths of sufficient length have associated source uncertainty scalar values that are not calculated as standard Shannon ‘entropies’ across some probability distribution.

It is possible to envision the equivalence classes of behavioral/developmental paths as defined by the ‘game’ an organism is playing: growing from inception, foraging for food or habitat, evading predation, wound healing, mating/reproducing, and similarly for fully mechanical systems. Paths within each ‘game’ are taken as equivalent. As Weinstein (1996) shows in detail, this division of developmental/behavioral paths defines a groupoid. From a human perspective, sets of behavioral paths associated with baseball, football, soccer, rugby, tennis, and so on, are easily discernible and placed in appropriate equivalence classes.

Recall, as well, the conundrum of the ergodic decomposition of nonergodic information sources. It is formally possible to express a nonergodic source as the composition of a sufficient number of ergodic sources, much as it is possible to reduce planetary orbits to a Fourier sum of circular epicycles, obscuring the basic

dynamics. Hoyrup (2013) discusses the problem further, finding that ergodic decompositions are not necessarily computable. Here, we finesse the matter by focusing only on the values of the source uncertainties associated with dynamic paths.

### 3. Dynamics of cognitive systems

Since we assume the system to be nonergodic, each possible grammatical/syntactical high probability ‘statement’  $x_j = \{x_j^1, x_j^2, \dots\}$  emitted by the dual information source will have its own value of source uncertainty, say  $H_j$ .

Since, as Feynman (2000) insists, information is a form of free energy, we can write a Boltzmann pseudoprobability for based on enumeration of high probability developmental pathways  $x_j, j = 1, 2, \dots$  available to the system – each with uncertainty  $H(x_j) \equiv H_j$  so that

$$P_j = \frac{\exp[-H_j/g(Z)]}{\sum_k \exp[-H_k/g(Z)]} \quad (3.1)$$

Again,  $H_j$  is the source uncertainty of the high probability path  $j$ , not assumed represented as a ‘Shannon entropy’ since do not restrict dynamics to ergodic sources.  $g(Z)$  is a temperature analog scalar – possibly a scalar matrix function  $g(\mathbf{Z})$  – that must be calculated from first principles.

This step directly imposes a version of the usual ‘free energy’ in statistical physics as constructed from a partition function (Landau and Lifshitz 2007), using Feynman’s (2000) central insight that ‘information’ is a form of free energy.

This argument is a generalization of the fundamental assumption behind the Shannon-McMillan Theorem of information theory (Khinchin 1957): in the limit of ‘infinite length’, it is possible to divide the full set of individual dynamic paths into two distinct equivalence classes. The first is a small set of high probability paths consonant with a characteristic ‘grammar’ and ‘syntax’ making sense within the venue of the system of interest and its environment. The second set is a much larger one of paths having vanishingly low probability that are not consonant with grammar and syntax, a set of measure zero. Characterization of such ‘grammar’ and ‘syntax’ will almost never be trivial, as was the case of the Genetic Code (Marshall 2014).

Matters become rapidly more complicated.

The temperature-analog characterizing the system,  $g(Z)$  in Eq.(3.1), can be calculated via a first-order Onsager nonequilibrium thermodynamic approximation built from the partition function, i.e., the denominator of Eq.(3.1) (de Groot and Mazur 1984).

We next define an ‘iterated free energy’ Morse Function  $F$  (Pettini 2007) as



$$\begin{aligned}
 \exp[-F/g(Z)] &\equiv \sum_k \exp[-H_k/g(Z)] \equiv h(g(Z)) \\
 F(Z) &= -\log[h(g(Z))]g(Z) \\
 g(Z) &= -\frac{F(Z)}{\text{RootOf}\left(e^X - h\left(-\frac{F(Z)}{X}\right)\right)} \tag{3.2}
 \end{aligned}$$

where  $X$  is a dummy variate and the sum is over all possible high probability developmental paths of the system, again, those consistent with an underlying grammar and syntax. Recall that system paths not consonant with grammar and syntax constitute a set of measure zero that has very many more members than the set of high probability paths.

The differentiation of dynamic paths into high and low probability equivalence classes of behaviors represents a first groupoid ‘symmetry breaking’, in the sense of Weinstein (1996). Recall that equivalence classes lead canonically to groupoids, which are a generalization of ‘ordinary’ groups in which there may not be products defined between any two elements, and thus there may be very many unit elements. Wallace (2022a) elevates this fact to a ‘big bang’ analog for the emergence of life.

Recall that Feynman (2000) – following Bennett – makes the direct argument that information itself is to be viewed as a form of free energy, using Bennett’s clever ideal machine that turns a message directly into work. By contrast, we are concerned here with an iterated, rather than a direct, construction.

$F$ , now a free energy, then becomes subject to symmetry-breaking transitions as  $g(Z)$  varies (Pettini 2007). These symmetry changes, however, are not as associated with physical phase transitions as represented by standard group algebras. Such symmetry changes represent transitions from playing one ‘game’ to playing another. For example, a cognitive system may engage in foraging behaviors that trigger a predatory attack by another system. Then the game changes from ‘foraging’ to ‘escape’.

Thus ‘cognitive phase changes’ involve shifts between equivalence classes of high probability developmental/behavioral pathways that are represented as groupoids. Again, this is a generalization of the group concept such that a product is not necessarily defined for every possible element pair, although multiple products with multiple identity elements are defined (Brown 1992; Cayron 2006; Weinstein 1996).

Dynamics emerge via a first-order Onsager approximation akin to that of nonequilibrium thermodynamics (de Groot and Mazur 1984) in the gradient of an entropy measure constructed from the ‘iterated free energy’  $F$  of Eq.(3.2). Recall from the Onsager approximation that  $\partial Z/\partial t \approx \partial S/\partial Z$ . The full development is

$$\begin{aligned}
 S(Z) &\equiv -F(Z) + ZdF(Z)/dZ \\
 \partial Z/\partial t &\approx dS/dZ = f(Z) \\
 f(Z) &= Zd^2F/dZ^2 \\
 g(Z) &= \\
 &\frac{-C_1Z - \left(\int \frac{f(Z)}{Z} dZ\right)Z + C_2 + \int f(Z) dZ}{\text{RootOf}\left(e^Q - h\left(-\frac{C_1Z + \left(\int \frac{f(Z)}{Z} dZ\right)Z - C_2 - \int f(Z) dZ}{Q}\right)\right)} \tag{3.3}
 \end{aligned}$$

where the last relation follows from an expansion of the third part of Eq.(3.3) using the second expression of Eq.(3.2).

Several central matters emerge:

- Since ‘RootOf’ may have complex number solutions, the temperature analog  $g(Z)$  enters the realm of the ‘Fisher Zeros’ characterizing phase transition in physical systems (e.g., Dolan et al. 2001; Fisher 1965; Ruelle 1964 Sec. 5). Such phase transitions lie at the base of the punctuated accession to consciousness in the Baars, and related, Global Workspace Models. Indeed, the ‘RootOf’ construction in Eqs.(3.2) and (3.3) actually generalizes the Lambert W-function (e.g., Yi et al. 2010; Mezo and Keady 2015).

- Information sources are not microreversible, that is, palindromes are highly improbable, e.g., ‘eht’ has far lower probability than ‘the’ in English, so that there are no ‘Onsager Reciprocal Relations’ in higher dimensional systems. The necessity of groupoid symmetries appears to be driven by this directed homotopy.

- Further, there will always be a delay in the rate of provision of  $Z$ , so that, in Eq.(3.3), for example,  $f(Z) = \beta - \alpha Z(t)$  – an exponential model having  $Z(t) = (\beta/\alpha)(1 - \exp[-\alpha t])$  – where  $Z \rightarrow \beta/\alpha$  at a rate determined by  $\alpha$ . Other dynamics are possible, such as the ‘Arrhenius’,  $Z(t) = \beta \exp[-\alpha/t]$ , with  $f(Z) = (Z/\alpha)(\log(Z/\beta))^2$ . However  $Z$  is constructed from the components  $\mathcal{C}$ ,  $\mathcal{Q}$  and  $\mathcal{M}$ , so here, it is the scalar resource rate  $Z$  itself that counts.

Suppose, in the first expression of Eq.(3.3), it is possible to approximate the sum across the high probability paths with an integral across some probability distribution  $\rho(x)$  over the range  $x = 0 \rightarrow \infty$ . Then, by the change-of-variables theorem,

$$\exp[-F/g(Z)] \approx \int_0^\infty \rho[H/g(Z)]dH = \int_{0/g}^{\infty/g} \rho(u)g(Z)du = g(Z) \tag{3.4}$$

since  $\int_0^\infty \rho(u)du = 1$ .

$g(Z)$  must be real-valued and positive. Then

$$\begin{aligned} F(Z) &= -\log[g(Z)]g(Z) \\ g(Z) &= -F(Z)/W[n, -F(Z)] \end{aligned} \quad (3.5)$$

where  $W$  is the ‘simple’ Lambert  $W$ -function that satisfies  $W[n, x] \exp[W[n, x]] = x$ . It is real-valued only for  $n = 0, -1$  and only over limited ranges of  $x$  in each case.

In theory, specification of any two of the functions  $f, g, h$  permits calculation of the third.  $h$ , however, is determined – fixed – by the internal structure of the larger system. Similarly, ‘boundary conditions’  $C_1, C_2$  are externally-imposed, further sculpting dynamic properties of the ‘temperature’  $g(Z)$ , and  $f$  determines the rate at which the composite essential resource  $Z$  can be delivered. Both information and other free energy/material resources are rate-limited.

A rate-of-cognition can be calculated according to the usual chemical kinetics formalism (e.g., Laidler 1987):

$$\begin{aligned} L &= \frac{\sum_{H_j > H_0} \exp[-H_j/g(Z)]}{\sum_k \exp[-H_k/g(Z)]} \\ L &= \frac{\int_{H_0}^{\infty} \rho[H/g(Z)]dH}{\int_0^{\infty} \rho[H/g(Z)]dH} = \\ &= \frac{1}{g(Z)} \int_{H_0}^{\infty} \rho[H/g(Z)]dH \end{aligned} \quad (3.6)$$

where  $H_0$  is the ‘activation level’ of the cognitive process.

The punctuated phase transitions implied by the *RootOf* constructs in Eqs.(3.2) and (3.3) carry directly over into cognition rate calculations.

We provide a ‘simple’ example.

Suppose we have a two-state  $2N$ -level system about a ‘detection level’ so that  $H_{\pm} = H_0 \pm \delta$  for some small  $\delta > 0$ . Thus the first expressions of Eq.(3.2) and Eq.(3.6) become

$$\begin{aligned} \exp[-F/g(Z)] &= N [\exp[-(H_0 + \delta)/g(Z)] + \exp[-(H_0 - \delta)/g(Z)]] \\ &= N \exp[-H_0/g(Z)] 2 \cosh[\delta/g(Z)] \\ L &= \frac{1}{1 + \exp[2\delta/g(Z)]} \end{aligned} \quad (3.7)$$

Solving the first expression in Eq.(3.7) for  $F[g(Z)]$  and then using the second and third expressions of Eq.(3.3) for an ‘exponential’ model  $dZ/dt = f(Z) = \beta - \alpha Z$  produces a complicated equation that, when approximated to third order in  $\delta$  becomes

$$\begin{aligned}
 & -Z (\ln(2) + \ln(N)) \left( \frac{d^2}{dZ^2} g(Z) \right) + \\
 & Z \left( \frac{\frac{d^2}{dZ^2} g(Z)}{2g(Z)^2} - \frac{\left( \frac{d}{dZ} g(Z) \right)^2}{g(Z)^3} \right) \delta^2 \approx \beta - \alpha Z
 \end{aligned} \tag{3.8}$$

This can be explicitly solved for  $g(Z)$ , and hence for  $L(Z)$ . The solutions to this approximation have two boundary conditions which we call  $C_1$  and  $C_2$ .

Figure 1a shows a ‘Yerkes-Dodson’ cognition rate graph fixing  $N = 1$ ,  $\alpha = 1$ ,  $\delta = 1/10$ . Remember that, for an ‘exponential’ model,  $Z \rightarrow \beta/\alpha$ .  $\beta$  then varies as the ‘arousal’, where we have taken  $C_1 = -1$  and  $C_2 = -1/4$ .

The first part of figure 1a is the classic inverted-U signal transduction characteristic of Yerkes-Dodson studies, where increasing ‘arousal’ at first raises attentional focus, and then degrades it. The second part of 1a, a step function, characterizes onset of ‘panic’ once a threshold in arousal is exceeded.

We can, via a stochastic version of the second part of Eq.(3.3), examine the effect of increasing ‘noise’ on this system, beyond ‘arousal’. In particular, we fix  $\beta$  at the maximum of the inverted-U, here  $\beta = 3$ , and use the Ito Chain Rule to study cognition rate dynamics based on the stochastic differential equation

$$\begin{aligned}
 dZ_t &= f(Z_t)dt + \sigma Z_t dB_t \\
 &= (\beta - \alpha Z_t)dt + \sigma Z_t dB_t
 \end{aligned} \tag{3.9}$$

where the second term represents classic ‘volatility’ under Brownian noise.

We then use the Chain Rule to calculate the nonequilibrium steady state average for the cognition rate,  $\langle dL_t \rangle = 0$ , giving, in figure 1b, the ‘solution set’  $\{\sigma, Z\}$  under the conditions of figure 1a. For the general base relation, application of the Ito Chain Rule to  $Z^2$  finds the condition for stability in variance is  $\sigma^2/2 < \alpha$ . Here, with  $\alpha = 1$ , a bifurcation phase transition becomes manifest well before the critical value  $\sigma = \sqrt{2}$ .

The Mathematical Appendix extends the model to ‘embodiment,’ a central matter in robotics, by incorporating cognition and control under the single aegis of the Rate Distortion Theorem, modulo the Feynman/Bennett identification of information as a form of free energy.

#### 4. What this model is and is not

Although Tononi’s ‘Integrated Information’ approach to consciousness invokes the term ‘information’ as a kind of shibboleth (e.g., Tononi et al. 2016; Tegmark 2016), nowhere does an asymptotic limit theorem corresponding to the construct appear. In contrast to that work, what we do here is very much situated within the four basic asymptotic limit theorems of information and control theories. Indeed, the-

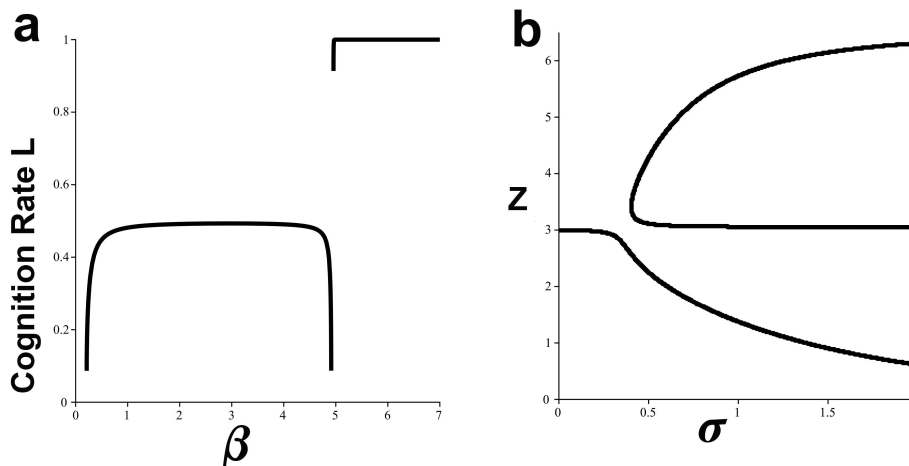


Fig. 1. **a.** A ‘Yerkes-Dodson’ cognition rate graph based on Eqs.(3.7) and (3.8). Here,  $N = 1$ ,  $\alpha = 1$ ,  $\delta = 1/10$ . Recall that  $Z \rightarrow \beta/\alpha$ .  $\beta$  then varies as the ‘arousal’. The boundary conditions are  $C_1 = -1$ ,  $C_2 = -1/4$ . The first part of the figure is the classic inverted-U signal transduction characteristic of Yerkes-Dodson studies. The second is a step function to ‘panic’ once an arousal threshold is exceeded. **b.** The solution set  $\{\sigma, Z\}$  for the nonequilibrium steady state relation  $\langle dL_t \rangle = 0$  based on figure 1a, taking  $\beta = 3$ . A bifurcation phase transition is evident well before the variance instability point  $\sigma = \sqrt{2}$ .

oretical progress in consciousness studies may depend on recognition of additional asymptotic limit theorems.

Similarly, although we invoke both information-as-free-energy, and an iterated free energy Morse Function based on that invocation, what we do is not at all like Friston’s ‘Free Energy Principle’ (Bogacz 2017), seen as literally driving the world. There is no comprehensive ‘principle’ here, any more than there is in Onsager’s approach to nonequilibrium thermodynamics: first order models can fail badly. More specifically, the formalisms above are not written in stone, are not a ‘General Relativity of Consciousness’. They are an attempt to derive statistical tools for the analysis of experimental and observational data on cognitive processes and their dynamics, up to and including consciousness. Think of analogs to t-tests and regression equations, but crafted by the theorems of information and control theories.

Nonetheless, what stands out here in particular are the punctuations inherent to the *RootOf* constructs of Eqs.(3.2) and (3.3), to be characterized as exact analogs to Fisher Zero phase transitions in physical systems, but involving symmetry-breaking in groupoids, rather than groups. Punctuation is one of the most fundamental observations in the study of consciousness, via the in-and-out of the attention ‘spotlight’. We can, however, also explore punctuation from a different, but parallel, perspective, reconsidering matters from the front of the elephant, as it were, using a network model.

## 5. Punctuation redux

Suppose a set of cognitive modules within a particular system in an embedding environment becomes linked to solve a series of problems. A canonical example might be riding a trail bicycle over a particularly rough course at high speed, then to be followed by wound healing. The 'no free lunch' arguments of Wolpert and MacReady (1997) become significant. The essential point (English 1996) is that if an optimizer has been tuned to the most effective possible structure for a particular kind of problem or problem set, it will necessarily be worst for some other problem set, which must then have a different function optimizer for optimality. As Wallace (2012) puts it,

Another way of stating this conundrum is to say that a computed solution is simply the product of the information processing of a problem, and, by a very famous argument, information can never be gained simply by processing. Thus a problem  $X$  is transmitted as a message by an information processing channel,  $Y$ , a computing device, and recoded as an answer. By [a direct] 'tuning theorem' argument [parallel to the Coding Theorem of information theory], there will be a channel coding of  $Y$  which, when properly tuned, is itself most efficiently 'transmitted', in a sense, by the problem – the 'message'  $X$ . In general, then, the most efficient coding of the transmission channel, that is, the best algorithm turning a problem into a solution, will necessarily be highly problem-specific. Thus there can be no best algorithm for all sets of problems, although there will likely be an optimal algorithm for any given set.

Something analogous has long been known, but from a distinctly different viewpoint (Chiang and Boyd 2004): Shannon (1959) found a curious and provocative duality between the properties of an information source with a distortion measure and those of a channel. This duality is enhanced if we consider channels in which there is a cost associated with the different letters. Solving this problem corresponds, in a sense, to finding a source that is right for the channel and the desired cost. In a somewhat dual way, evaluating the rate distortion function for a source corresponds to finding a channel that is just right for the source and allowed distortion level.

We will use a related Rate Distortion perspective in the Mathematical Appendix to study the necessary 'embodiment' of cognition and control in robotic systems.

From these considerations, it becomes clear that different challenges facing an environmentally-embedded cognitive system must be met by different arrangements of cooperating lower level cognitive modules. It is possible to make an abstract picture of this based only on the network of linkages between the information sources dual to the physiological and learned unconscious cognitive modules (UCM) that may become entrained into address of those challenges. The network of lower level cognitive modules is reexpressed in terms of the information sources dual to them.

Given two distinct problem classes (e.g., bicycle riding vs. wound healing), there must be two markedly different wirings of the information sources dual to the available UCM.

The possible expansion of a closely-linked set of information sources dual to the UCM into a global broadcast depends, for this model, on the underlying network topology of the dual information sources and on the strength of the couplings between the individual components of that network.

The simplest model of such phenomena is instantiated by phase transitions in a random network.

The appearance of the Lambert W-function in the arguments above – for the simple case  $h(g(Z)) = g(Z)$  – is of interest. The fraction of nodes within the ‘giant component’ of a random network of  $N$  nodes – here, taken as interacting information sources dual to unconscious cognitive processes – can be described in terms of the probability of contact between nodes,  $p$ , as (Newman 2010)

$$\frac{W(0, -Np \exp[-Np]) + Np}{Np} \quad (5.1)$$

giving the results of figure 2.

Note, in particular, the threshold for highly punctuated onset of a single giant component in the random network case. This sort of dynamic is a central matter for the high-speed neural processes that are the foundation of consciousness, seen here as a necessarily stripped-down example of more general multiple tunable spotlight physiological phenomena that can do much more, but only if they operate at far slower rates: gene expression, immune function, tumor suppression, and so on (Wallace 2022).

An important feature here is the topological tunability of the threshold dynamics implied by the two limiting cases, the star-of-stars-of-stars vs. the random network.

Lambert W-functions thus appear to suggest existence of an underlying formal network structure. For our purposes here – neural structures and their machine analogs – we can envision the underlying abstract network to be a set of information sources dual to unconscious cognitive phenomena within the brain. These become linked by ‘ $Np$ ’ crosstalk, in the context of a tunable topology that shifts somewhere between the two limits of the figure.

Figure 6 of Dehaene and Changeux (2011) shows something similar.

Previous sections have abducted results from nonequilibrium thermodynamics to consciousness theory, applicable to nonergodic, as well as ergodic, models of cognition. Here, we abduct the Kadanoff renormalization treatment of physical phase transitions (Wallace 2012), applying it to a reduced version of the iterated ‘free energy’ Morse Function of Eq.(3.2).

Although a more general argument can be made, representing embodied consciousness in itself, for simplicity, we project down on to the internal subsystem dominated by  $\mathcal{C}$ , the internal system bandwidth, envisioning a number of internal

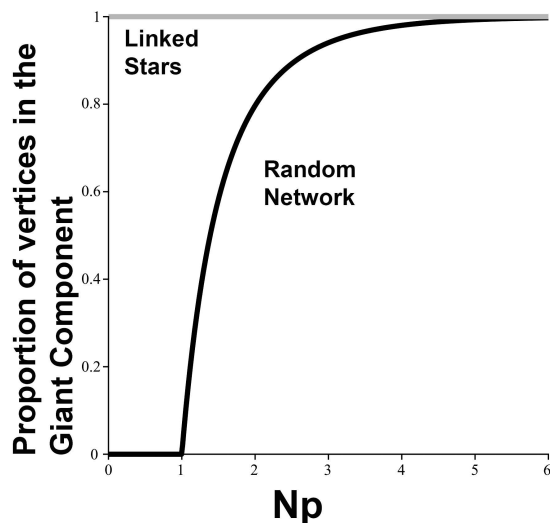


Fig. 2. Proportion of  $N$  interacting information sources dual to unconscious cognitive processes that are entrained into a ‘giant component’ global broadcast as a function of the probability of contact  $p$  for random and stars-of-stars-of-stars topologies. Tuning topologies determines the threshold for ‘ignition’ to a single, large-scale, global broadcast.

cognitive submodules as connected into a topologically identifiable network having a variable average number of fixed-strength crosstalk linkages between components. The mutual information measure of crosstalk can continuously change, and it becomes then possible to conduct a parameterized renormalization in a now-standard manner (e.g., Wilson 1971; Wallace 2022, Sec. 6.6).

The internal modular network linked by information exchange has a topology depending on the magnitude of interaction. Define an interaction parameter, a real number  $\omega > 0$ , and examine structures characterized in terms of linkages set to zero if crosstalk is less than  $\omega$ , and renormalized to 1 if greater than or equal to  $\omega$ . Each  $\omega$  defines, in turn, a network ‘giant component’ (Spenser 2010), linked by information exchange greater than or equal to it.

Now invert the argument: a given topology of interacting submodules making up a giant component will, in turn, define some critical value  $\omega_C$  such that network elements interacting by information exchange at a rate less than that value will be excluded from that component, will be locked out and not ‘consciously’ perceived.

$\omega$  is a tunable, syntactically dependent, detection limit depending on the instantaneous topology of the giant component of linked cognitive submodules defining, by that linkage, a ‘global broadcast’.

For ‘slow’ systems (Wallace 2012) – immune response, gene expression, institutional process – as opposed to the 100 ms time constant of higher animal consciousness, there can be many such ‘global workspace’ spotlights acting simultaneously.



Such multiple global broadcasts, indexed by the set  $\Omega = \{\omega_1, \omega_2, \dots\}$ , lessen the likelihood of inattentive blindness to critical signals, both internal and external. The immune system, for example, engages simultaneously in pathogen and malignancy attack, neuro-immune dialog, and routine tissue maintenance (Cohen 2000).

## 6. A red flag

The previous sections have not, in fact, presented ‘a’ theory of consciousness. Rather, they have explored a spectrum of models and modeling approaches that might be adapted as tools for the empirical and observational study of conscious phenomena involving organisms, machines, and their increasingly ubiquitous composites. This may seem disappointing, but it reflects an important underlying reality: there never can be ‘a’ theory of consciousness. Data are to be found across a broad range of human studies, most particularly regarding the role of culture in determining brain hard-wiring for basic perception. We follow closely chapter 13 of Wallace and Wallace (2016).

Nisbett et al. (2001), consonant with a long line of research (e.g., Markus and Kitayama 1991; Heine 2001), review an extensive range of empirical studies regarding basic cognitive differences between individuals raised in East Asian and Western cultural heritages. These are characterized, respectively, as ‘holistic’ and ‘analytic’. Their argument is roughly as follows:

- Social organization directs attention to some aspects of the perceptual field at the expense of others.
  - What is attended to influences metaphysics.
  - Metaphysics guides tacit epistemology, that is, beliefs about the nature of the world and causality.
  - Epistemology dictates the development and application of some cognitive processes at the expense of others.
    - Social organization can directly affect the plausibility of metaphysical assumptions, such as whether causality should be regarded as residing in the field vs. in the object.
    - Social organization and social practice can directly influence the development and use of cognitive processes such as dialectical vs. logical ones.
    - Thus tools of thought embody a culture’s intellectual history, tools have theories built into them, and users accept these theories, often unknowingly, when they use these tools.

Masuda and Nisbett (2006), somewhat later, argue that research on perception and cognition suggests 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 suggest that *there can be cultural variation in what may seem to be*

*basic perceptual processes*. Nisbett and Miyamoto (2005) argue that fundamental perceptual processes are influenced by culture.

These findings establish a dynamic relationship between the cultural context and perceptual processes, implying strongly that perception can no longer be regarded as consisting of processes that are universal across all people at all times.

Wallace (2007), somewhat in parallel, explores analogous dynamics involving inattentional blindness and culture.

Broadly, then, within human populations there can be no single model of ‘consciousness’. This suggests, in turn, that those seeking to build conscious machines should be prepared to explore a very wide dynamic landscape using a very adaptable toolbox: ‘horses for courses’.

## 7. Implications for ‘machine consciousness’

Consciousness in higher animals is a necessarily stripped-down version of such complex cognition/regulation dyads as gene expression, protein folding, wound healing, immune maintenance/pathogen fighting, and similar institutional level phenomena, all of which have the ‘luxury’ of relatively large characteristic response times (Wallace 2012). These larger processes can assemble a shifting, tunable multiple-workspace ‘workforce’ to address patterns of challenge or opportunity. Consciousness, under evolutionary selection pressures, has emerged to operate with a 100 ms time constant across ‘wet chemistry’ neurons, modulo their order of magnitude higher rate of consumption of metabolic free energy than other tissues. While this is no small thing, such high speed requires single-workspace simplicity, backed-up by ‘hindbrain’ conditioned emotional responses that can be learned (or inherited) to short-circuit lengthy conscious ruminations. Even very ‘stupid’ animals have really good emotional tuning.

This constellation of evolutionary circumstance has some implications for machine design.

First, it should not be particularly difficult to construct a fast, single workspace ‘conscious machine’ that closely mimics the human tunable neuronal global workspace system. Tie it to an AI that has ‘learned’ hyperrapid pattern responses to some characteristic set of challenges, and you have an ‘emotional’ conscious machine.

Second, a particularly clever designer may want to mimic the more capable multiple-workspace systems of gene expression, protein folding, immune response, institutional function, and so on. High speed fighter aircraft, built as inherently unstable to permit rapid maneuver, already use odd numbers of parallel control systems to ‘vote’ on control dynamics. Such systems are resistant to inattentional blindness and related failings of overfocus and thrashing.

Multiple-consciousness tunable global workspace machines, particularly when coupled with high-speed AI ‘emotions’, will likely strongly dominate single-workspace conscious machines under virtually all circumstances of fog-of-war con-

frontation. See Wallace (2021) for a somewhat parallel argument focused on AI alone. But such entities will surely be nothing like having the family cat on your lap.

Machines are constructed to purpose, built around – and reflecting – the jobs they are to do: ‘horses for courses’. Sometimes, it is true, machines made for one purpose can serve another. Although it indeed seems possible to build machines that parallel much of what we know of higher animal consciousness, their realms of actual utility are far from clear, discounting the usual dystopian fantasies and allied social constructions.

## 8. Mathematical Appendix

### *A brief reformulation*

A centrality of control theory can be reformulated as a rate distortion problem (Cover and Thomas 2006) involving the sending a ‘control signal’ along the ‘noisy channel’ of the system being controlled, and then using a scalar distortion measure to compare the ‘message’ sent with the actual behavior of the system itself. Recall that, for a particular channel and a particular scalar average distortion measure  $D$ ,  $R(D)$  is defined the minimum channel capacity needed to keep average distortion less than or equal to  $R$ .  $R(D)$  can be show to be convex in  $D$ , and this is a central property. Typically,  $D \rightarrow 0$  only as  $R(D) \rightarrow \infty$ .

We again take  $R(D)$  as our basic (here, uniterated) free energy measure, and define an ‘entropy’ as

$$S = -R(D) + DdR/dD \quad (8.1)$$

subject to the ‘Onsager’ relation

$$dD/dt \propto dS/dD = Dd^2R/dD^2 \quad (8.2)$$

For the ‘worst case’ Gaussian channel, under the square distortion measure (Cover and Thomas 2006)

$$\begin{aligned} R(D) &= \frac{1}{2} \log_2(\sigma^2/D), \quad D \leq \sigma^2 \\ R(D) &= 0, \quad D > \sigma^2 \end{aligned} \quad (8.3)$$

Plugging in,

$$\begin{aligned} dD/dt &= \frac{1}{2 \log(2) D(t)} \\ D(t) &\propto \sqrt{t} \end{aligned} \quad (8.4)$$

where the last expression should be recognized as the standard diffusion result.

We can even obtain a simple version of the Data Rate Theorem from this formalism.

In the absence of imposed ‘control’, in a large sense, Eq. (8.4) states that – for a Gaussian channel – distortion simply grows under ‘ordinary’ diffusion.

This result, almost incidentally, validates the generalized Onsager nonequilibrium formalism used above. The method further serves for a heuristic derivation of the relation between information and control theories, the Data Rate Theorem (Nair et al. 2007) characterizing the minimum control information  $\mathcal{H}$  needed to stabilize an inherently unstable system.

Let us impose sufficient stabilizing ‘control free energy’ at some rate  $M(\mathcal{H})$  on an inherently unstable system diffusing from control according to some monotonic function of time, depending on the nature of the ‘control channel’. A necessary condition for stability is

$$\begin{aligned} dD/dt &= \mu D d^2 R/dD^2 - M(\mathcal{H}) \leq 0 \\ M(\mathcal{H}) &\geq \mu D d^2 R/dD^2 \geq 0 \\ \mathcal{H} \geq \mathcal{H}_0 &\equiv \max\{M^{-1}(\mu D d^2 R/dD^2)\} \end{aligned} \quad (8.5)$$

Here, max represents the maximum value, noting that, if  $M$  is monotonic increasing, so is the inverse function  $M^{-1}$ . By the convexity of  $R(D)$ ,  $d^2 R/dD^2 \geq 0$  (Cover and Thomas 2006).

### **Embodiment**

‘Embodiment’ – the melding of control with information across an intelligent system – emerges quite directly, in terms of Eqs.(3.4)-(3.6), since  $R : 0 \rightarrow \infty$ . That is, we again impose an appropriate probability density function  $\rho(\omega, x)$  defined on  $x : 0 \rightarrow \infty$ , where  $\omega$  is a parameter set, and write

$$\begin{aligned} \exp[-F/g(Z)] &= \int_0^\infty \rho(\omega, R/g(Z))dR = g(Z) \\ F(Z) &= -\log[g(Z)]g(Z) \\ g(Z) &= \frac{-F(Z)}{W[n, -F(Z)]} \\ L &= \frac{1}{g(Z)} \int_{R_0}^\infty \rho(\omega, R/g(Z))dR \end{aligned} \quad (8.6)$$

$W(n, x)$  is the Lambert W-function of order  $n$  and  $R_0$  is a minimum detection threshold.

The standard physics approach takes  $\rho(\omega, R) = \exp[-R]$ , leading to the cognition rate  $L = \exp[-R_0/g(Z)]$ . In contrast, as a somewhat perverse example, we instead take the more versatile chi distribution as the base, having distribution function

20 *Rodrick Wallace*

$$\rho(k, X) = \frac{X^{k-1} e^{-\frac{X^2}{2}}}{2^{\frac{k}{2}-1} \Gamma\left(\frac{k}{2}\right)} \quad (8.7)$$

with mode  $\sqrt{k-1}$  for  $k \geq 1$ . We first set  $k = 5$ , as in figure 2a, with the mode at  $X = 2$ .

The last expression in Eq.(8.6) gives

$$L = \frac{\Gamma\left(\frac{k}{2}, \frac{R_0^2}{2g^2}\right)}{\Gamma\left(\frac{k}{2}\right)} \\ = -\frac{\sqrt{2} \left( \operatorname{erf}\left(\frac{\sqrt{2}}{g}\right) \sqrt{\pi} g^3 \sqrt{2} - \sqrt{\pi} g^3 \sqrt{2} - 4 e^{-\frac{2}{g^2}} g^2 - \frac{16 e^{-\frac{2}{g^2}}}{3} \right)}{2\sqrt{\pi} g^3} \quad (8.8)$$

for  $k = 5$ ,  $R_0 = 2$ .

Taking  $dZ/dt = f(Z) = \beta - \alpha Z$  in Eq.(3.3), so that  $Z(t) \rightarrow \beta/\alpha$ , gives

$$F(Z) = \ln(Z) Z\beta - Z\beta - \frac{Z^2\alpha}{2} + C_1 Z + C_2 \quad (8.9)$$

where  $C_1$  and  $C_2$  are characteristic boundary conditions.

Again, we take  $R_0$  in Eq.(8.6) at the mode of figure 3a, i.e.,  $X = 2$ . Then, setting  $\alpha = 1$ ,  $C_1 = -1$ ,  $C_2 = 1$ , and using the Lambert W-function of order zero gives the cognition rate in figure 3b as a function of increasing  $\beta$ . Recall that the zero-order Lambert W-function  $W(0, x)$  is real-valued only for  $x \geq -\exp[-1]$ . In consequence, the cognition rate displays a characteristic inverted-U signal transduction pattern analogous to the classic Yerkes-Dodson effect that describes cognition under increasing levels of ‘arousal’.

Further, imposing a stochastic differential equation on  $Z$  as

$$dZ_t = (\beta - \alpha Z_t)dt + \sigma Z_t dB_t \quad (8.10)$$

where  $dB_t$  is Brownian noise, permits calculation of  $dL_t$  using the Ito Chain Rule. Figure 3c fixes  $\beta = 3$ , the peak of figure 3b, and presents the solution set  $\{Z, \sigma\}$  to the averaging relation  $\langle dL_t \rangle = 0$ , again, using the Ito Chain Rule. The resulting equations challenge the computer algebra program. Sufficient ‘noise’  $\sigma$  drives the system to failure.

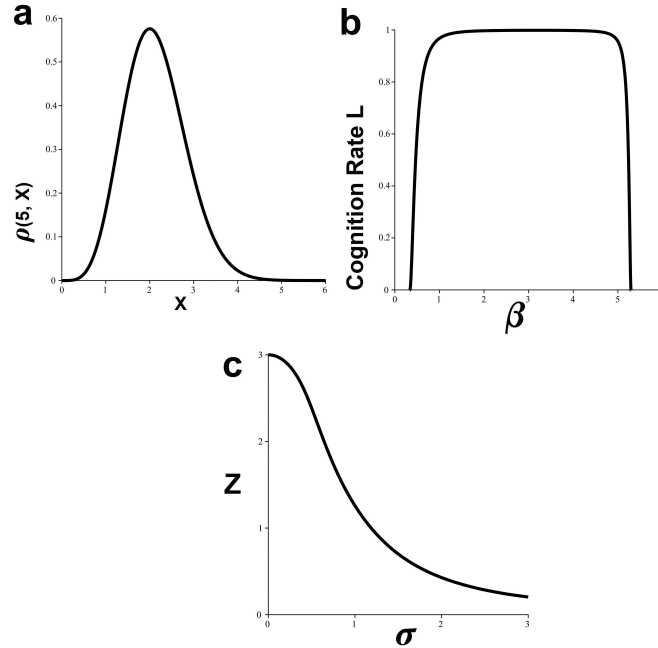


Fig. 3. **a.** Chi distribution of order  $k = 5$ , having a mode at  $X = 2$ . **b.** Cognition rate  $L(\beta)$  for the distribution of **a**, setting  $R_0 = 2$ , the distribution mode, assuming an exponential model  $dZ/dt = \beta - \alpha Z$ , with  $\alpha = 1$   $C_1 = -1$ ,  $C_2 = 1$  in the definition of  $F$ . **c.** Solution set  $\{Z, \sigma\}$  for the relation  $\langle dL_t \rangle = 0$  under the Ito Chain Rule, setting the system at  $\beta = 3$  in figure 3b. Sufficiently large  $\sigma$  drives the system to failure.

Figure 4a shows the chi distribution function  $\rho(1/2, x)$ , bunched-up to the left, roughly like the ‘standard’ physics model with  $\rho(\omega, R/g) = \exp[-R/g]$ . Figure 4b shows the analogous inverted-U signal transduction result, setting  $R_0 = 1$  so that

$$L = \frac{\sqrt{2} \Gamma(\frac{3}{4}) \Gamma(\frac{1}{4}, \frac{1}{2g^2})}{2\pi} \tag{8.11}$$

Figure 4c displays the stochastic analysis for Eq.(8.11), i.e.,  $\langle dL_t \rangle = 0$ , again taking  $\beta = 3$ . Here, however, the solution set shows onset of a bifurcation phase transition at sufficient imposed noise  $\sigma$ , much like figure 1b.

It is not difficult, using the first part of Eq.(8.8), to extend figures 2 and 3 across three dimensions, exploring a similar development allowing  $k$  to vary for a particular fixed  $R_0$ . In particular, the bifurcation in figure 3c disappears for sufficiently large  $k$ , representing another groupoid equivalence class phase transition.

One inference from these results is that regression model analogs based on the asymptotic limit theorems of information and control theories that describe cognition/consciousness will be as difficult to create, validate, and use as more ‘ordinary’

22 REFERENCES

statistical tools based on the Central Limit and other asymptotic limit theorems of probability theory.

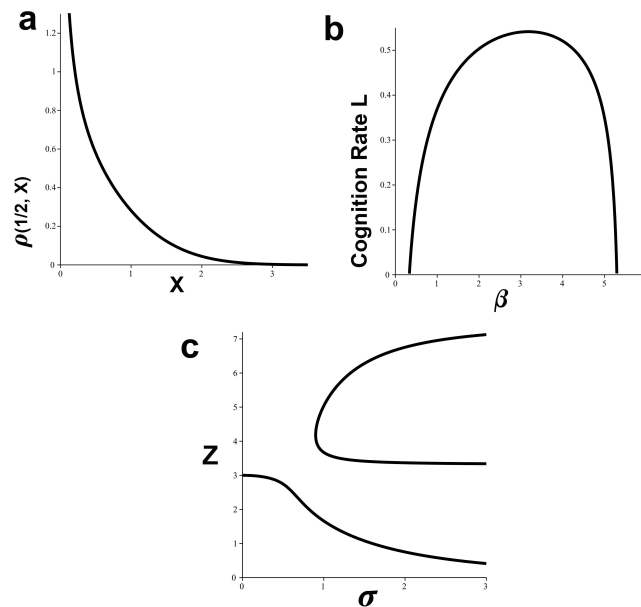


Fig. 4. **a.** Chi distribution of order  $k = 1/2$ , with ‘mode’ at zero. **b.** Signal transduction cognition rate calculation for  $dZ/dt = f(Z) = \beta - \alpha Z$ , with  $\alpha, C_1, C_2$  as in figure 3. The detection threshold is  $R_0 = 1$ . **c.** Solution set  $\{Z, \sigma\}$  for  $\langle dL_t \rangle \geq 0$ . In contrast with figure 3c, sufficiently large  $\sigma$  triggers a bifurcation phase transition.

**References**

Adams, F., 2003, The informational turn in philosophy, *Minds and Machines*, 13:471-501.

Atlan, H., I. Cohen, 1998, Immune information, self-organization, and meaning, *International Immunology*, 10:711-717.

Baars, B., 1988, *A Cognitive Theory of Consciousness*, Cambridge University Press, New York.

Baars, B., 2005, Global workspace theory of consciousness: toward a cognitive neuroscience of human experience, *Progress in Brain Research*, 150:45-53.

Baars, B., S. Franklin, 2003, How conscious experience and working memory interact, *Trends in Cognitive Science*, 7:166-172.

Baars, B., N. Gage, 2010, *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, Elsevier, Amsterdam.

Bogacz, R., 2017, A tutorial on the free-energy framework for modelling perception and learning, *Journal of Mathematical Psychology* 76:198211

- Brown, R., 1992, Out of line, Royal Institute Proceedings 64:207-243.
- Cayron, C., 2006, Groupoid of orientational variants, Acta Crystallographica Section A, A62:21040.
- Champagnat, N., R. Ferriere, S. Meleard, 2006, Unifying evolutionary dynamics: From individual stochastic process to macroscopic models, Theoretical Population Biology, 69:297-321.
- Chiang, M., S. Boyd, 2004, Geometric programming duals of channel capacity and rate distortion, IEEE Transactions on Information Theory, 50:245255.
- Chella, A., A. Cangelosi, G. Metta, S. Bringsjord, eds., 2019, Consciousness in Humanoid Robots, Lausanne: Frontiers Media, doi: 10.3389/978-2-88945-866-0.
- Cohen, I., 2000, Tending Adam's garden: evolving the cognitive immune self, Academic, New York.
- Cover, T., J. Thomas, 2006, Elements of Information Theory, Second Edition, Wiley, New York.
- de Groot, S., P. Mazur, 1984, Non-Equilibrium Thermodynamics, Dover, New York.
- Dehaene, S., J. Changeux, 2011, Experimental and theoretical approaches to conscious processing, Neuron 70:200227.
- Dembo, A., O. Zeitouni, 1998, Large deviations and applications, 2nd ed., Springer, New York.
- Dolan, B., W. Janke, D. Johnston, M. Stathakopoulos, 2001, Thin Fisher zeros, Journal of Physics A 34:6211-6223.
- Dretske, F., 1981, Knowledge and the Flow of Information, MIT Press, Cambridge, MA.
- Dretske, F., 1988, Explaining Behavior, MIT Press, Cambridge, MA.
- Dretske, F., 1994, The explanatory role of information, Philosophical Transactions of the Royal Society A, 349:59-70.
- English, T., 1996, Evaluation of evolutionary and genetic optimizers: no free lunch. In: Fogel, L., P. Angeline, T. Back (eds) Evolutionary programming V: proceedings of the fifth annual conference on evolutionary programming. MIT Press, Cambridge, pp 163169
- Feynman, R., 2000, Lectures on Computation, Westview Press, New York.
- Khinchin, A., 1957, The Mathematical Foundations of Information Theory, Dover, New York.
- Fisher, M., 1965 Lectures in Theoretical Physics vol. 7, University of Colorado Press, Boulder.
- Heine, S., 2001, Self as cultural product: an examination of East Asian and North American selves, Journal of Personality 69:881-906.
- Hoyrup, M., 2013, Computability of the ergodic decomposition, Annals of Pure and Applied Logic 164:542-549.
- Laidler, K., 1987, Chemical Kinetics, Third Edition, Harper and Row, New York.
- Landau, L., E. Lifshitz, 2007, Statistical Physics, 3rd Ed. Part 1, Elsevier, New York.
- Markus, H., S. Kitayama, 1991, Culture and the self-implications for cognition,



## 24 REFERENCES

- emotion, and motivation, *Psychological Review* 98:224-253.
- Marshall, J., 2014, The genetic code, *PNAS* 111:5760.
- Masuda, T., R. Nisbett, 2006, Culture and change blindness, *Cognitive Science* 30:381-399.
- Maturana, H., F. Varela, 1980, *Autopoiesis and Cognition: The realization of the living*, Reidel, Boston.
- Mezo I., G. Keady, 2015, Some physical applications of generalized Lambert functions, arXiv:1505.01555v2 [math.CA] 22 Jun 2015.
- Nair, G., F. Fagnani, S. Zampieri, R. Evans, 2007, Feedback control under data rate constraints: an overview, *Proceedings of the IEEE* 95, 108138.
- Newman M (2010) *Networks: an introduction*. Oxford University Press, New York.
- Nisbett, R., K. Peng, C. Incheol, A. Norenzayan, 2001, Culture and systems of thought: holistic vs. analytic cognition, *Psychological Review* 108:291-310.
- Nisbett, R., Y. Miyamoto, 2005, The influence of culture: holistic versus analytic perception, *Trends in Cognitive Science* 9:467-473.
- Pettini, M., 2007, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, New York.
- Ruelle, D., 1964, Cluster property of the correlation functions of classical gases, *Reviews of Modern Physics*, April, 580-584.
- Shannon, C., 1959, Coding theorems for a discrete source with a fidelity criterion, *Institute of Radio Engineers International Convention Record*, 7:142163.
- Spenser, J., 2010, The giant component: a golden anniversary, *Notes of the American Mathematical Society* 57:720724.
- Tegmark, M., 2016 Improved measures of integrated information, *PLOS Computational Biology* <https://doi.org/10.1371/journal.pcbi.1005123>.
- Tononi, G., M. Boly, M. Massimini, C. Koch, 2016, Integrated information theory: from consciousness to its physical substrate, *Nature Reviews Neuroscience* 17:450461.
- Wallace, R., 2007, Culture and inattentive blindness. *Journal of Theoretical Biology* 245:378-390.
- Wallace, R., 2012, Consciousness, crosstalk, and the mereological fallacy: an evolutionary perspective, *Physics of Life Reviews* 9:426453
- Wallace, R., 2021. How AI founders on adversarial landscapes of fog and friction, *Journal of Defense Modeling and Simulation* doi: 10.1177/1548512920962227.
- Wallace, R., 2022, *Consciousness, Cognition and Crosstalk: The evolutionary exaptation of nonergodic groupoid symmetry-breaking*, Springer Briefs in Applied Sciences and Technology, New York.
- Wallace, R., 2022a, Major transitions as groupoid symmetry-breaking in nonergodic prebiotic, biological and social information systems, *Acta Biotheoretica* 70:27 <https://doi.org/10.1007/s10441-022-09451-5>.
- Wallace, R., D. Wallace, 2016, *Gene Expression and its Discontents: The social production of chronic disease*, Second Edition, Springer, New York.
- Weinstein, A., 1996, *Groupoids: unifying internal and external symmetry*, *Notices of the American Mathematical Association*, 43:744-752.

- Wilson, K., 1971, Renormalization group and critical phenomena I, Renormalization group and the Kadanoff scaling picture, *Physical Review B* 4:31743183
- Wolpert, D., W. MacReady, 1997, No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computing* 1:6782
- Yi, S., P.W. Nelson and A.G. Ulsoy, 2010, *Time-Delay Systems: Analysis and Control Using the Lambert W Function*, World Scientific, New Jersey.