Failure of Real-Time Multi-Component, Multi-Level Cognitive Systems on Clausewitz Landscapes

Rodrick Wallace
The New York State Psychiatric Institute
Rodrick.Wallace@nyspi.columbia.edu, rodrick.wallace@gmail.com

July 21, 2018

Abstract

Real-time, multi-component cognitive enterprises – institutional, machine, or man/machine hybrid – acting under constraints of time, resources, internal communications, and intelligence across different scales and levels of organization, can be confounded by an adversary with sufficient situational understanding and resources, or by the evolution of appropriate circumstances in response to the system's own actions. Consequently, there can be no free lunch for AI combat systems, driverless cars on intelligent roads, or other such enterprises, tasked with the control of real-time critical processes. Intelligent adversaries, resource constraints, 'design flaws', inattentional blindness, command stupidity, coevolutionary process, and the like, confront all real-time, multi-component cognitive systems. AI systems will not be spared simply by being labeled 'high tech' in marketing campaigns.

1 Introduction

Cognitive systems, from individual organisms, to social colonies of lower animals, up through human social institutions, machine systems, and 'cockpit' man-machine hybrids, are often in conflict or cooperation with other similar agents in the search for, and exploitation of, essential resources. When negotiation and trade – in a large sense that may range from cellular mutualism to international treaties – fail, then cognitive systems are driven to real-time contention on a landscape most singularly marked by analogs to the 'fog-of-war' and 'friction' characterized by such practitioners of the art and science of war as Carl von Clausewitz, John Boyd, Mao Tse-Tung, and Vo Nguyen Giap.

Recent advances in 'deep learning' and 'reinforcement learning' neural networks have revolutionized pattern recognition and permitted machines to greatly

exceed human players in the turn-based 'combat analogs' of Chess and Go, perfect information games for which there is neither fog-of-war nor friction. More recently, genetic fuzzy algorithms have exceeded human capabilities in small-unit air combat simulations, where the fog-of-war is relatively 'thin' and time constraints are not stacked across scale and level of organization (Ernst et al. 2016). Technology company marketing efforts tout these examples and offer a leap-of-faith worthy of Herman Melville's Confidence Man, suggesting that such systems provide 'the answer', ranging across autonomous vehicles through 'cyber combat' and automated first-strike systems. A sophisticated reader may perhaps remember when 'smoking is safe', 'polyvinyl chloride is no worse in a fire than wood', 'fat is worse than sugar', 'antibiotics in animal feed do not cause resistant pathogens', 'fossil fuel is ok and global warming a hoax', and so on.

Here, we examine in some detail the dynamics of real-time, resource-constrained, multi-component, multi-level cognitive systems – of any nature – embedded in a Clausewitz landscape for which fog-of-war and friction are dominant themes.

2 John Boyd's OODA Loop

The US military strategist John Boyd, generalizing from his own experiences of the 'thin, warm' fog-of-war associated with mid-20th century air combat, came to understand the intimate relation between cognition and control that is inherent to contention and conflict. His work developed from a relatively simple sequential dynamic of Observation, Orientation, Decision and Action, to the the elaborate cognition and control model of his later presentations, which Osinga (2007) describes as follows:

> The OODA loop model as [later] presented by Boyd... represents his view on the process of individual and organizational adaptation in general, rather than only the military-specific command and control decision-making process that it is generally understood to depict. It refers to [a] conceptual spiral... to the process of learning, to doctrine development, to command and control processes and to the Popperian/Kuhnian ideas of scientific advance. The (neo-)Darwinists have their place, as do Piaget, Conant, Monod, Polanyi and Hall, while Prigogine and Goodwin are incorporated through Boyd's concluding statement in the final slide that follows [his presentation of] the OODA loop picture:
>
> "The key statements of this presentation, the OODA Loop Sketch and related insights represent an evolving, open-ended, far from equilibrium process of self-organization, emergence and natural selection."
>
> This relates the OODA loop clearly to Complex Adaptive Systems, the role of schemata and to the process of evolution and adaptation. Once again it shows that where the aim is 'to survive and prosper' in a non-linear world dominated by change, novelty and un-

certainty, adaptation is the important overarching theme in Boyd's strategic theory.

Figure 1, from Osinga (2007), and adapted from Boyd's presentation, shows his fully-developed cognitive/control process model.

In the following sections, via the asymptotic limit theorems of information and control theories, we will formally parse Fig.(1) back and forth, first from cognitve, and then from control perspectives, with special emphasis on the vulnerabilities of AI systems via their use of 'anytime' algorithms that permit 'good enough' solutions under time constraint.

The results, for 'thick, cold' fog-of-war, are no more encouraging than the outcomes of US military efforts in Vietnam, Iraq, and Afghanistan.

3 A first model for failure of cognition in real-time conflict

We begin by adapting of an Arrhenius chemical reaction rate relation to model both the rate at which cognition proceeds, and the efficiency of that cognition (Wallace 2016). One central point is to treat the rate at which some index of essential resources, Z, is delivered to the system of interest as a temperature analog. Details can be found in the reference. Another essential idea is, following Feynman (2000), that the Rate Distortion Function R of the inherent cognitive channel connecting a cognitive system's intent with its actual effect (e.g., Cover and Thomas 2006) is a 'free energy' measure, permitting a Boltzmann-like probability argument leading to a normalized, dimensionless, rate of cognition

$$P[R \geq R_0] = \frac{\int_{R_0}^{\infty} \exp[-R/\kappa Z] dR}{\int_0^{\infty} \exp[-R/\kappa Z] dR} = \exp[-R_0/\kappa Z] \tag{1}$$

where κ is a local constant and R_0 is the threshold channel capacity necessary for initiation of a cognitive function, for recognition that a 'signal', in a large sense, must elicit a response. In real systems this value will never be zero, and in global broadcast settings (Wallace 2017) will involve sharp punctuation similar to the onset of consciousness.

The effectiveness – rate of response – and efficiency of a cognitive function indexed by j can thus be respectively measured as

$$\begin{aligned} F_j(Z_j) &= \exp[-K_j/Z_j] \\ f_j(Z_j) &= \frac{\exp[-K_j/Z_j]}{Z_j} \end{aligned} \tag{2}$$

for an associated local 'detection level' parameter K_j. See figure 2. At an 'infinite' rate of resource supply, effectiveness tops out at 1, in this model. Efficiency, however, reaches a maximum at $Z = K$, and declines monotonically thereafter, implying early onset of draconian cost-benefit constraints.

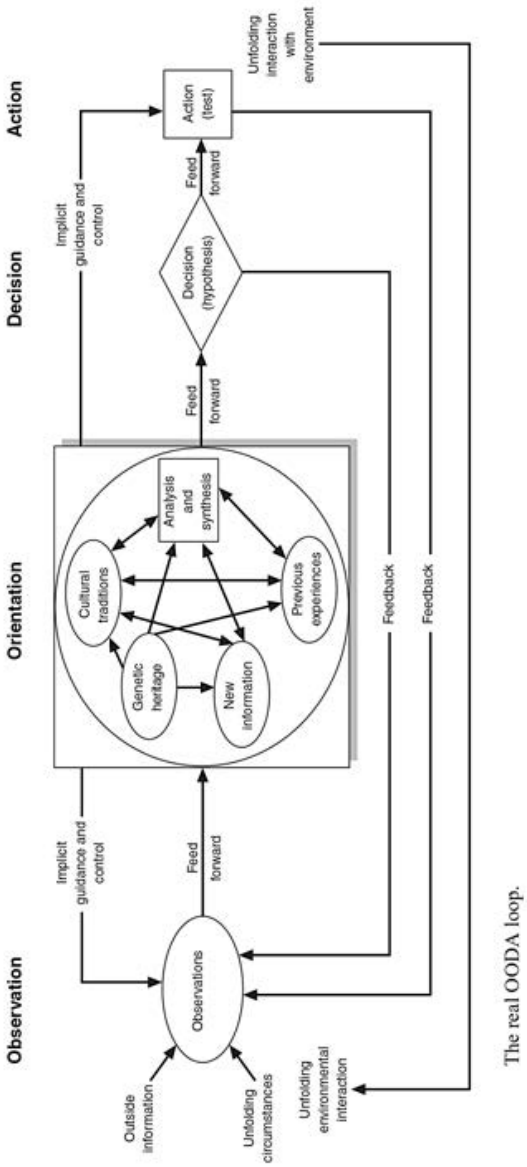

Figure 1: Adapted from Osinga (2007). John Boyd's fully-developed OODA loop convolutes cognition with control, under the conditions of uncertainty inherent to conflict on a Clausewitz landscape strongly dominated by friction and the fog-of-war.

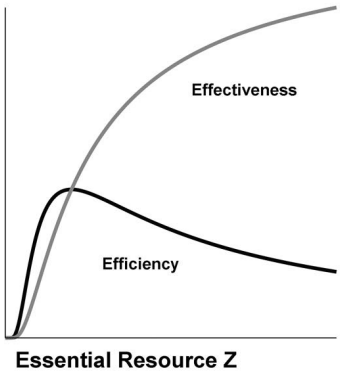

Figure 2: Effectiveness, defined as the rate of system response, and efficiency of a cognitive system, as functions of the rate of availability of an essential resource Z. While 'effectiveness' does indeed rise with increasing resource rate, to a maximum of 1, in this model, efficiency tops out at $Z = K$ and declines thereafter, suggesting onset of serious cost-benefit constraints.

What is Z?. Some $j = 1, ..., n$ cognitive modules must work together within an institution, a machine system, or the usual man/machine composite, a 'cockpit' of some size. These crosstalking-modules must be provided with material resources at some rate M_j. In addition, they must communicate with each other, reflected at each module by some 'bandwidth' C_j representing the channel capacity available to the module. Further, the embedding environment, represented here as an 'external' information source H, must impose signals on each module at some rate H_j. Then Eq.(2) becomes

$$f_j(M_j, C_j, H_j) = \frac{\exp[-\frac{K_j}{M_j C_j H_j}]}{M_j C_j H_j} \tag{3}$$

Across the institution/machine/cockpit there will be influential resource constraints:

$$M = \sum_{j=1}^{n} M_j, \; C = \sum_{j=1}^{n} C_j, \; H = \sum_{j=1}^{n} H_j \tag{4}$$

Assume n cognitive modules and m constraints.

Given some set of cognitive modules, we impose a linear scalarization of a multi-objective optimization problem using unit weighting (Hwang and Massud 1979), and apply a standard Lagrangian optimization as

$$L = \sum_{j=1}^{n} f_j(\Pi_{k=1}^{m} x_k^i) - \sum_{k=1}^{m} \lambda_k (X_k - \sum_{j}^{n} x_j^i) \tag{5}$$

where the x are the basic variates, the X their limiting sums, and the λ are dual 'cost variates' usually treated as 'undetermined multipliers'. In Eqs.(3) and (4), $m = 3$. For a Western 'tactical-operational-strategic' culture-bound model, $n = 3$.

Algebraic regularities permit a projection down onto a single variate, here taken as the rate of communication between modules. After some manipulation, a symmetry emerges as

$$x_p^j = \frac{\lambda_p}{\lambda_q} x_q^j = \frac{X_p}{X_q} x_q^j \tag{6}$$

Then

$$f_i(\Pi_{j=1}^m x_j^i) = f_i(\frac{[x_q^i]^m}{X_q^{m-1}} \Pi_{p \neq q} X_p) \tag{7}$$

We have, effectively, recapitulated Puget's (2005) protocol for using symmetries of an optimization problem to collapse it down to an irreducible set of solutions. We can do this here because $f(\Pi_{j=1}^m x_j)$ is symmetric in all elements of the appropriate permutation group, so the solution can be projected down onto a 'unit group' with a single element. There is, as a result, only a single independent solution.

For $p \neq q$ we next normalize to $X_p \equiv 1$, and examine a cognitive module and its associated apparatus, taking each as having the same threshold K in Eq.(3). Assuming different K_j will explode the algebra without, for our limited purposes here, measurably increasing the insight.

Focusing on the the rate of communication between crosstalking cognitive modules within the overall system gives

$$f(C_i) = \frac{\exp[-\frac{K}{C_i^3}]}{C_i^3} \tag{8}$$

For a three-level cultural model, the expression of interest then becomes

$$\frac{\exp[-\frac{K}{C_1^3}]}{C_1^3} + \frac{\exp[-\frac{K}{C_2^3}]}{C_2^3} + \frac{\exp[-\frac{K}{(1-C_1-C_2)^3}]}{(1-C_1-C_2)^3} \tag{9}$$

Different thresholds for detection, different values of K, generate the sequence of surfaces shown in figure 3. C_1 and C_2 here represent, respectively, the 'channel capacities' for communication of the 'tactical' and 'operational' levels. Recall that the values of K represent the detection thresholds for signals sent or received: *only a small range of values for the threshold variable K will not be highly pathological* in this model.

For the relatively small non-pathological range of the detection threshold K, it is easy to see that maximum efficiency is at the peak represented by $C_1 = C_2 = C_3 = 1/3$, given this particular realization of the model. That is, here, under the relatively small range of excitation thresholds leading to non-pathological dynamics, efficiency peaks when communication bandwidth is

shared equally across 'tactical', 'operational', and 'strategic' levels of organization that are characteristic of Western cultural artifacts, organizations, machines, or composite 'cockpit' entities. Different values of the detection thresholds for K_i would alter this point of greatest efficiency.

Scalarization based on the synergism of rate $\times$ efficiency simply doubles the value of the K_i.

There are other implications.

Defining the reduced Lagrangian as

$$L = \sum_i f(C_i) + \lambda(1 - \sum_i C_i) \tag{10}$$

the conditions for maximization becomes

$$\begin{aligned} \partial f/\partial C_i - \lambda &= 0 \\ \sum_i C_i &= 1 \end{aligned} \tag{11}$$

providing a system of equations that can be explicitly solved.

The relation $\partial L/\partial C_i = \lambda$ is of particular interest. In a physical system, λ represents an inverse temperature (e.g., Schrodinger 1989, Section 2), and in an economic model, a price. For Eq.(8),

$$-3\,\frac{{C_i}^3 - K}{{C_i}^7}\mathrm{e}^{-\frac{K}{{C_i}^3}} = \lambda \tag{12}$$

Figure 4 shows this result as a function of some C_i, for the three-level model.

As Jin et al. (2008) note, pathological values of λ – the 'cost' or 'inverse temperature' – imposed by external circumstance can obviate even the non-pathological parts of the sequence shown in figure 3. That is, negative λ represents a resource drain that cannot be indefinitely sustained, while the demands represented by λ outside the response range available to the system simply cannot be met in any optimal manner.

Recall the scalarizations of Eq.(2), where now

$$Z \equiv \Pi_i z_i \tag{13}$$

so that Z_i is a symmetric product over the set of essential resources. Recall also that $0 \leq F \leq 1$ and $0 \leq f \leq \exp[-1]/\mathcal{K}$, having its maximum at $Z = \mathcal{K}$.

We assume, as the result of strategic decisions, that Z undergoes withdrawal, escalation, or stability at some rate $dZ/dt = \Omega(Z, t)$. Thus Ω may be negative, positive, or zero, depending on command judgment, at a presumably fixed threshold $\mathcal{K}$ that is characteristic of the system.

A real-world enterprise, however, is subject to the effects of 'noise', broadly interpreted, and, consequently, a stochastic differential equation (Protter 2005)

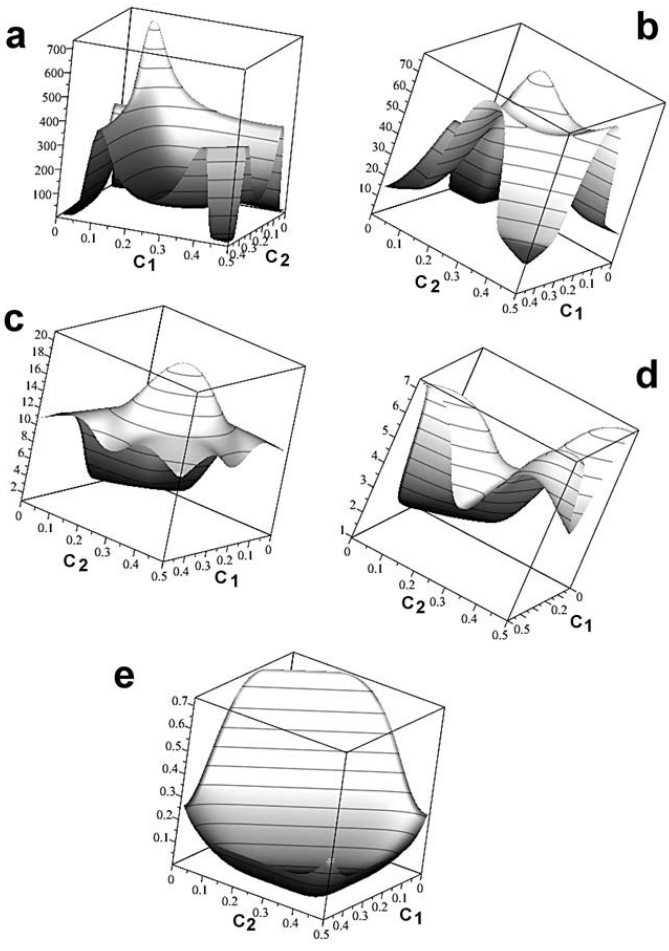

Figure 3: Efficiency surfaces for Eq.(9). C_1 and C_2 are, respectively, internal communication channel capacities available to the 'tactical' and 'operational' levels, so that $C_3 = 1-C_1-C_2$. For figures 2a to 2e, the detection thresholds are, respectively, $K = 0.001$, 0.01, 0.05, 0.1, 0.5. For the non-pathological modes b and c, it is easy to see that peak efficiency occurs when $C_1 = C_2 = C_3 = 1/3$, for this particular system in which all activation thresholds are the same. That is, efficiency peaks when there is equal flow of communication across the three levels of organization. Differing activation thresholds for the three components would impose a different optimal pattern for information transfer.

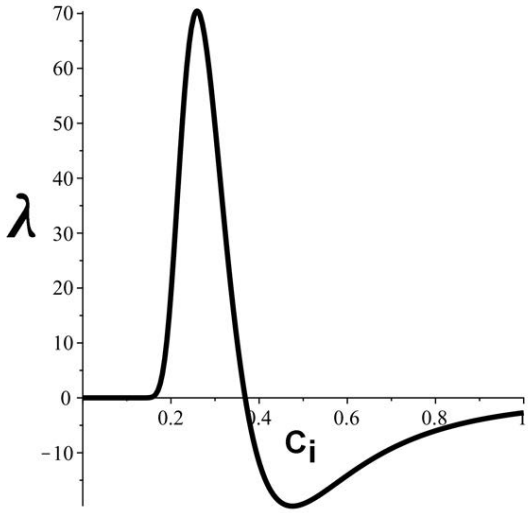

Figure 4: 'Inverse temperature' or 'price' for the model c of figure 3. If λ is not in the range of the figure, no optimization is possible. Negative λ represents an unstable 'excited state' in a physical model, or 'loss' in an economic model.

for the time dependence of the function f can be written as

$$df_t = (df/dZ \times dZ/dt)dt + \sigma f_t dW_t =$$
$$\frac{\exp[-\mathcal{K}/Z]}{Z}[(\frac{\mathcal{K}}{Z^2} - \frac{1}{Z})\Omega(Z,t)dt + \sigma dW_t] =$$
$$f_t[(\frac{\mathcal{K}}{Z^2} - \frac{1}{Z})\Omega(Z,t)dt] + \sigma f_t dW_t = f_t dY_t^f \tag{14}$$

and similarly for F, so that

$$dF_t = F_t[\frac{\mathcal{K}}{Z^2}\Omega(Z,t) + \sigma dW_t] = F_t dY_t^F \tag{15}$$

where dW_t is taken here as Brownian noise, the Y_t^X are stochastic processes, and we neglect matters of resource limitations.

The relations clearly have an unstable nonequilibrium steady states at $f(0) = 0$ and $F(0) = 0$.

The final forms of Eqs.(14) and (15) are precisely those required of the Doleans-Dade exponential, leading to extension of the model under non-Brownian 'colored noise' (Protter 2005). The solution can be formally written as

$$\mathcal{E}(f) \propto \exp(Y_t - \frac{1}{2}[Y_t, Y_t]),\ t \geq 0 \tag{16}$$

where the convergence is in probability and $[Y_t, Y_t]$ is the quadratic variation (Protter 2005) of the process Y_t, necessarily ≥ 0, that grows monotonically in time. For Brownian white noise this is just $\sigma^2 t$.

Heuristically, by the Mean Value Theorem, if $dY/dt < (1/2)d[Y,Y]/dt$, the system declines exponentially in time to zero. This is a counterintuitive result that can be greatly extended.

We have, in Eqs.(3) and (13), leading to the expressions of Eqs.(14) and (15), imposed a particular functional form for the effectiveness, efficiency, and the stochastic dynamics, of a cognitive enterprise facing adversarial challenges.

More comprehensive results, however, are available, via the work of Appleby et al. (2008).

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

$$d\mathcal{F}_t = g(\mathcal{F}_t)dt + h(\mathcal{F}_t)dW_t \tag{17}$$

subject to the Lipschitz condition that, for a positive constant C, and for all $\mathcal{F}(Z_2)$ sufficiently near to some fixed point $\mathcal{F}(Z_1)$,

$$|g(\mathcal{F}(Z_1)) - g(\mathcal{F}(Z_2))| \leq C|\mathcal{F}(Z_1) - \mathcal{F}(Z_2)| \tag{18}$$

We further assume that $\mathcal{F}(0) = 0$, 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., drives $\mathcal{F}(Z,t)$ from some initial positive value down to zero. This generalizes the Doleans-Dade exponential of Eq.(16) to a much broader class of functions. That is, sufficient 'noise' of a proper form can collapse a very large set of cognitive effectiveness and efficiency measures to zero.

Under protracted evolutionary selection pressure – extended, repeated, interludes of conflict – a sophisticated and persistent adversary will either randomly explore, or cognitively construct, a proper function h.

Again, constraints on Z will lead to stochastic optimization models.

4 A second model for failure of cognition in real-time conflict

Following closely Wallace (2018b), another possible approach to institutional cognition under conflict involves a network of lower level information sources 'dual' to individual cognitive modules whose probability of interaction is measured by the strength of the crosstalk between them. Cognition implies choice, choice involves reduction of uncertainty, and reduction of uncertainty implies the existence of an information source defined by the cognitive process under study. The argument is direct (e.g., Wallace 2012, 2017).

For a random network of cognitive modules, it is well known that a 'giant component' involving most of the network nodes emerges in a punctuated manner once the strength of interaction exceeds a well-defined threshold. By contrast, a linked star-of-stars of stars configuration has a zero threshold: everything is connected. Network topology counts.

A random network analysis assumes a variable average number of fixed-strength linkages between the information sources composing the network nodes. Mutual information, by contrast, can continuously vary in magnitude, suggesting the need for a parameterized renormalization. The modular network structure linked by crosstalk has a topology depending on the degree of interaction. The central argument, following Wallace (2012) closely, is as follows. Define an interaction parameter, say ω, a real positive number, and look at geometric structures defined by linkages set to zero if crosstalk mutual information is less than that value, and set to unity if greater than ω. A given ω will define a regime of giant components of network elements – cognitive submodules – linked by mutual information greater than or equal to it.

The argument can be inverted. Some given topology for the giant component will be define a critical value ω_C such that network elements interacting by crosstalk mutual information less than this value will be unable to participate, that is, will be locked out and not 'consciously' perceived on the time scale of the shifting 'spotlight' global broadcast of interest. As a result, ω is a tunable, syntactically dependent, detection limit that depends on the instantaneous topology of the corresponding shifting, tunable, giant component of linked cognitive modules defining the dynamic global broadcasts required by the no-free-lunch constraints. Wallace (2012) shows how this argument can be extended to a multiple set of simultaneous global broadcasts, characterized by a vector $\Omega = (\omega_1, ..., \omega_m)$.

All global broadcast systems, because of their necessary dynamic flexibility, are inherently unstable in the control theory sense and require fairly draconian regulation, as constrained by the Data Rate Theorem (Wallace 2017, 2018a).

An almost-classic phase transition model for such structures emerges if it is possible to identify equivalence classes of a system's developmental pathways, for example 'functional' and 'doomed'. This allows definition of a symmetry groupoid for the developmental process (Wallace 2017; Weinstein 1996; Golubitsky and Stewart 2006). A groupoid is a generalization of an algebraic group in which a product is not necessarily defined between each element. The simplest example is, perhaps, a disjoint union of separate groups, but sets of equivalence classes also define a groupoid. See Wallace (2012, 2017) or Weinstein (1996) for details.

A 'free energy' can be defined that is liable to an analog of Landau's classical spontaneous symmetry breaking, in the Morse Theory sense (Pettini 2007; Matsumoto 2002). Under symmetry breaking, higher 'temperatures' are associated with more symmetric higher energy states in physical systems. Cosmology makes much of such matters in the first moments after the 'big bang': different physical phenomena break out as the universe cools. Here, for the control of cognitive processes, decline in $|\Omega|$ can result in sharply punctuated collapse from higher to lower symmetry states, resulting in serious cognitive failure.

This is worked out as follows.

Given a 'dual' information source associated with the inherently unstable cognitive system of interest, an equivalence class algebra can be constructed by choosing different system origin states and defining the equivalence of subse-

quent states at a later time by the existence of a high probability path connecting them to the same origin state. Disjoint partition by equivalence class, analogous to orbit equivalence classes in dynamical systems, defines a symmetry groupoid associated with the cognitive process (Wallace 2017).

The equivalence classes across possible origin states define a set of information sources dual to different cognitive states available to the inherently unstable cognitive system or coalition of systems of interest. These create a large groupoid, with each orbit corresponding to an elementary 'transitive' groupoid whose disjoint union is the full groupoid. Each subgroupoid is associated with its own dual information source, and larger groupoids must have richer dual information sources than smaller.

Take X_{G_i} as the system's dual information source associated with groupoid element G_i. A Morse Function can now be built using the temperature analog $|\Omega|$.

Take $H(X_{G_i}) \equiv H_{G_i}$ as the Shannon uncertainty of the information source associated with the groupoid element G_i. It is possible to define a Boltzmann-like pseudoprobability as

$$P[H_{G_i}] \equiv \frac{\exp[-H_{G_i}/|\Omega|]}{\sum_j \exp[-H_{G_j}/|\Omega|]} \tag{19}$$

The sum is over the different possible cognitive modes of the full system.

The 'free energy' Morse Function $\hat{F}$ is then defined as

$$\exp[-\hat{F}/|\Omega|] \equiv \sum_j \exp[-H_{G_j}/|\Omega|]$$
$$\hat{F} = -|\Omega| \log[\sum_j \exp[-H_{G_j}/|\Omega|]] \tag{20}$$

Given the underlying groupoid generalized symmetries associated with high-order cognition, as opposed to simple control theory, it is possible to apply a version of Landau's symmetry-breaking approach to phase transition (Pettini 2007). The shift between such symmetries should remain highly punctuated in the 'temperature' $|\Omega|$, but in the context of what are likely to be far more complicated groupoid rather than group symmetries.

Arguing by abduction from physical theory, there should be only a few possible phases, with sharp and sudden transitions between them as $|\Omega|$ decreases.

5 Failure of control in real-time conflict

A different approach to these problems is via recent developments linking control and information theories. The essential insight is that cognitive systems under fog-of-war conditions are inherently unstable, much as a vehicle driving at night on a twisting, pot-holed roadway. The vehicle requires, in addition to a competent driver, both very good headlights and highly responsive steering.

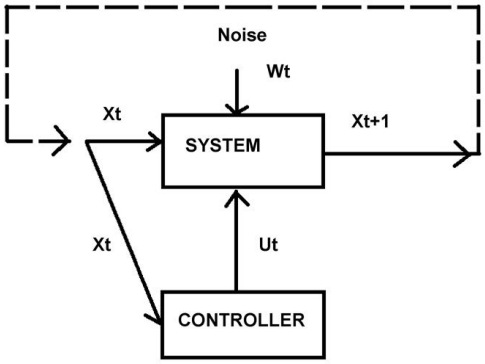

Figure 5: Model of a control system under the constraints of the Data Rate Theorem. Compare with Boyd's model of Fig.(1).

The Data Rate Theorem (DRT) (Nair et al. 2007) establishes the minimum rate at which externally-supplied control information must be provided for an inherently unstable system to remain stable.

The first approximation assumes a linear expansion near some nonequilibrium steady state, so that an n-dimensional vector of system parameters at time t, x_t, determines the state at time $t+1$ according to the expression

$$x_{t+1} = \mathbf{A}x_t + \mathbf{B}u_t + W_t \tag{21}$$

See Fig.(5).

$\mathbf{A}, \mathbf{B}$ are taken as fixed $n \times n$ matrices. u_t is the vector of control information, and W_t is an n-dimensional vector of Brownian 'white noise'. The DRT then states that the minimum rate at which control information $\mathcal{H}$ must be provided for system stability is

$$\mathcal{H} > \log[|\det[\mathbf{A}^m]|] \equiv a_0 \tag{22}$$

Assuming $m \leq n$, $\mathbf{A}^m$ is the subcomponent of $\mathbf{A}$ having eigenvalues ≥ 1. The right hand side of Eq.(22) is then to be taken as the rate at which the system generates 'topological' information.

Stability is lost if the inequality of Eq.(22) is violated. For a night-driving example, if the headlights go out, a twisting roadway cannot be navigated. A different approach to the DRT uses the Rate Distortion Function that will vary according to the nature of the control channel (Wallace 2017). The central argument revolves around the inherent convexity of all Rate Distortion Functions.

Eqs. (21) and (22) seem to instantiate something like the insight of the US military strategist John Boyd regarding a necessary continuous cycle of interaction with the environment, assessing and responding to its constant changes. Boyd's main assertion is that victory in short-duration combat is assured by forcing circumstances to change more rapidly than an adversary can respond.

This triggers fatal destabilization by making the rate at which topological information is generated greater than the rate at which the adversary can counter with useful control information.

While no cognitive system is immune to such attack, as Wallace (2018a, Ch. 6) argues, there are 'Pyrrhic Victory' dynamics operating on longer scales: An adversary can lose every battle and still win the war.

Clausewitz's 'fog-of-war' refers to the inevitability of limited intelligence regarding battlefield conditions, and his 'friction' to the inevitable difficulty of imposing control, due to weather, terrain, time lags, attrition, difficulty in resupply and logistics, and so on. For night driving on bad roads, this might involve a synergism between poor headlights and unresponsive steering.

Real-time cognitive systems facing conflict will have many such constraints acting simultaneously and synergistically. We thus suppose a nonsymmetric $n \times n$ correlation-like matrix ρ having elements $\rho_{i,j}$ representing those constraints and their pattern of interaction. Such matrices have n 'invariants', $r_i, i = 1..n$, that remain fixed when principal component transformations similar to rotation are applied to data, and we construct an invariant scalar measure based on the standard polynomial relation

$$p(\gamma) = \det(\rho - \gamma I) = \gamma^n + r_1\gamma^{n-1} + ...r_{n-1}\gamma + r_n \tag{23}$$

det is the determinant, γ a parameter, and I is the $n \times n$ identity matrix. The first invariant is the matrix trace, and the last is $\pm$ the determinant. Based on the n invariants, we can then define an appropriate composite scalar index $\Gamma = \Gamma(r_1, ..., r_n)$ as a monotonic increasing real function, a downward projection analogous to the Rate Distortion Manifold of Glazebrook and Wallace (2009) or to the Generalized Retina of Wallace and Wallace (2016).

We make a heuristic extension of the condition of Eq.(22) as

$$\mathcal{H}(\Gamma) > f(\Gamma)a_0 \tag{24}$$

A Black-Scholes approximation (Wallace 2017 Section 7.10) finds that, in first order, $\mathcal{H}(\Gamma)$ has a typical linear form, $\mathcal{H} \approx \kappa_1\Gamma + \kappa_2$. Expanding $f(\Gamma)$ to similar first order, so that $f(\Gamma) = \kappa_3\Gamma + \kappa_4$, the limit condition becomes

$$\mathcal{T} \equiv \frac{\kappa_1\Gamma + \kappa_2}{\kappa_3\Gamma + \kappa_4} > a_0 \tag{25}$$

where we will call $\mathcal{T}$ the 'Clausewitz temperature' of the system. For $\Gamma = 0$ the stability condition is $\kappa_2/\kappa_4 > a_0$. At large Γ this becomes $\kappa_1/\kappa_3 > a_0$. If $\kappa_2/\kappa_4 \gg \kappa_1/\kappa_3$, the stability condition may be violated at high Γ. See figure 6.

The results of Jin et al. (2008), and of the 'cognition' section above, suggest that this relation might well be rewritten using $|\lambda|$ from Eq.(12) and figure 4, or $|\Omega|$ from Eq.(19), to replace Γ, connecting the two strands of argument:

$$\begin{aligned} \mathcal{T} &= \frac{\kappa_1|\lambda| + \kappa_2}{\kappa_3|\lambda| + \kappa_4} > a_0 \\ \mathcal{T} &= \frac{\kappa_1(1/|\Omega|) + \kappa_2}{\kappa_3(1/|\Omega|) + \kappa_4} > a_0 \end{aligned} \tag{26}$$

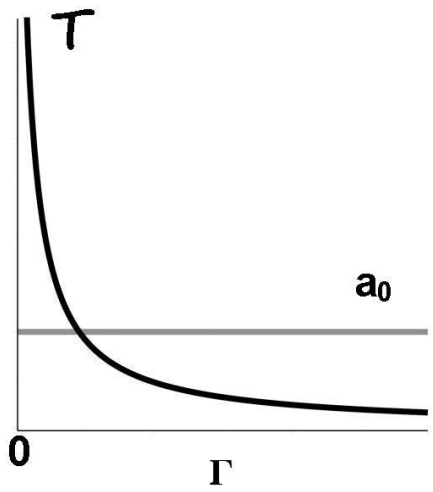

Figure 6: The horizontal line is the DRT limit a_0. If $\kappa_2/\kappa_4 \gg \kappa_1/\kappa_2$, at an intermediate value of Γ the Clausewitz temperature $\mathcal{T}$ falls below criticality, and control must fail.

6 Failure of 'anytime' algorithms in real-time conflict

'Anytime' algorithms, in the words of Zilberstein (1996)

> ...give intelligent systems the capability to trade deliberation time for quality of results. This capability is essential for successful operation in domains such as signal interpretation, real-time diagnosis and repair, and mobile robot control. What characterizes these domains is that it is not feasible (computationally) or desirable (economically) to compute the optimal answer...

Here, the 'essential resource' as studied in Section 3 is time-of-computation, under constraint.

In general, however, anytime algorithms are not necessarily stable, and sufficient conditions for stability must be established on a case-by-case basis, sometimes using draconian stochastic Lyapunov function methods that are far from transparent but are essentially similar to recent results regarding the stabilization/destabilization of stochastic differential equations (e.g., Quevedo and Gupta 2012; Dang et al. 2018; Appleby et al. 2008).

And thereupon hangs a tale.

We consider a hierarchical, sequential command system under which there is a constraint on the total time available for computation across individual components j, that is, $\sum_i t_i = T$.

Following Zilberstein (1996), the 'quality' of the calculation approaches an asymptotic limit with time, as in Fig.(7), albeit with some considerable variance.

Ashok and Patra (2010) examine a similar problem over 100, 200 and 250 cities and fit their results to an Arrhenius function, as in Section 3 of this work. See Fig.(8).

We first examine the Arrhenius model, taking $F = \beta \exp[-\alpha/t]$, so that $dF/dt = (\beta\alpha/t^2)\exp[-\alpha/t] = (\alpha/t^2)F$.

The Lagrangian becomes

$$L = \sum_i \beta_i \exp[-\alpha_i/t_i] + \lambda(T - \sum_i t_i) \tag{27}$$

The optimization relation is

$$\frac{\beta_i \alpha_i}{t_i^2} \exp[-\alpha_i/t_i] = \lambda \tag{28}$$

This is shown in Fig.(9), and is similar to Fig.(4). The necessary conditions for stability are $\lambda \geq 0$, $\lambda \leq 0.54134...\beta/\alpha$. That is, the 'cost' or 'inverse temperature' parameter λ is seen as imposed by the embedding conflict mechanisms. Following Jin et al. (2008), if the imposed 'costs' are beyond the ability of the system to respond – if the value of λ is outside the range of the figure – no solution is possible, and the system fails under load. This is essentially another version of the John Boyd mechanism of 'getting inside the command loop' of an opponent.

The stochastic version of the Arrhenius model is then

$$dF_t = F_t[\frac{\alpha}{t^2}dt + \sigma dW_t] \rightarrow \sigma F_t dW_t \tag{29}$$

Clearly, this model, or other versions without volatility, *in the presence of noise*, converge to some diffusion process. Indeed, for this model, application of the Ito chain rule to F^2 gives

$$Var(F) \rightarrow \exp[\sigma^2 t] \tag{30}$$

The expression for $\mathcal{E}(F)$, via application of the Ito chain rule to $\log(F)$, rises to a peak at $t = \sqrt{2\alpha}/\sigma$, but declines thereafter as $\exp[-\sigma^2 t]$. In a first iteration, then, for the Arrhenius model, the maximum possible value of F under conditions of noise is no longer β, but rather $\beta \exp[-\sigma\sqrt{\alpha/2}]$, rapidly declining toward zero.

Only for $\sigma = 0$ does an Arrhenius anytime model formally and fully converge, taking this perspective, although the result for $\mathcal{E}(F)$ suggests that, *at low noise levels*, useful results may still be found.

This is not a very satisfactory result, and we study other possible models, next examining an 'exponential' relation, having the form $F = (\beta/\alpha)(1 - \exp[-\alpha t])$ so that $dF/dt = \beta \exp[-\alpha t]$.

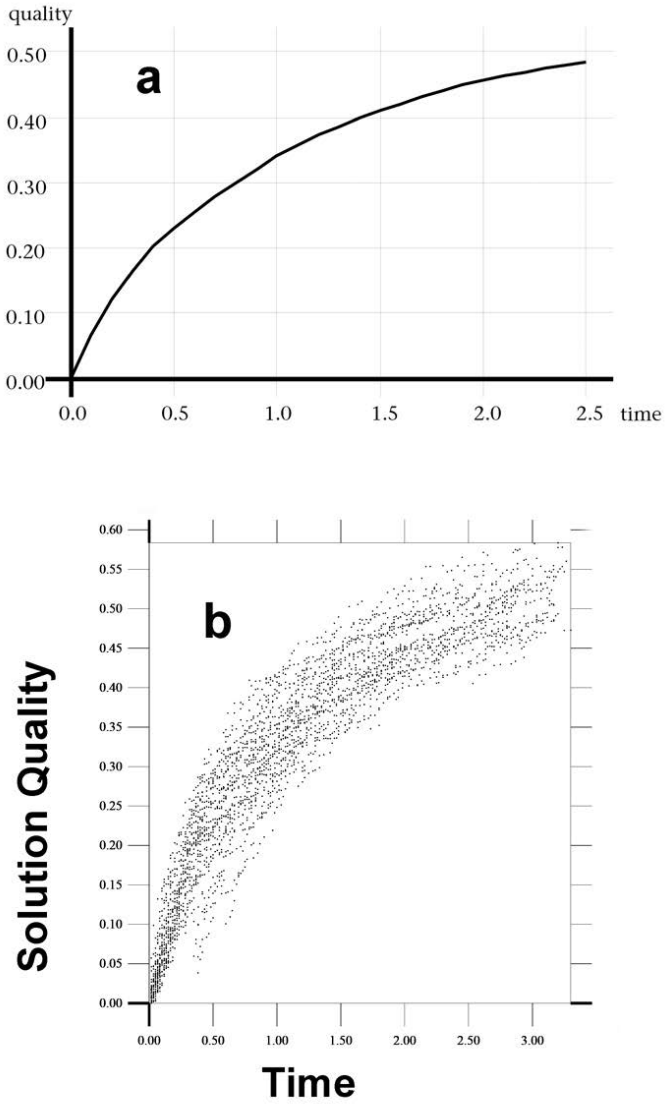

Figure 7: From Zilberstein (1996). (a). Generic form of a quality-of-calculation result as a function of time under a generalized anytime algorithm. (b). Empirical realizations of quality measures at time t of a randomized tour-improvement algorithm for the traveling salesman problem over 50 cities. Under what conditions does the variance explode?

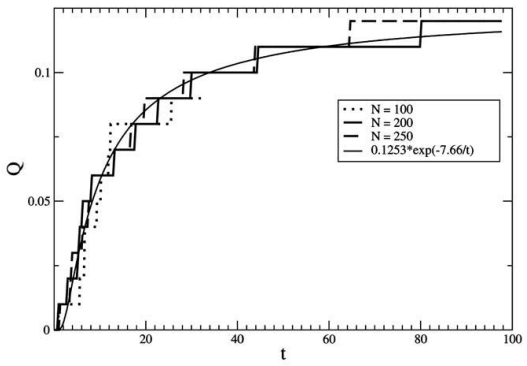

Figure 8: From Ashok and Patra (2010). Quality measure for their anytime algorithm for the cities problem, N=100, 200, 250. The fitted line is an Arrhenius function.

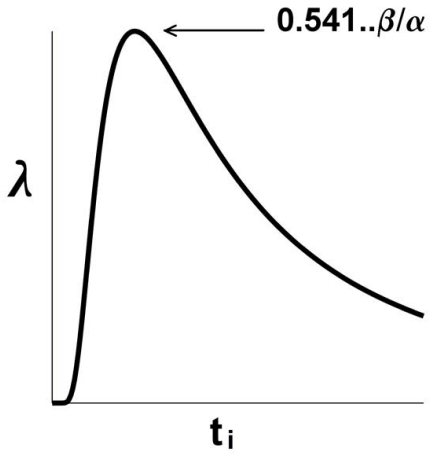

Figure 9: Inverse temperature/price for an Arrhenius anytime algorithm following Eq.(28). If λ is beyond the range of the figure, no solution is possible.

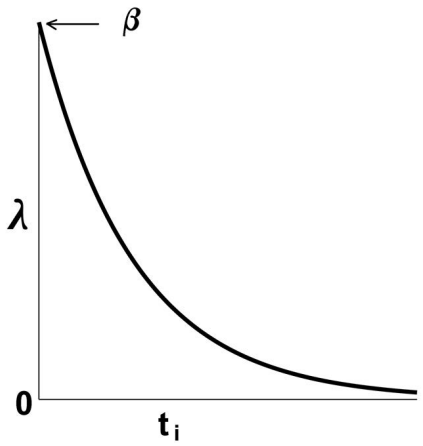

Figure 10: 'Inverse temperature' or 'price' for the exponential anytime algorithm model of Eq.(31). Here, the value of λ is being imposed by the embedding circumstances of conflict. However, if $\lambda > \beta$ or $\lambda < 0$, no solution to the optimization problem is possible, and the system fails. Again, negative values of λ represent 'excited states' or 'losses' in physical or economic models.

The corresponding Lagrangian is

$$L = \sum_i \frac{\beta_i}{\alpha_i}(1 - \exp[-\alpha_i t_i]) + \lambda(T - \sum_i t_i) \tag{31}$$

The ratio β_i/α_i determines the 'topping out' level, while α_i determines the rate at which that level is approached, under a constraint for the total time allowed.

The conditions for optimization become

$$\beta_i \exp[-\alpha_i t_i] = \lambda \tag{32}$$

This relation is shown in Fig.(10). The conditions at $t = 0$ and $t \to \infty$ imply that the necessary conditions for system stability are $\lambda \leq \beta$ and $\lambda \geq 0$.

Again, a stochastic approach is possible. Taking $dF/dt = \beta \exp[-\alpha t]$, the SDE model is

$$dF_t = (\beta - \alpha F_t)dt + \sigma F_t dW_t \tag{33}$$

where, again, dW_t is assumed to be Brownian noise.

Application of the Ito Chain Rule to F^2 allows a variance calculation as

$$Var(F) = \left\{ \frac{\beta}{\alpha - \sigma^2/2}(1 - \exp[(\sigma^2/2 - \alpha)t]) \right\}^2 - \frac{\beta^2}{\alpha^2}(1 - \exp[-\alpha t])^2$$

$$if\ \alpha > \sigma^2/2$$
$$\to \infty\ otherwise \tag{34}$$

Somewhat like the Doleans-Dade results of Eq.(16), the variance converges only if $\alpha > \sigma^2/2$. That is, while the algorithm works at low noise, above a critical value it becomes explosively unstable, in this model. Recalling Appleby et al. (2008), explosive variability, in the context of the 'topping out' of the effectiveness measure, implies that zero can become an absorbing state.

We envision λ as an externally-imposed constraint, and not merely an 'undetermined multiplier', and that t_i in Fig.(10) is truncated on the right by that imposed value. Thus, in Eq.(34), $t = \log(\beta/\lambda)/\alpha$ and we have the variance relation

$$Var(F) =$$
$$\frac{\beta^2}{\left(\alpha - 1/2\,\sigma^2\right)^2}\left(1 - \mathrm{e}^{\frac{1/2\,\sigma^2-\alpha}{\alpha}\ln\left(\frac{\beta}{\lambda}\right)}\right)^2 - \frac{\beta^2}{\alpha^2}\left(1 - \frac{\lambda}{\beta}\right)^2 \tag{35}$$

which again explodes as $\sigma^2/2 \to \alpha$.

Thus both 'external' parameters, the 'inverse temperature' and the 'noise', provide different necessary conditions for algorithm stability.

Somewhat curiously, Zilberstein and Russell (1995, Sec. 4.1), as an example, explicitly invoke a two-component sequential exponential model in a Lagrangian optimization. They do not, however, examine the 'phase transitions' that occur when $\lambda \to \beta$ or when $\sigma^2/2 \to \alpha$.

A logistic model, in deterministic form

$$F = \frac{C_1}{1 + \exp[-C_2 t]} - \frac{C_1}{2} \tag{36}$$

leads to the SDE

$$dF_t = (\beta - \alpha F_t^2)dt + \sigma F_t dW_t \tag{37}$$

Using the Ito chain rule again on F^2 gives the variance as

$$Var(F) =$$
$$1/16\,\frac{\left(\tanh\left(1/4\,t\sqrt{\sigma^4 + 16\,\beta\,\alpha}\right)\sqrt{\sigma^4 + 16\,\beta\,\alpha} + \sigma^2\right)^2}{\alpha^2}$$
$$-\frac{\left(\tanh\left(t\sqrt{\beta\,\alpha}\right)\right)^2\beta}{\alpha} \tag{38}$$

This increases as the fourth power of σ.

The optimization condition is

$$\beta_i[1 - \tanh\left(t_i\sqrt{\beta_i\alpha_i}\right)^2] = \lambda \tag{39}$$

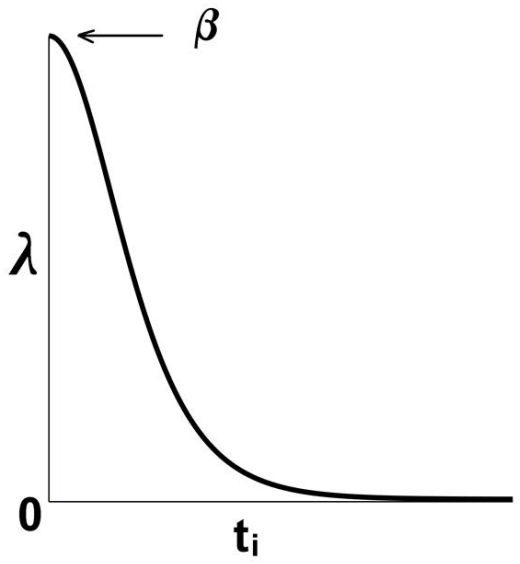

Figure 11: 'Inverse temperature' or 'price' plot for the logistic anytime algorithm model of Eq.(31). Again, the necessary conditions for stability are $\lambda \leq \beta$ and $\lambda \geq 0$. Imposing a value for λ also imposes the value of t_i.

as displayed in figure 11. Again, λ is an externally-imposed 'inverse temperature' or 'price' parameter. Eq.(39) again imposes the necessary stability conditions that $\lambda \leq \beta$ and $\lambda \geq 0$.

Within that range, as for the exponential model above, the value of t_i can be solved for in terms of the parameters α, β and λ and inserted into the variance relation, giving

$$Var(F) = 1/16\,\frac{1}{\alpha^2}\left(\tanh\left(1/4\,\frac{\sqrt{\sigma^4+16\,\beta\,\alpha}}{\sqrt{\beta\,\alpha}}\operatorname{arctanh}\left(\frac{\sqrt{\beta\,(\beta-\lambda)}}{\beta}\right)\right)\sqrt{\sigma^4+16\,\beta\,\alpha}+\sigma^2\right)^2 - \frac{\beta-\lambda}{\alpha} \tag{40}$$

which, again, grows as the fourth power of σ. And again, the term containing $\sqrt{(\beta-\lambda)/\beta}\,)$ implies the necessary condition that $\lambda \leq \beta$.

Simulation of the Arrhenius, exponential, and logistic models of algorithm quality vs. time is possible using the ItoProcess function of the Finance package

in the computer algebra program Maple 2018.1. The simulations are adjusted to converge on unit quality, and have the same values for the 'noise' parameter σ. As Fig.(12) shows, however, they display different patterns of scatter. Indeed, the Arrhenius model was found, at best, to be marginally stable under simulation, as perhaps implied by the analytic results.

More general SDE models of anytime algorithms, involving expressions having the form

$$dF_t = G(F_t)dt + \sigma F_t dW_t \tag{41}$$

can be explored analytically or numerically case-by-case. However, recall the arguments at the end of Section 3. The approach of Appleby et al. (2008) is analogous in spirit to those of Quevedo and Gupta (2012) and Dang et al. (2018) and permits a deeper and more general insight.

Relations of the form

$$dF_t = G(F_t)dt + h(F_t)dW_t \tag{42}$$

having an unstable equilibrium at $F_t = 0$ can *always* be driven to zero in probability by the 'proper' choice of the noise function $h(F_t)$. That is, an h-function can always be found that stabilizes an unstable equilibrium.

There is, here, something of a paradox: What are presumed to be well-behaved anytime algorithms may become seriously, and likely unexpectedly, unstable under proper stochastic 'fog-of-war' challenge, i.e., when what may be fairly rigorous sufficient conditions for algorithm stability are violated.

Figure 7b, taken from Zilberstein (1996), catches something of the inherent variability of the method, but absent the full fog-of-war increment we impose here. Again, see Quevedo and Gupta (2012) and Dang et al. (2018) for another approach to 'sufficient' conditions for stochastic anytime algorithm stability. And again, their methods are essentially similar to the approach of Appleby et al. (2008), applied specifically to the stability of stochastic differential equations.

These results represent, perhaps, a different version of the John Boyd mechanism, involving the injection of uncertainty into an opponent's command loop, in contrast to simply operating faster that the opponent can respond.

The time constraint that defines anytime algorithms in, for example, Eq.(31) can be formally placed in the larger context of the optimization results of the second section by extending the definition of F in terms of the symmetric variate $Z_j = M_j C_j H_j t_j$, here taking the relatively well-behaved exponential model as a test case:

$$\begin{aligned} F_j(M_j, C_j, H_j, t_j) = \sum_j \frac{\beta_j}{\alpha_j}(1 - \exp[-\alpha_j Z_j]) \\ M = \sum_{j=1}^{n} M_j,\ C = \sum_{j=1}^{n} C_j, \\ H = \sum_{j=1}^{n} H_j,\ T = \sum_{j=1}^{n} t_j \end{aligned} \tag{43}$$

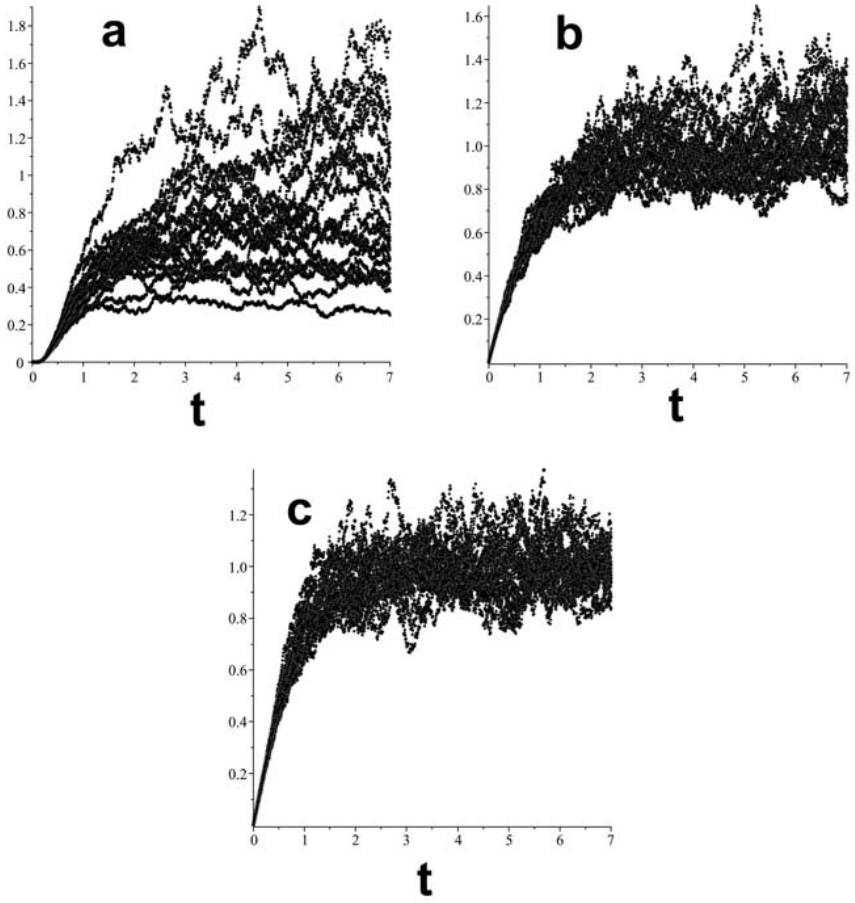

Figure 12: Twenty simulations of the Arrhenius (a), exponential (b) and logistic (c) models for anytime algorithm convergence as a function of time. The vertical axis represents solution quality and the horizontal time. Although the value of the 'noise' parameter σ is the same across all three, the Arrhenius model displays considerable instability, as implied by the analytic result of Eq.(29) that $dF_t \rightarrow \sigma F_t dW_t$.

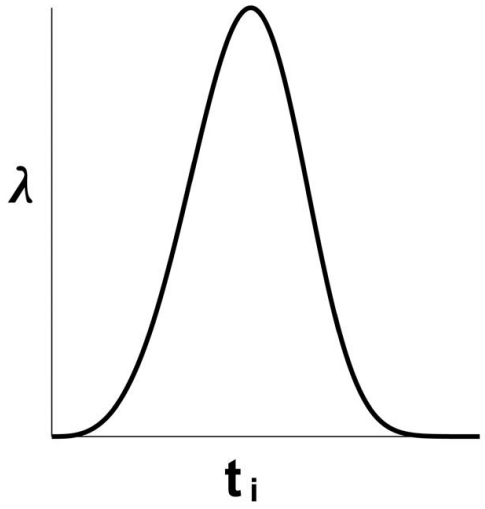

Figure 13: Inverse temperature or price plot for the hierarchical exponential model of an anytime algorithm under constraints of C, M and H from Section 3 in addition to the time constraint. Again, λ is externally-imposed, and the necessary conditions for stability are $\lambda \leq 1.522773...\beta/(T\alpha)^{1/4}$ and $\lambda \geq 0$. The multi-variate time-constrained hierarchical system is much different from a single-level time constraint, as shown in Fig.(6).

We normalize the non-time variable sums to unity, assuming $\sum_j x_j = X \equiv 1$, and then apply the symmetry result of Eq.(6), taking all α and β as the same across the subsystems. This gives a Lagrangian of order four in t as

$$L = \sum_i \frac{\beta}{\alpha}(1 - \exp[-\alpha t_i^4/T^3]) + \lambda(T - \sum_i t_i) \tag{44}$$

leading to the optimization condition

$$4\beta t_i^3/T^3 \exp[-\alpha t^4/T^3] = \lambda \tag{45}$$

This is shown in figure 13 and should be contrasted with Fig.(10), the result for the simple exponential model. As above, Eq.(45) imposes necessary conditions for system stability. The peak is found here by setting $\partial^2 L/\partial t^2 = 0$. This gives the necessary conditions for stability as $\lambda \leq 1.522773...\beta/(T\alpha)^{1/4}$ and $\lambda \geq 0$.

Some consideration shows that maximization of L in Eq.(44) is far, far more complicated than was the case for Eq.(9), an effort that led to the ambiguities and complexities of Fig.(3). That is, optimizing time constraints when 'other factors' must be simultaneously optimized – or at least taken into account – is not at all straightforward and, as in Fig.(3), will likely lead to a large set of deceptively counterintuitive synergisms.

This is no small matter. Fig.(14) shows two contrasting examples over the same time constraint. The first is a one-level two-component system that only optimizes time, for which one sets $\alpha = \beta = 1$ and then seeks to maximize the expression

$$(1 - \exp[-t]) + (1 - \exp[-(T - t)]) \tag{46}$$

For $T = 30$ this is shown in Fig.(14a).

A two-component, three-level plus time system, under the same conditions, must maximize the expression

$$(1 - \exp[-t^4/T^3]) + (1 - \exp[-(T - t)^4/T^3]) \tag{47}$$

This is shown, again for $T = 30$, in Fig.(14b).

The first case is much as expected. For only a time constraint, the maximum occurs when time is equally shared between the two components. Given three additional constraints, as in Section 3, if the time constraint T is below a critical value representing a phase transition, the system as-a-whole becomes highly pathological. That is, adding optimization requirements in addition to time-of-computation will impose highly punctuated destabilization if the time limit T is sufficiently constrained. This result can probably be proved in general and is likely already hidden in the literature (e.g., Ashok and Patra 2010).

One possible approach to a deeper analysis is to look at the solutions of the relation $\lambda = \partial L/\partial t_1 = \partial L/\partial t_2$ for $t_2 = T - t_1$ in the expression of Eq.(47), here over the range $0 \leq T \leq 30$. This is done in Fig.(15), using the *implicitplot* function of Maple 2018.

The central node $t_1 = t_2 = T/2$ is evident throughout, but only beyond the bifurcation point at the time constraint $T \approx 12$ does that partition represent a maximum value to the relation of Eq.(47). That is, for time constraints below the bifurcation point, a version of the John Boyd destabilization mechanism dominates system dynamics. What this implies depends on the observer's relation to the underlying conflict.

A stochastic extension is possible.

Imposing the full, symmetric, time constraint as determined by the Lagrangian formalism, the appropriate SDE becomes fourth-order in time,

$$dF_t = 4\beta t^3/T^3 \exp[-\alpha t^4/T^3]dt + \sigma F_t dW_t \rightarrow \sigma F_t dW_t \tag{48}$$

And as for the Arrhenius case of Eq.(29), sufficient time drives the system to unrestrained volatile diffusion: use of the Ito Chain Rule on F^2 shows the variance grows explosively, as $\exp[\sigma^2 t]$, or as $\exp[\sigma^2 T/2]$ along the node of Fig.(15), placing a serious limit on the operation of multivariate hierarchical, time-constrained systems that is not necessarily present in the simple one-stage time systems of Fig.(9).

The possible values of t_i are, as above, determined by solving Eq.(45).

If the time constraint T is too short, however, one is confronted by the instability shown in Fig.(14b). The tradeoff appears singularly brutal. A similar result, of course, follows use of the multidimensional Arrhenius model.

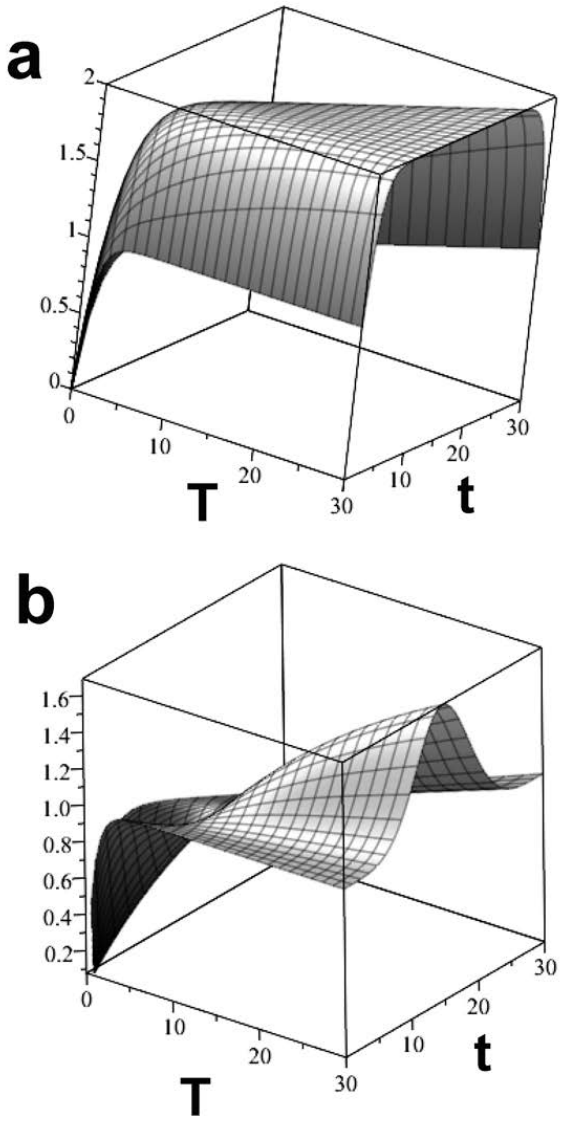

Figure 14: (a). Under only a time constraint, the Lagrange maximization is nominal, with time shared equally between the two sequential components. (b) In the presence of three constraints in addition to that of time, only under a sufficiently large time constraint T does the full system show anything approximating a 'normal' behavior, displaying a phase transition at a critical value of T.

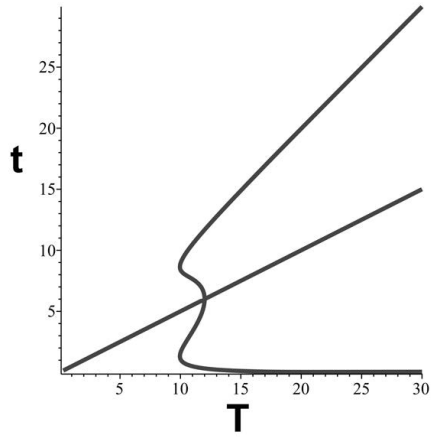

Figure 15: Solutions to the equation $\lambda = \partial F/\partial t_1 = \partial F/\partial t_2$, $t_2 = T - t_1$ for the relation of Eq.(47) as shown in Fig.(14b). While the node $t_1 = t_2 = T/2$ is evident, only beyond the bifurcation phase transition at $T \approx 12$ does that partition represent a maximum. Below that value of the time constraint the system is driven into instability, much in the sense of the military tactician John Boyd.

Here, we have found, quite generally, that the inverse temperature λ and the noise σ constitute independent environmental burdens imposed by the externalities of conflict and that these parameters, in turn, impose necessary conditions on the stability of anytime algorithms. Violation of those conditions will constitute some version of the John Boyd mechanism, leading to system failure. One possible inference is that λ represents the inverse temperature of the fog-of-war, and σ is a measure of its density.

A different but related series of models can be made by viewing an anytime algorithm as the transmission of a 'problem' into a 'solution' via a channel defined by computation in the presence of noise.

We do this first under a time constraint that can be seen as an analog to a power constraint under conditions of noise, if only because the total energy consumed by the computer will be proportional to the time it runs. Then, assuming a Gaussian channel characterizes computation, we can obtain a 'channel capacity' for the calculation in terms of the amount of time allowed – defining the total signal power – following the literature (e.g., Cover and Thomas 2006):

$$C = \beta \log(1 + \frac{t}{\sigma}) \tag{49}$$

under the time constraint $T = \sum_j t_j$ and a noise level σ. Assuming independence of the individual subcomponents, this produces a Lagrangian that directly

incorporates noise

$$L = \sum_j \beta_j \log(1 + \frac{t_j}{\sigma_j}) + \lambda(T - \sum_j t_j) \tag{50}$$

leading to the optimization condition as

$$\frac{\beta_j}{\sigma_j(1 + t_j/\sigma_j)} = \lambda \tag{51}$$

Eq.(51) is convex in t and has a maximum possible value of β/σ at $t = 0$. If the product $\lambda\sigma$ is greater than β, if the synergism between inverse temperature and the density of the fog-of-war is sufficient, no optimal solution is possible, and the system fails.

Replacing t/σ in Eq.(49) by $(t/\sigma)^n$, where $n > 1$, the sufficient condition for optimization failure becomes

$$\lambda\sigma > \beta(n-1)^{(n-1)/n} \tag{52}$$

If the times t_j themselves are stochastic variables, then Eq.(50) can be averaged across the appropriate distributions. For example, taking exponential distributions for individual components, so that

$$< t_j > \equiv \tau_j = \int_0^\infty \omega_j \exp[-\omega_j t] dt = 1/\omega_j \tag{53}$$

leads to Lagrangians for $n = 1$ and $n = 2$ as

$$L = \sum_j \beta_j \, \mathrm{e}^{\frac{\sigma_j}{\tau_j}} Ei\left(1, \frac{\sigma_j}{\tau_j}\right) + \lambda \left(T - \sum_j \tau_j\right)$$

$$L = \sum_j \beta_j \left(\mathrm{e}^{\frac{2\,i\sigma_j}{\tau_j}} Ei\left(1, \frac{i\sigma_j}{\tau_j}\right) + Ei\left(1, \frac{-i\sigma_j}{\tau_j}\right)\right) \mathrm{e}^{\frac{-i\sigma_j}{\tau_j}} + \lambda \left(T - \sum_j \tau_j\right) \tag{54}$$

where $Ei(1, x)$ is the exponential integral of order 1 in the variable x.

After some development, the respective sufficient conditions for optimization failure are $\lambda\sigma > \beta$ and $\lambda\sigma > 0.8136836...\beta$, in contrast to Eq.(52) where the first expression sufficies for both cases.

Even in the absence of time constraints, anytime algorithms can display various forms of stochastic instability. Indeed, Quevedo and Gupta (2012) and others take an inverse approach, studying sufficient conditions for stability of such systems. A 'computational' view, by contrast, centers on treating the solution quality as an inverse measure of distortion between a 'problem' sent as a message onto a 'solution' via the computation. We can then invoke the Rate Distortion Theorem that mandates a minimum channel capacity – a free energy measure in the sense of Feynman (2000) – for a given average distortion.

For a Gaussian channel under the squared distortion measure $D \equiv E[(X - E[X])^2]$ the Rate Distortion Function can be expressed as

$$R(D) = b \log[\omega^2/D] \tag{55}$$

where ω is taken here as the noise inherent to the computation, as opposed to the noise imposed by embedding fog-of-war and friction burdens, which we will continue to call σ.

Since $R(D)$ is a free energy measure, we can construct a corresponding entropy as $S = R - DdR/dD$, and then build an initial nonequilibrium Onsager diffusion-like model in the gradient of the entropy as $dD/dt = -\mu dS/dD$, giving $D(t) \propto \sqrt{\mu t}$ as in ordinary diffusion. See any reference on nonequilibrium thermodynamics. This result permits construction of a stochastic differential equation in terms of the external noise σ. Details of the method and appropriate references can be found in the appendix to Wallace (2017). The essential matter lies in the inherent convexity of *all* Rate Distortion functions in *any* appropriate distortion measure (Cover and Thomas 2006).

The resulting SDE, after some manipulation, has the form

$$dD_t = [\frac{\mu b}{D_t} - f(t)]dt + \sigma D_t dW_t \tag{56}$$

where μ is a diffusion coefficient and $f(t)$ represents the Arrhenius, the exponential, the logistic, or any similar model that converges as $t \to \infty$. In the absence of noise σ, the inverse distortion, a measure of calculation quality, converges on $f(\infty)/(\mu b)$. In the presence of noise, however, application of the Ito chain rule to D^2 finds a necessary condition for stability in variance over all such models as

$$\lim_{t \to \infty} f(t) \geq \sigma \sqrt{2\mu b} \tag{57}$$

Sufficient fog-or-war or frictional noise – large enough σ – violates this necessary constraint.

The convexity of the Rate Distortion Function implies a similar result for all other channels. For example, the 'natural' channel, having $R(D) \propto \omega^2/D$, leads to the necessary condition

$$\lim_{t \to \infty} f(t) \geq \frac{3}{2} \sqrt[3]{\mu \omega^2 \sigma^4} \tag{58}$$

7 A third model for failure of cognition in real-time conflict

The previous section introduced time constraint in the operation of a particular form of cognitive process, the 'anytime' algorithms that are supposed to provide 'good enough' solutions at any point during computation. As described in Section 3, institutions in real-time conflict – military or otherwise – must often reach decisions under some combination of resource, intelligence, and internal

bandwidth constraints, in addition to time constraints. We now extend the arguments of Section 3 to include time limitations.

Recall the first part of Eq.(2), representing the rate of cognitive response as determined by an 'essential resource'. Adding a time constraint to the M-C-H model of Section 3 leads to $Z_j = M_j C_j H_j t_j$, where we now set all the sums equal to 1, except that for time, so that $T = \sum_j t_j$. Following the arguments of Eqs.(6) and (7), but now *in terms of the time constraint*, leads to the Lagrangian

$$L = \sum_j \exp[-K_j T^3/t_j^4] + \lambda(T - \sum_j t_j) \tag{59}$$

The analog to Eq.(12) is then

$$\frac{4K_j T^3 \exp[-K_j T^3/t_j^4]}{t_j^5} = \lambda \tag{60}$$

producing a more sharply peaked version of Fig.(13). The 'inverse temperature' condition is then $0 \leq \lambda \leq 1.5147.../(KT^3)^{1/4}$.

The corresponding SDE becomes

$$dF_t = F_t \left[\frac{4KT^3}{t^5}dt + \sigma dW_t\right] \rightarrow \sigma F_t dW_t \tag{61}$$

Again, as in the case of Eq.(29), this SDE is unstable in the presence of noise, having $\mathcal{E}(F) \rightarrow \exp[-\sigma^2 t]$ and $Var(F) \rightarrow \exp[\sigma^2 t]$.

Two-component models closely replicate the results of Fig.(12b) and Fig.(13). Like the Arrhenius model for anytime algorithms, forcing the system to operate under a time constraint less than that defined by the bifurcation point of Fig.(15) triggers failure: the John Boyd mechanism again.

In addition, high dimensional Arrhenius models appear to suffer inherent stochastic instabilities analogous to those explored in Fig.(12a).

The real world is not a game of Go.

8 Discussion

There are, then, many ways in which a multi-level, hierarchical cognitive system – institutional, autonomous machine, or man/machine 'cockpit/centaur' composite – can fail on Clausewitz landscapes of contention that are strongly dominated by 'friction' and 'fog-of-war'.

From the first model, failure can involve response thresholds that become too small or too large, conditions a, d, and e of figure 3. That is, the hierarchical system seems to work only in a narrow range of K-values. Behavior leading to failure might include letting some cognitive adversary define the conditions of engagement, or failing to respond to such an adversary's actions in a timely manner.

A second failure mode for the first model, even if K is kept within acceptable range, is to constrain internal bandwidth available to culturally-defined 'tactical' and 'operational' levels of organization, placing the system in an efficiency mode that is unsustainable in the long term, regardless of the ability to respond appropriately.

A third failure mode also seems to be externally driven, involving 'temperature' or 'cost' demands that are outside the response range of the system.

The inference is that both K and λ can be imposed by an adversary, while limiting communication channels available to tactical and operational levels will guarantee failure even under the best of circumstances.

A fourth failure mode for the first model involves the impact of a fog-of-war 'noise', leading, under sufficiently protracted conflict, to escalation under resource constraints that triggers an erosion much like a population extinction.

The first model might be extended in various ways. For example, one can examine the effects of a distribution in excitation thresholds K. Inverted-U distributions, having distinct positive modes, give results similar to what we examine here. By contrast, reverse-J distributions like the exponential, that have a strong mode at zero, lead to a canonical pathological response surface much like figure 3a or 3e. Other extensions would be to algebraically derive the pattern of figure 3, showing the range of inherently destabilizing activation thresholds, and an explicit stochastic optimization treatment of the matters raised in the second section. The latter extension is not trivial.

A second model, taking a more classic 'network' approach to institutional cognition, focuses on highly punctuated phase transitions driven by the failure of crosstalk – information exchange – between submodules, based on collapse of groupoid symmetries associated with large-scale cognitive function (Wallace 2012; 2017; 2018b).

Imposition of the Data Rate Theorem linking control and information theories provides a different portrait of system failure under conflict, finding consonance with the first two models via Eq.(26). This approach is perhaps most directly analogous to that of the US military strategist John Boyd, involving 'getting inside the command loop' of an adversary (Wallace 2018a).

A relatively direct line of argument, somewhat surprisingly, implies catastrophic failure of AI's increasingly important anytime algorithms under sufficient stochastic variation, implementing both Clausewitz's 'friction' or 'fog-of-war' and Boyd's command loop circumventions in armed conflict. Something of this already lies buried in the literature, involving elaborate and highly restrictive sufficiency conditions for algorithm stability that are seldom fully explored or fully explained.

Indeed, the anytime algorithm analysis leads to a reconsideration of failure of cognition in the hierarchical multi-level model studied in Section 3, suggesting that instabilities related to time constraint will be synergistic with those of material and information limitations: command loop delays interacting with fog-of-war and friction.

The bottom line is that all real-time, multi-component, multi-level cognitive enterprises – institutional, autonomous machine, or man/machine hybrid

– acting under constraints of time, resources, internal communications, and intelligence information, can be confounded by activities of a cognitive adversary having sufficient situational awareness and resources, or by the evolution of embedding environments under the selection pressure generated by the enterprises themselves (Wallace 2018a, Ch.6). There will be no free lunch for AI systems, and these will not and cannot be exempt from challenge and frequent failure in the real-time real world, in spite of corporate marketing rhetoric to the contrary.

The real world is not a game of Go, and those familiar with the Theranos example (Carreyrou 2018; Couzin-Frankel 2018) may perhaps see a similar pattern emerging around the headlong rush to use AI in real-time critical systems. It will not be sufficient for tech companies to follow their usual practice of beta testing unstable systems on human populations for refinement. Over Clausewitz landscapes, ultimately, there are always significant casualties and collateral damage, but such losses are massively compounded by arrogance, blindness, and command stupidity.

References

Appleby, J., X. Mao, A. Rodkina, 2008, Stabilization and destabilization of nonlinear differential equations by noise, IEEE Transactions on Automatic Control, 53:126-132.

Ashok, B., T. Patra, 2010, Locating phase transitions in computationally hard problems, PRAMANA Journal of Physics, 75:549-563.

Carreyrou, J., 2018, Bad Blood: Secrets and Lies in a Silicon Valley Startup, Knopf, New York.

Couzin-Frankel, J., 2018, The rise and fall of Theranos, Science, 360:720.

Cover, T., J. Thomas, 2006, Elements of Information Theory, Second Edition, Wiley, New York.

Dang, T., K. Ling, D. Quevedo, 2018, Stability analysis of event-triggered anytime control with multiple control laws, arXiv:1804.07876v1 [math.OC] April 2018.

Ernst, N., D. Carroll, C. Schumacher, M. Clark, K. Cohen, G. Lee, 2016, Genetic fuzzy based artificial intelligence for unmanned combat areal vehicle control in simulated air combat missions, Journal of Defense Management, 6:144 (online).

Feynman, R., 2000, Lectures on Computation, Westview Press, New York.

Glazebrook, J.F., R. Wallace, 2009, Rate distortion manifolds as models for cognitive information, Informatica, 33:309-345.

Golubitsky, M., I. Stewart, 2006, Nonlinear dynamics and networks: the groupoid formalism, Bulletin of the American Mathematical Society, 43:305-364.

Hwang, C., A. Masud, 1979, Multiple Objective Decision Making, Methods and Applications, Springer, New York.

Jin, H., Z. Hu, X. Zhou, 2008, A convex stochastic optimization problem arising from portfolio selection, Mathematical Finance, 18:171-183.

Matsumoto, Y., 2002, An Introduction to Morse Theory, American Mathematical Society, Providence, RI.

Nair, G., F. Fagnani, S. Zampieri, R. Evans, 2007, Feedback control under data rate constraints: an overview, Proceedings of the IEEEE, 95:108-137.

Osinga, F., 2007, Science, Strategy and War: The Strategic Theory of John Boyd, Routledge, London.

Pettini, M., 2007, Geometry and topology in hamiltonian dynamics and statistical mechanics, Springer, New York.

Protter, P., 2005, Stochastic Integration and Differential Equations, Second Edition, Springer, New York.

Puget, J., 2005, Breaking all value symmetries in surjection problems, pp. 490-504 in P. Van Beek (Ed.), CP 2005, LNCS 3709, Springer, New York.

Quevedo, D., V. Gupta, 2012, Sequence-based anytime control, IEEE Transactions on Automatic Control, 58:377-390.

Schrodinger, E., 1989, Statistical Thermodynamics, Dover, New York.

Wallace, R., 2012, Consciousness, crosstalk, and the mereological fallacy: an evolutionary perspective, Physics of Life Reviews, 9:426-453.

Wallace, R., 2016, 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., 2017, Computational Psychiatry: A systems biology approach to the epigenetics of mental disorders, Springer, New York.

Wallace, R., 2018a, Carl von Clausewitz, the Fog-of-War, and the AI Revolution: The real world is not a game of Go, Springer, New York.

Wallace, R., 2018b, Culture and the trajectories of developmental pathology: insights from control and information theories, Acta Biotheoretica, 66:79-112.

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 Mathatical Association, 43:744752.

Zilberstein S., S. Russell, 1995, Approximate Reasoning Using Anytime Algorithms. In: Natarajan S. (ed.) Imprecise and Approximate Computation. The Springer International Series in Engineering and Computer Science (Real-Time Systems), Vol 318. Springer, Boston, MA

Zilberstein, S., 1996, Using anytime algorithms in intelligent systems, AI Magazine, 17:73-83.