On the Inherent Instability of Biocognition: Toward New Probability Models and Statistical Tools

Rodrick Wallace [1] , **Irina Leonova [2,3]** and **Saikat Gochhait [3,4,]**

[1] The New York State Psychiatric Institute, 1051 Riverside Dr, New York, NY 10032, USA; rodrick.wallace@nyspi.columbia.edu
[2] Faculty of Social Sciences, Lobachevsky University, 603950 Nizhny Novgorod, Russia; irina.leonova@unn.ru
[3] Neuroscience Research Institute, Samara State Medical University, 89 Chapaevskaya str., 443001 Samara, Russia
[4] Symbiosis Institute of Digital and Telecom Management, Symbiosis International Deemed University, Symbiosis Knowledge Village, Village- Lavale, Tahasil- Mulshi, Pune 412115, India
[*] Correspondence: saikat.gochhait@sidtm.edu.in

Abstract: A central conundrum enshrouds biocognition: almost all such phenomena are inherently unstable and must be constantly controlled by external regulatory machinery to ensure proper function, in much the same sense that blood pressure and the 'stream of consciousness' require persistent delicate regulation for the survival of higher organisms. Here, we derive the Data Rate Theorem of control theory that characterizes such instability via the Rate Distortion Theorem of information theory for adiabatically stationary nonergodic systems. We then outline a novel approach to building new statistical tools for data analysis based on those theorems, focusing on groupoid symmetry-breaking phase transitions characterized by Fisher Zero analogs.

1. Introduction

Although, as Maturana and Varela [1] put it, all organisms—with and without a nervous system—are cognitive, and living per se is a process of cognition, the assertion, as it were, only sings half the Mass. Many, perhaps most, cognitive phenomena, from blood pressure control through gene expression, immune function, consciousness, and driving a fast vehicle at night on a twisting, pot-holed roadway, are inherently unstable and must be heavily regulated to prevent catastrophe. For humans in particular, the 'stream of consciousness' must always be contained within social and cultural as well as more common neuropsychological 'riverbanks' to successfully confront and circumvent powerful selection pressures (e.g., [2]).

Cognition implies choice, choice reduces uncertainty, and the reduction in uncertainty implies the existence of an information source 'dual' to the cognitive process under study [2–6]. The argument is both direct and unambiguous. Here, we will assume such information sources are piecewise adiabatically stationary, so that on a trajectory 'piece', the system is sufficiently close to nonergodic stationary for the appropriate asymptotic limit theorems to work sufficiently well. A useful analog is the Born–Oppenheimer approximation in molecular QM, where nuclear motions are taken as slow enough for electron structures to essentially equilibrate.

Relatively recent work relates the asymptotic limit theorems of information theory to control theory via the Data Rate Theorem (e.g., [7]). Following the linear approximation used in that work, consider a reduced model of a control system as in Figure 1.

An initial n-dimensional vector of system parameters at time t is represented as x_t. The system state at time $t + 1$ is then—near a presumed nonequilibrium steady state—approximated by the first-order relation:

$$x_{t+1} \approx \mathbf{A}x_t + \mathbf{B}u_t + W_t \tag{1}$$

where $\mathbf{A}$ and $\mathbf{B}$ are fixed n-dimensional square matrices, u_t is a vector of control information, and W_t is an n-dimensional vector of Brownian white noise. According to the Data Rate Theorem, if $\mathcal{H}$ is a rate of control information sufficient to stabilize an inherently unstable control system, then it must be greater than a minimum, $\mathcal{H}_0$,

$$\mathcal{H} > \mathcal{H}_0 \equiv \log(\| \det[\mathbf{A}^m]\|) \tag{2}$$

where det is the determinant of the subcomponent $\mathbf{A}^m$ – with $m \leq n$ – of the matrix $\mathbf{A}$ having eigenvalues ≥ 1. $\mathcal{H}_0$ is defined as the rate at which the unstable system generates its own 'topological information'.
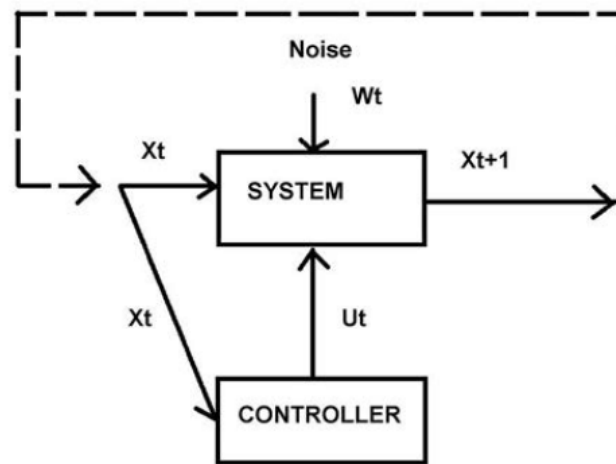

Figure 1. A simplified model of an inherently unstable system stabilized by a control signal U_t. A singular feature of this perspective, as a reviewer has pointed out, is that not only does it separate the car and driver from the road, but it differentiates the road from the car and driver, which may be of considerable importance if one's focus is highway maintenance.

If this inequality is violated, stability fails: blood pressure skyrockets, the immune system eats its host, conscious higher animals suffer hallucinations, and a speeding vehicle goes off the road.

A singular feature of such a control theory perspective, as a reviewer pointed out, is that not only does it separate a car and driver from the road, but it differentiates a road from cars and drivers, which may be of considerable importance if one's central focus is highway maintenance. Further as a reviewer has noted, the channel can be rendered noise-free by importing more information through expanding the environment—the controller—so that the system/controller interaction occurs at the boundary that separates them. This is, however, seldom possible in real-world, real-time interactions.

Here, we will approach these matters from a somewhat unusual direction that permits significant generalization and engineering application, in a large sense. We will derive the Data Rate Theorem from the Rate Distortion Theorem via some counterintuitive but characteristic formal legerdemain.

2. Remembering Brownian Motion

Einstein's treatment of Brownian motion for N particles, via the simple diffusion equation

$$\partial\rho(x,t)/\partial t = \mu\partial^2\rho(x,t)/\partial x^2 \tag{3}$$

where x is distance and μ a diffusion coefficient, produces a solution in terms of the Normal Distribution,

$$\rho(x,t) = \frac{N}{\sqrt{4\pi\mu t}} \exp[-x^2/(4\mu t)] \tag{4}$$

After some manipulation,

$$\sqrt{<x^2>} \propto \sqrt{t} \tag{5}$$

A central observation of Bennett, as represented by Feynman [8], is that a 'simple ideal machine' permits the extraction of free energy from an information source. The Rate Distortion Theorem [9] expresses the minimum channel capacity $R(D)$ needed for a transmitted signal along a noisy channel to be received with average distortion less than or equal to some scalar measure D. A 'worst case', in many respects, is the Gaussian channel for which, under the square distortion measure,

$$R(D) = \frac{1}{2} \log_2(\sigma^2/D), \;\; D \le \sigma^2$$
$$R(D) = 0, \;\; D > \sigma^2 \tag{6}$$

where σ is a 'noise' parameter.

If, following Feynman [8], we identify information as a form of free energy, it becomes possible—*in a purely formal manner*—to construct an 'entropy' S for the system in the standard way of ordinary thermodynamics via the Legendre transform

$$S \equiv -R(D) + D dR/dD \tag{7}$$

and to then impose a first-order nonequilibrium thermodynamics Onsager approximation [10] as the simple diffusion equation

$$dD/dt \approx \mu dS/dD = \mu D d^2 R/dD^2 \tag{8}$$

where 'μ' is taken as a kind of diffusion coefficient. Using Equation (6), we can directly calculate $D(t)$ as

$$dD/dt = \frac{\mu}{2D(t)\log(2)}$$
$$D(t) \propto \sqrt{t} \tag{9}$$

Thus, based on the definition of 'entropy' in Equation (7) as a Legendre transform and the use of the Onsager entropy gradient approximation of Equation (8), $D(t) \propto \sqrt{t}$, *as if the 'target' undergoes simple Brownian motion* in the absence of further data. Parenthetically, as will be shown below in Section 4, this surprising result suggests a direction for generalization to less simple systems, i.e., build an appropriate free energy, define an 'entropy' as its Legendre transform, and then impose the Onsager entropy gradient approximation to derive system dynamics. Taking the next step in the full argument, for a simple deterministic system, we can impose 'control free energy' at a rate $M(\mathcal{H})$, depending, in a monotonic increasing but possibly nonlinear manner, on a control information rate $\mathcal{H}$. Then, for the existence of a nonequilibrium steady state—a kind of stability having, at least, a fixed value of distortion—dD/dt must decline to zero,

$$dD/dt = \mu D d^2 R/dD^2 - M(\mathcal{H}) \le 0$$
$$M(\mathcal{H}) \ge \mu D d^2 R/dD^2 \ge 0$$
$$\mathcal{H} \ge \mathcal{H}_0 \equiv \max\{M^{-1}(\mu D d^2 R/dD^2)\} \tag{10}$$

where max represents the maximum value, and noting that, if M is monotonic increasing, so is the inverse function M^{-1}. By convexity, $d^2 R/dD^2 \ge 0$ [9]. For the Gaussian channel—

in the absence of further noise—at nonequilibrium steady state where $dD/dt \equiv 0$, then $D \propto 1/M(\mathcal{H})$.

Again, an example might be continuous radar or lidar illumination of a moving target with transmitted energy rate M, defining a maximum D inversely proportional to M. If illumination fails, then D will increase as $\sqrt{t}$, representing classic diffusion from the original tracking trajectory.

We have, in the last expression of Equation (10), replicated something of Equation (2), but there is far more structure concealed here.

3. A General Model

Recall again that the Rate Distortion Function $R(D)$ is always convex in D, so that $d^2R/dD^2 \geq 0$ [9,11]. Further, following [11,12], and others, for a nonergodic process, the Rate Distortion Function can be calculated as an average across the RDF of the ergodic components of that process, and can thus be expected to remain convex in D. We reconsider Figure 1 and now examine the 'expected transmission' of a signal $X_t \to \hat{X}_{t+1}$, (in the presence of an added 'noise' Ω), but received as a de facto signal X_{t+1}. That is, there will be some deviation between what is ordered and what is observed, measured under a scalar distortion metric as $d(\hat{X}_{t+1}, X_{t+1})$ and averaged as

$$D = \sum Pr(\hat{X}_{t+1})d(\hat{X}_{t+1}, X_{t+1}) \tag{11}$$

where Pr is the probability of $\hat{X}_{t+1}$ and the 'sum' may represent a generalized integral. We have constructed an adiabatically, piecewise stationary information channel for the control system and can invoke a Rate Distortion Function $R(D)$. Following Equation (10), we can write a general stochastic differential equation [13]

$$dD_t = \left(\mu D_t[d^2R/dD^2]_t - M(\mathcal{H})\right)dt + \Omega h(D_t)dW_t \tag{12}$$

where μ is a 'diffusion coefficient' and $\Omega h(D)$ is the 'volatility' under Brownian noise dW_t. As above, an average $<D>$ can be calculated from the relation

$$< dD_t >= \mu D_t[d^2R/dD^2]_t - M(\mathcal{H}) = 0$$

so that, again, something like Equation (10) can be said to hold. However, stability of stochastic systems is a far richer landscape than for deterministic systems. That is, we are now interested in stability under stochastic volatility, and there are many possible characterizations of it. More specifically, we want to calculate the nonequilibrium steady-state properties of a general function $Q(D_t)$ given Equation (11), i.e., the relation $< dQ_t >= 0$. This can be conducted using the Ito Chain Rule [13], leading to

$$\left(\mu D d^2R/dD^2 - M(\mathcal{H})\right)dQ/dD + \frac{\Omega^2}{2}h(D)^2 d^2Q/dD^2 = 0 \tag{13}$$

Solving for $M(\mathcal{H})$,

$$M(\mathcal{H}) = \mu D\left(\frac{d^2}{dD^2}R(D)\right) + \frac{\Omega^2 h(D)\left(\frac{d^2}{dD^2}Q(D)\right)}{2\left(\frac{d}{dD}Q(D)\right)} \tag{14}$$

giving general results that are not confined to a particular algebraic form for $R(D)$.

As a 'worst case' example, we treat the Gaussian channel of Equation (6) in second order.

Assuming the simple volatility function $h(D) = D$ and taking $Q(D) = D^2$ allows determination of a general condition for stability in variance for the Gaussian channel from Equation (14) as

$$M(\mathcal{H}) = \frac{\Omega^2}{2}D + \frac{1}{D\log(4)} \geq \frac{\Omega}{\sqrt{\log(2)}}$$
$$\mathcal{H} \geq M^{-1}\left(\frac{\Omega}{\sqrt{\log(2)}}\right) \tag{15}$$

where the last inequality in the first expression can be found from a simple minimization argument. Calculations for other kinds of channels are similar.

Another possible approach views the control signal $\mathcal{H}$ itself as the fundamental distortion measure between intended and observed behaviors, so that Equation (11) becomes

$$H \equiv \sum Pr(\hat{X}_{t+1})\mathcal{H}(\hat{X}_{t+1}, X_{t+1}) \tag{16}$$

where $\mathcal{H}$ is the rate of control information needed to stabilize the inherently unstable system and the other factors are as above. We can now carry through the analysis as driven by Equation (12), but with D replaced by H and the rate of control free energy is taken as $M(H)$. This gives the necessary condition of Equation (14), but in terms of H, it provides a different picture of stability dynamics based on the 'average control distortion' H rather than on $\mathcal{H}$ itself, regardless of the typically monotonic increasing nature of $M(H)$.

4. Extending the Perspective

The focus on distortion and channel capacity in our reinterpretation of Figure 1 permits a simplified one-dimensional analysis. We continue in one dimension but branch the argument.

Via the Legendre transform and Onsager approximations of Equations (7) and (8), and through the SDE of Equation (12), it has been possible to derive, in Equations (14) and (16), a fair version of the Data Rate Theorem result of Equation (2). A constraint on the approach is that, because information transmission is not microreversible—for example, in English, the term 'the' has a much higher probability than 'eht', and in Chess, one cannot uncheckmate a King—there can be no 'Onsager reciprocal relations' in multidimensional variants of Equations (8) and (12).

The treatment of multiple parallel processes—branching—in this formalism is of particular interest if somewhat subtle.

Here, we see control dynamics—the set of possible 'Data Rate Theorems'—as highly context-dependent, supposing that different 'road conditions' will require different analogs to, or forms of, the DRT. That is, the transmission of control messages as represented in Equation (11) involves equivalence classes of possible control sequences $X \equiv \{X_t, X_{t+1}, X_{t+2}, \ldots\}$. For example, driving a particular stretch of road slowly on a dry, sunny morning is a different 'game' than driving it at high speed during a midnight snowstorm.

We characterize all possible such 'games' in terms of equivalence classes G_j of the control path sequences X, each class associated with a particular 'game' represented by an information source having uncertainty $H_{G_j} \equiv H_j$.

In this model, $R(D)$, characterizing the particular 'vehicle', plays a different role. Here, we fix the maximum acceptable average distortion and impose the corresponding R as a temperature analog via an iterated model. To do this, we construct a pseudoprobability

$$P_j = \frac{\exp[-H_i/g(R)]}{\sum_j \exp[-H_i/g(R)]} \tag{17}$$

where the 'temperature' $g(R)$ must be calculated from first principles.

The denominator in Equation (17) can be taken as a statistical mechanical partition function to derive a free energy F as

$$\exp[-F/g(R)] = \sum_k \exp[-H_k/g(R)] \equiv A(g(R))$$

$$F(R) = -\log[A(g(R))]g(R) \tag{18}$$

Next, again, it is possible to formally define an 'entropy' in terms of a Legendre transform on F and to again impose a dynamic relation like that of the first-order Onsager treatment of nonequilibrium thermodynamics [10]. Then, in general (dropping the 'diffusion coefficient' μ),

$$S(R) \equiv -F(R) + RdF/dR$$

$$\partial R/\partial t \approx dS/dR = Rd^2F/dR^2 = f(R)$$

$$F(R) = \int \int \frac{f(R)}{R} dRdR + C_1 R + C_2$$

$$g(R) = \frac{-F(R)}{RootOf(\exp[Z] - A(-F(R)/Z))} \tag{19}$$

Several points:

(1) In particular, note the first-order Onsager nonequilibrium thermodynamics approximation in the second expression. This is a very crude model that cannot be expected to have universal applicability. Higher-order and multivariate versions will be of greater, but still limited, use.

(2) $f(R)$ represents the 'friction' inherent to any control system, e.g., $dR/dt = f(R) = \beta - \alpha R$, $R(t) \to \beta/\alpha$, while the RootOf construction generalizes the Lambert W-function, seen by carrying through a calculation setting $A(g(R)) = g(R)$.

(3) As 'RootOf' may have complex number solutions, the temperature analog $g(R)$ now imposes 'Fisher Zeros' analogous to those characterizing the phase transition in physical systems (e.g., [14–16]). Phase transitions in the cognitive process range from the punctuated onset of conscious signal detection to Yerkes–Dodson 'arousal' dynamics. See Wallace [2] for worked-out examples.

(4) The set of equivalence classes G_j defines a groupoid in the classic manner (e.g., [17]), and the Fisher Zero construction represents groupoid symmetry-breaking for cognitive phenomena that is analogous to, but different from, the group symmetry-breaking associated with the phase transition in physical processes.

Groupoid symmetry-breaking phase transitions represent one extension of the basic DRT.

(5) One possible extension of these results is via a 'reaction rate' treatment abducted from chemical physics [18]. The 'reaction rate' L, taking the minimum possible source uncertainty across the 'game' played by the control system as H_0, is then

$$L = \frac{\sum_{H_j \geq H_0} \exp[-H_j/g(R)]}{\sum_k \exp[-H_k/g(R)]} \equiv \mathcal{L}(H_0, g(R)) \tag{20}$$

where the sums may be generalized integrals and the denominator can be recognized as the partition function $A(g(R))$ from Equation (18).

Wallace [2] uses similar models to derive a version of the Yerkes–Dodson 'inverted-U' model for cognition rate vs. arousal.

Again, Fisher Zero analogs in $g(R)$ characterize cognition rate phase transitions.

(6) We have, by careful design, been able to restrict the development to one dimension via focus on the Rate Distortion Function $R(D)$ alone. The extension of the theory to higher dimensions, for example, writing $R(D, \mathbf{Z})$, where $\mathbf{Z}$ is an irreducible vector of

resource rates, is not entirely straightforward [19] (Chapter 5), nor is the extension to more complicated versions of the Onsager models [20].

(7) The argument extends directly to stationary nonergodic systems, where time averages are not ensemble averages, assuming that sequences can be broken into small high-probability sets consonant with underlying forms of grammar and syntax, and a much larger set of low-probability sequences not so consonant [2,19,20]. Equation (17) is then path-by-path, as source uncertainties can still be defined for individual paths [21]. The 'game' equivalence classes still emerge directly, leading to groupoid symmetry-breaking phase transitions [2,22].

Finally, it is possible to explore the stochastic stability of the system of Equation (19) in much the same manner as was performed in the previous section. Here, the driving stochastic relation—analogous to Equation (12)—becomes

$$dR_r = f(R_t)dt + \Omega h(R_t)dW_t \tag{21}$$

where, again, $\Omega h(R)$ is a volatility term in the Brownian noise dW_t.

The application of the Ito Chain Rule to a stochastic stability function $Q(R)$ produces an analog to Equation (13), $< dQ_t >= 0$, as

$$f(R_t)\frac{\mathrm{d}}{\mathrm{d}R_t}Q(R_t) + \frac{\Omega^2(h(R_t))^2 \frac{\mathrm{d}^2}{\mathrm{d}R_t^2}Q(R_t)}{2} = 0 \tag{22}$$

Examining stability in second order and with 'simple' volatility, so that $Q = R^2$ and $h(R) = R$, and assigning an 'exponential' relation as $dR/dt = f(R) = \beta - \alpha R$ gives the variance as

$$Var = \left(\frac{\beta}{\alpha - \Omega/2}\right)^2 - (\beta/\alpha)^2 \tag{23}$$

which explodes as $\Omega^2 \to 2\alpha$.

A different viewpoint permits R in Equation (22) to be unconstrained, thus imposing selection pressure via the possible forms of $f(R)$ and F, as determined by Q and h.

5. Discussion

The 'simple' Rate Distortion Theorem arguments of Sections 2 and 3, driven by the essential convexity of the Rate Distortion Function for stationary systems, lead to very general deterministic and stochastic forms of the Data Rate Theorem. That theorem, usually derived as an extension of the Bode Integral Theorem (e.g., [7], and references therein), places control theory and information theory within the same milieu, one that encompasses much of the cognitive phenomena so uniquely characterizing the living state (e.g., [1]).

More generally, there cannot be cognition—held within the confines of information theory—without a parallel regulation, held within the confines of control theory. Both theories are constrained by powerful—and apparently closely related—asymptotic limit theorems.

Section 4 extends the underlying concept to multiple 'selection pressures' via equivalence classes of control path sequences. This leads to an iterated model in which a partition function, the denominator of Equation (17), serves as the basis for an iterated free energy, its associated Legendre transform entropy, and an iterated Onsager approximation in the gradient of that entropy construct, driving system dynamics. A stochastic stability analysis can be conducted in a standard manner. Ultimately, the extended theory focuses on groupoid symmetry-breaking phase transitions characterized by Fisher Zero analogs.

In sum, we have outlined a novel approach to building new statistical tools for data analysis based on the asymptotic limit theorems of control and information theories. An interested reader should be able to take this material and run with it, recognizing that all such statistical tools—much like Onsager approximations in nonequilibrium thermodynamics—inevitably have limited ranges of reliability and applicability. Such recognition serves

to evade many of the intellectual thickets surrounding contemporary Grand Unifying Theories, e.g., [23–27].

References

1. Maturana, H.; Varela, F. *Autopoiesis and Cognition: The Realization of the Living*; Reidel: Boston, MA, USA, 1980.
2. Wallace, R. *Consciousness, Cognition and Crosstalk: The Evolutionary Exaptation of Nonergodic Groupoid Symmetry-Breaking*; Springer: New York, NY, USA, 2022.
3. Atlan, H.; Cohen, I. Immune information, self-organization, and meaning. *Int. Immunol.* **1998**, *10*, 711–718. [CrossRef] [PubMed]
4. Wallace, R. *Consciousness: A Mathematical Treatment of the Global Neuronal Workspace Model*; Springer: New York, NY, USA, 2005.
5. Wallace, R. Consciousness, crosstalk, and the mereological fallacy: An evolutionary perspective. *Phys. Life Rev.* **2012**, *9*, 426–453. [CrossRef] [PubMed]
6. Wallace, R. On the variety of cognitive temperatures and their symmetry-breaking dynamics. *Acta Biotheor.* **2020**, *68*, 421–439. [CrossRef] [PubMed]
7. Nair, G.; Fagnani, F.; Zampieri, S.; Evans, R. Feedback control under data rate constraints: An overview. *Proc. IEEE* **2007**, *95*, 108–137. [CrossRef]
8. Feynman, R. *Lectures on Computation*; Westview Press: New York, NY, USA, 2000.
9. Cover, T.; Thomas, J. *Elements of Information Theory*, 2nd ed.; Wiley: New York, NY, USA, 2006.
10. de Groot, S.; Mazur, P. *Nonequilibrium Thermodynamics*; Dover: New York, NY, USA, 1984.
11. Effros, M.; Chou, P.; Gray, R. Variable-rate source coding theorems for stationary nonergodic sources. *IEEE Trans. Inf. Theory* **1994**, *40*, 1920–1925. [CrossRef]
12. Shields, P.; Neuhoff, D.; Davisson, L.; Ledrappier, F. The Distortion-Rate function for nonergodic sources. *Ann. Probab.* **1978**, *6*, 138–143. [CrossRef]
13. Protter, P. *Stochastic Integration and Dierential Equations*, 2nd ed.; Springer: New York, NY, USA, 2005.
14. Dolan, B.; Janke, W.; Johnston, D.; Stathakopoulos, M. Thin Fisher zeros. *J. Phys. A* **2001**, *34*, 6211–6223. [CrossRef]
15. Fisher, M. *Lectures in Theoretical Physics*; University of Colorado Press: Boulder, CO, USA, 1965; Volume 7.
16. Ruelle, D. Cluster property of the correlation functions of classical gases. *Rev. Mod. Phys.* **1964**, *36*, 580–584. [CrossRef]
17. Weinstein, A. Groupoids: Unifying internal and external symmetry. *Not. Am. Math. Assoc.* **1996**, *43*, 744–752.
18. Laidler, K. *Chemical Kinetics*, 3rd ed.; Harper and Row: New York, NY, USA, 1987.
19. Wallace, R. *Deception and Delay in Organized Conflict: Essays on the mathematical theory of Maskirovka*; Springer: New York, NY, USA, 2022.
20. Wallace, R. Toward a formal theory of embodied cognition. *BioSystems* **2021**, *202*, 104356. [CrossRef] [PubMed]
21. Khinchin, A. *Mathematical Foundations of Information Theory*; Dover: New York, NY, USA, 1957.
22. Wallace, R. *Essays on Strategy and Public Health: The Reconfiguration of Power Relations*; Springer: New York, NY, USA, 2022.
23. Friston, K. A free energy principle for a particular physics. *arXiv* **2019**, arXiv:1906.10184.
24. Aguilera, M.; Millidge, B.; Tschantz, A.; Buckley, C. How particular is the physics of the free energy principle? *Phys. Life Rev.* **2022**, *40*, 24–50. [CrossRef]
25. Biehl, M.; Pollock, F.; Kanai, R. A technical critique of some parts of the free energy principle. *Entropy* **2021**, *23*, 293. [CrossRef] [PubMed]
26. Colombo, M.; Wright, C. Explanatory pluralism: An unrewarding prediction error for free energy theorists. *Brain Cogn.* **2017**, *112*, 3–12. [CrossRef]
27. Colombo, M.; Wright, C. First principles in the life sciences: The free energy principle, organicism, and mechanism. *Synthese* **2021**, *198* (Suppl. 14), S3463–S3488. [CrossRef]