The Planetary Science Journal, 6:176 (13pp), 2025 July https://doi.org/10.3847/PSJ/ade310

Exo-Daisy World: Revisiting Gaia Theory through an Informational Architecture Perspective

Damian R Sowinski[1], Gourab Ghoshal[1,2], and Adam Frank[1]
[1] Department of Physics and Astronomy, University of Rochester, Rochester, NY 14627, USA; damian.sowinski@Rochester.edu
[2] Department of Computer Science, University of Rochester, Rochester, NY 14627, USA
Received 2024 November 14; revised 2025 May 27; accepted 2025 June 4; published 2025 July 23

Abstract

The Daisy World model has long served as a foundational framework for understanding the self-regulation of planetary biospheres, providing insights into the feedback mechanisms that may govern inhabited exoplanets. In this study, we extend the classic Daisy World model through the lens of semantic information theory (SIT), aiming to characterize the information flow between the biosphere and planetary environment—what we term the information architecture of Daisy World systems. Our objective is to develop novel methodologies for analyzing the evolution of coupled planetary systems, including biospheres and geospheres, with implications for astrobiological observations and the identification of agnostic biosignatures. To operationalize SIT in this context, we introduce a version of the Daisy World model tailored to reflect potential conditions on M dwarf exoplanets, formulating a system of stochastic differential equations that describe the coevolution of the daisies and their planetary environment. Analysis of this exo-Daisy World model reveals how correlations between the biosphere and environment intensify with rising stellar luminosity and how these correlations correspond to distinct phases of information exchange between the coupled systems. This rein control provides a quantitative description of the informational feedback between the biosphere and its host planet. Finally, we discuss the broader implications of our approach for developing detailed ExoGaia models of inhabited exoplanetary systems, proposing new avenues for interpreting astrobiological data and exploring biosignature candidates.

Unified Astronomy Thesaurus concepts: Exoplanet evolution (491); M dwarf stars (982); Astronomical simulations (1857); Exoplanet surface variability (2023); Exoplanet surface characteristics (496); Star-planet interactions (2177); Solar-planetary interactions (1472); Planetary science (1255)

1. Introduction

Understanding the long-term evolution of planetary biospheres is a crucial frontier in astrobiology. The search for life on exoplanets requires the identification of biosignatures, which rely on life having significantly altered the spectroscopic properties of a planet. Thus, exoplanetary life searches focus not on detecting individual organisms but on identifying the collective effects of life on the planetary system—what we refer to as exo-biospheres.

The study of biosignatures is therefore inseparable from the study of biospheres. Given that exoplanets may be observed at any point in their evolutionary history, it becomes essential to understand how and when biospheres reach a *mature* state, wherein they exert strong feedback on the planetary geospheres (atmosphere, hydrosphere, cryosphere, and lithosphere). This same argument applies to *technosignatures* and *technospheres* (i.e., intelligence; J. Zalasiewicz et al. 2017; A. Frank et al. 2022), but in this work, we focus exclusively on nontechnological life.

In this regard, there has been considerable progress made in astrobiology regarding the relationship between life and its host planets. Such progress is captured in a number of excellent books on the subject, which review the current state of the field (V. S. Meadows et al. 2020; M. Lingam & A. Loeb 2021; M. Lingam & A. Balbi 2024; see chapters 4 and 5 of the first reference in particular). Of particular interest is the question of habitability, which is defined in terms of an environment enabling a population of organisms to persist over geological timescales (C. S. Cockell et al. 2024). Habitability is, however, not a one-way street. Within environmental and ecosystem studies, it has long been recognized that communities of living systems can alter their environments just as much as those environments drive biological evolution. Given that exoplanet astrobiology requires the identification of whole-biosphere effects on an exoplanet's spectral output, this recognition of the *coevolution* of biospheres and geospheres becomes of central importance to the search for life (M. Lingam & A. Loeb 2021). In particular, understanding how life can exert feedbacks on its host world that then shape that world's evolution in ways beneficial to life represents a frontier in understanding the long-term habitability (i.e., inhabitation) of exoplanets.

One of the earliest and still useful models of biospheres exerting strong feedbacks on their planet's evolution is the Daisy World model. This simple dynamical model is associated with the Gaia theory of James Lovelock and Lynn Margulis, who fully articulated the range of potential biospheric effects on a planet through their Gaia theory (J. E. Lovelock & L. Margulis 1974; L. Margulis & J. E. Lovelock 1974; J. E. Lovelock & A. J. Watson 1982; L. Onori & G. Visconti 2012). Initially termed self-regulating Earth system theory, the Gaian model posited by Lovelock and Margulis emphasized the dense feedback loops between the biosphere and the abiotic planetary system, allowing a planet to maintain habitability over long timescales. It is useful for our purposes to briefly review Gaia theory.

Gaia theory faced considerable resistance upon its introduction. Criticisms focused on issues of teleology in evolution (R. Dawkins 1982) as well as questions surrounding how natural selection could work for entire biospheres (W. F. Doolittle 2017). In addition, while a number of feedback mechanisms for full homeostasis have been suggested, none have yet been conclusively validated empirically for the Earth (see T. Tyrrell 2013 for a modern critique of the Gaia hypothesis). While some of these questions remain, Gaia theory's most lasting contribution was, perhaps, to highlight the idea of *coevolution* between the biosphere and geospheres. It was in this way that many of the fundamental principles of Gaia theory became repackaged as Earth systems science (ESS), which forms the foundation of modern approaches to Earth's evolutionary history (W. Steffen et al. 2020). In particular, ESS adopted Gaia theory's recognition of the biosphere as a key driver of planetary evolution—a concept that astrobiology also applies in the search for exoplanetary biosignatures.

Progress in astrobiology thus depends on advancing our understanding of how biospheres coevolve with planetary geospheres (atmosphere, hydrosphere, cryosphere, lithosphere). This requires untangling the complex network of Gaian (or semi-Gaian) feedbacks that regulate planetary systems. As a complex adaptive system (D. Krakauer 2024), the biosphere's dynamics span multiple spatial and temporal scales, with both upward and downward causal chains.

In this work, we explore an alternative approach to characterizing biospheric complexity by combining physical descriptions of bio-geosphere feedback with information-theoretic measures, constructing what we call an *information narrative*. To do so, we investigate the simplest model of biospheric homeostasis: the well-known Daisy World model (A. J. Watson & J. E. Lovelock 1983). Daisy World is a toy model originally developed to illustrate Gaian self-regulation. The model consists of a planet with white and black daisies and a star with increasing luminosity over time. As the star's brightness changes, the daisy populations either grow or diminish, modifying the planet's albedo in such a way that its temperature remains within a habitable range for the daisies.

We introduce a stochastic variant of Watson and Lovelock's model, which we refer to as exo-Daisy World (eDW), to analyze aspects of its informational architecture. The original Daisy World model has been extensively studied and generalized in various ways, as highlighted in Wood's comprehensive 2008 review (A. J. Wood et al. 2008) and more recently in Savi and Viola's 2023 variant (M. A. Savi & F. M. Viola 2023). To our knowledge, our variant is the first to generalize Daisy World using coupled stochastic differential equations, which allow the application of information-theoretic measures to the model. Our overarching aim is to bring these information-centric methods to the study of biospheres, clarifying the processes underlying planetary coevolution.

Our manuscript is organized as follows. In Section 2, we provide a brief overview of semantic information theory (SIT), the information-theoretic framework we apply to the Daisy World model. In Section 3, we introduce our modified Daisy World model, incorporating stochastic variations in stellar luminosity. Section 4 presents the information measures drawn from SIT, which we apply to eDW. Section 5 outlines our results and their implications, followed by our conclusions in Section 6. The technical aspects of our model, the details of the numerics, and the approximations used are outlined in Appendix B.

2. SIT and Informational Architecture

Over the past three decades, innovative research avenues to explore life's self-production and self-maintenance have emerged from fields such as far-from-equilibrium thermodynamics and network theory. Both these perspectives have helped capture the essential nature of life as an emergent complex system. More recently, a critical new dimension has been added to these studies by focusing on life as a physical system that uses information (i.e., storage, copying, transfer and computation; S. I. Walker et al. 2016; M. D. Egbert & J. Pérez-Mercader 2018). By marrying an information-theoretic description with both network and thermodynamic constraints, life's information architecture emerges as a key conceptual construct for understanding self-production and self-maintenance, i.e., agency and autonomy.

Information's role in developing a *physics of life* faces a key challenge. Information theory, from Shannon's foundational work (C. E. Shannon 1948), focuses on syntactic measures, defining information via the combinatorics of strings without addressing the *semantic content* of information. While useful in engineering, this approach omits how life utilizes information meaningfully (G. Schlosser 1998; M. Mossio et al. 2009). For instance, understanding agency revolves around a system's ability to process information for its survival, often linked to its *viability*, a critical factor distinguishing living systems (X. E. Barandiaran & M. D. Egbert 2014; M. Egbert et al. 2023).

Although there are philosophical and computational discussions of semantic information (J. Barham 1996; D. Polani et al. 2001; C. L. Nehaniv et al. 2002; P. A. Corning 2007; T. W. Deacon 2007; E. Thompson & M. Stapleton 2009; D. R. Sowinski 2016; M. Gleiser & D. Sowinski 2018), a mathematically rigorous theory applicable to diverse scientific fields, including the origin of life, has been missing (J. E. Cohen 1977; J. A. Acebrón et al. 2005; N. Rooney et al. 2006; F. Cooper et al. 2013; J. C. Blain & J. W. Szostak 2014; S. Mimar et al. 2019, 2021; J. C. Xavier et al. 2020; D. Garcìa-Selfa et al. 2021). Recently (A. Kolchinsky & D. H. Wolpert 2018), a formalism for semantic information (SIT) was introduced using state spaces and probability distributions to assess mutual information between an agent and its environment, with persistence measured through a viability function. By scrambling information between systems of agents embedded in an environment, SIT quantifies how viability responds to altered information flow.

While promising, the formalism faces computational challenges, as it requires precise computation of high-dimensional state spaces and joint probability distributions. Thus, simplifying measures are needed for practical application, including effective definitions of viability, methods for scrambling information, and dimensional reduction techniques (I. G.-A. Pemartín et al. 2024). Recent studies applied SIT to models of foraging (D. R. Sowinski et al. 2023) and synchronization (D. R. Sowinski et al. 2024), capturing core biological behaviors like exploration, resource gathering, and collective behavior, providing a practical and computationally feasible road map for applying SIT to real-world applications. A recent perspective article has also outlined the broader

scientific motivations and empirical possibilities for SIT in the context of biology and planetary-scale systems (S. Bartlett et al. 2025).

A key result from D. R. Sowinski et al. (2023) was the identification of a viability plateau, where specific correlations between the agent and its environment had no measurable effect on survival, demonstrating that not all information is semantic. Beyond this threshold, increasing noise resulted in a decline in viability. This novel insight, which holds regardless of specific foraging strategies, highlights the potential for broader applications of SIT to other systems, including the context considered in this work.

The SIT framework can be summarized in the following steps.

1. Decompose the system's coarse-grained degrees of freedom (dof) into agent and environment components. Construct a *dynamical system* $(S, \mathcal{D})$, where S is the system state space and $\mathcal{D}$: $S \times T \to S$ is the dynamics on the state space for all times $t \in T$. The system must be decomposable into an *agent* and *environment*, $S = A \times E$, meaning that any state of the system can be written $s = (a, e)$, where $a = (a_1, a_2, \ldots) \in A$ and $e = (e_1, e_2, \ldots) \in E$ are, respectively, the agent and environment dynamical dof. We write $s(t) = (e(t), a(t))$ to represent the system evolution through state space and refer to the specific form of $\mathcal{D}$ responsible for $a(t)$ as the agent's *behavior*. In many cases, the separation of the system into agent and environment may be obvious (a foraging bacterium in a petri dish). In other cases, however, the separation may be more problematic. We note that this is a open question for information-theoretic studies of agential systems.
2. Identify and quantify the key biotic characteristics of the agent that contribute to its viability. The SIT formalism requires the definition of a *viability function* for the system. A dead agent represents the ground-state s_0 of the system. Thus, some measure of the *distance* to s_0 from a given viable state $s(t) = (e(t), a(t)) \neq s_0$ endogenously defines the *viability* of an agent.
3. Quantify the correlations present in the system, specifically between the agent and its environment. Correlations between agent and environment dof are manifest in the information architecture of the system. Information is quantified by a corpus of measures including mutual information $I(A\colon E)$ (T. M. Cover & J. A. Thomas 2006), cross entropy $H(A\colon E)$ (Z. Chen et al. 2022), and transfer entropy $\mathcal{T}_{E\to A}$ (T. Schreiber 2000). In practice, these quantities may not be calculable or efficiently computable in full; thus, coarse-graining of the system dof is required to treat the problem (M. Ibánez et al. 2024; Á. Tejero et al. 2025).
4. Investigate and articulate the relationship between biotic characteristics and these correlations. The goal of SIT is to identify how the correlations and viability defined in the last two steps relate to one another—specifically, to find which correlations matter to the agent. This is done by implementing a scrambling protocol on the correlations and examining the resultant change in viability due to the impact the scrambled correlations have on the dynamics of the system. This can be done using *interventions* to generate counterfactual histories. The procedure, however, becomes unmanageable for systems with more than several hundred states. Coarse-graining, if done correctly, identifies which coarse-grained dof should be investigated and scrambled. The point at which the scrambled viability diverges from the actual viability is the *semantic threshold* (D. R. Sowinski et al. 2023).

The formalism also necessitates a causal analysis across all possible interventions. However, as noted above, for all but the smallest systems, this approach is computationally infeasible. Therefore, in this study, *we omit the interventions step*, focusing instead on clarifying the correlation structure between the agent and environment as a function of external conditions, such as variations in stellar luminosity. Though we explicitly do not enact the final step of the full SIT program here, in future studies, the full ensemble of counterfactual trajectories based on scrambling initial system mutual information can be explored in the Daisy World or other biosphere–geosphere models.

3. The eDW Model

In its simplest incarnation, Daisy World is a planet with a fraction of its area, f, habitable by two species of flora, black and white daisies, occupying, respectively, f_B and f_W fractions of the planet's area. In what is to follow, the respective populations are indexed by $\alpha \in \{B, W\}$. The growth or decline of each flora is governed by a Malthusian law with a carrying capacity constraint, $f_W + f_B \leqslant f$; thus,

$$\frac{df_\alpha}{dt} = \beta(T_\alpha)(f - f_W - f_B)f_\alpha - \gamma_D f_\alpha. \tag{1}$$

Here γ_D is a decay rate (the same for both species), and $\beta(T)$ is the temperature-dependent growth rate. The latter peaks at some optimal temperature, $T_{\rm opt}$, and has a characteristic width ΔT. We note that this equation may be viewed in terms of the dynamics of a *metapopulation*—a system of interconnected local populations inhabiting a landscape of patches as formulated by R. Levins (1969).

The ground has an albedo A_G, while the flora have albedos $A_B < A_G$ and $A_W > A_G$. Consequently, the planetary albedo is the weighted sum,

$$A = A_G + \sum_\alpha (A_\alpha - A_G)f_\alpha. \tag{2}$$

Rein control is exerted by the flora through the positive and negative differences of each species' albedo with respect to the ground. The planetary temperature is constrained by the solar flux impingent on the planetary surface. For a stellar luminosity L and planetary orbital radius r, this is given by the Stefan–Boltzmann law,

$$T^4 = \frac{1 - A}{16\pi\sigma r^2}L, \tag{3}$$

where $\sigma \simeq 5.6703 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$. Due to their differing albedos, the flora are at different temperatures than the planetary average,

$$T_\alpha^4 = T^4 + q(A - A_\alpha), \tag{4}$$

which is governed by the coupling constant q. The magnitude of the differences in turn determines the growth rates of the respective species.

From a mathematical standpoint, the Daisy World model is a fully deterministic system governed by two equations of motion and a single constraint, Equation (3). This constraint assumes that the thermal equilibration timescale of the planetary surface (the time it takes the planet to respond to a change in stellar luminosity) is much shorter than the timescales associated with the growth and decline of the flora. As a toy model, however, it is not intended to directly represent the complexities of real planetary biospheres. For example, Earth experiences numerous stochastic events—such as volcanic eruptions and earthquakes—that are not captured in the deterministic framework of Daisy World. While incorporating such real-world stochasticity presents significant challenges, one approach to introducing variability is by modeling the full coupling of star and planet on comparable timescales. This approach is motivated by M dwarf planets, which present conditions where these timescales are more closely aligned. In this way, M dwarf systems serve as a useful heuristic for breaking the constraint in Equation (3). By coupling these timescales, we elevate temperature and stellar luminosity to dynamical dof in our eDW model. Fluctuations in stellar luminosity inject stochasticity into the system, providing a basis for the application of the SIT formalism to explore the informational dynamics of the model.

Considering M dwarfs, we note that flaring is one mechanism for luminosity fluctuations, with over 20% of M5 stars showing regular flaring events (R. R. Martínez et al. 2020). At even lower-mass scales, L dwarfs have variable brightness due to cloud formation on their surfaces. Using such variations as motivation for our toy model, we describe the luminosity of eDW's host star by an Ornstein–Uhlenbeck process,

$$\frac{dL}{dt} = \frac{1}{\tau_s}(\langle L\rangle - L) + \sqrt{\frac{2}{\tau_s}}\,\delta\langle L\rangle\,\eta. \tag{5}$$

This ensures that the mean luminosity is $\langle L\rangle$ with fluctuations of order $\delta\langle L\rangle$, where $\delta \ll 1$ and η is a heuristic white noise. The timescale of these stellar fluctuations is τ_s. We consider the exoplanet in the orbit of this star to have an atmosphere of scale height $h \ll R_p$ (the radius of the planet), density ρ, and specific heat c_V. A simple model for the planetary temperature dynamics assumes no work done on the atmosphere, resulting in

$$\frac{dT}{dt} = \frac{1-A}{16\pi r^2 h\rho c_V}L - \frac{\sigma}{h\rho c_V}T^4. \tag{6}$$

This gives the equilibration timescale $\tau_E = \rho h c_V/\sigma T_{\rm eq}^3$, and at equilibrium reproduces the Stefan–Boltzmann law, Equation (3). The derivation of both Equation (6) and τ_E can be found in Appendix A. Together Equations (1), (5), and (6) form a coupled set of stochastic differential equations describing the dynamics of eDW. The equations are closed with the provision of a growth rate.

In the original formulation of Daisy World, the growth rate $\beta(T)$ was introduced with a quadratic dependence on temperature, as follows:

$$\beta(T) = \begin{cases} \gamma_G\left(1 - 4\dfrac{(T-T_{\rm opt})^2}{\Delta T^2}\right) & |T - T_{\rm opt}| < \dfrac{\Delta T}{2} \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

with ΔT the growth rate bandwidth (the range of bearable temperatures for the daisies) and γ_G their maximal growth rate, which occurs at some optimal temperature, $T_{\rm opt}$. The behavior of such a dependence decreases away from the optimal temperature, and upon reaching 0, it abruptly stops. While this piecewise quadratic form is straightforward, in cases where differentiability is a concern, smooth functions like Gaussians have been employed as alternatives. When smoothness is not required, simpler forms such as triangular or rectangular window functions lead to sufficiently similar model behavior. For this study, we adopt a functional form that remains differentiable yet preserves a rectangular window shape:

$$\beta(T) = \gamma_G e^{-8\left(\frac{T-T_{\rm opt}}{\Delta T}\right)^4}. \tag{8}$$

This formulation offers both the smoothness needed for differentiability and the sharp transitions characteristic of window functions.

To better understand the timescales relevant to our stochastic eDW model, we provide some estimations that highlight why M dwarf planets serve as a useful exemplar. These stars exhibit luminosities ranging from 10^{-4} to $10^{-1}\ L_\odot$ (where $L_\odot = 3.846\times10^{26}$ W is the solar luminosity), and their masses fall between 0.075 and 0.6 $M_\odot$ (T. J. Henry & W.-C. Jao 2024). Exoplanets orbiting such stars typically have orbital radii of 0.05–0.25 au (M. Tuomi et al. 2019), resulting in orbital periods as short as several days. With planetary albedos assumed to be approximately 0.3, the equilibrated average surface temperatures of these exoplanets could span a wide range, from 50 to 600 K. Assuming atmospheric densities greater than Earth's (1–10 kg m^{-3}), specific heat capacities between 10^3 and 10^4 J kg^{-1} K^{-1}, and scale heights of 1–10 km, the corresponding equilibration timescales range from a single day to 10^6 yr. For exoplanets with temperatures near 200 K, this timescale is on the order of several days. Given that the orbital period provides an annual cycle for exoplanets with axial tilt or orbital eccentricity, we argue that this cycle also defines the timescale for biosphere dynamics. In the eDW model, this implies that $\beta(T_{\rm opt})^{-1} \sim \tau_E$, emphasizing the necessity of incorporating temperature dynamics into the model. It is important to note that we are not suggesting that M dwarf planets are expected to directly instantiate the eDW model; rather, we use these planets to illustrate the implications of timescale synchronization within the model.

Figure 1 illustrates the behavior of eDW across a range of average stellar luminosities (x-axes) and growth rate bandwidths (columns). Each point is taken from a single instance of the model; the description of the model dimensionalization, numerics, and ensemble statistics is fully detailed in Appendix B.2. The mean values of the daisy fraction, albedo, and planetary temperature (dashed lines in each of the panels) behave as in the original Daisy World model. The bandwidths of the growth rate are depicted in the bottom panels, manifesting as horizontal contour lines centered around the optimal temperature, $T_{\rm opt}$, with each contour corresponding to a successive 5% decline in growth rate. As the system transitions beyond the tolerable thermal range, the daisy populations (top panels) collapse, and planetary temperature dynamics become governed primarily by the bare surface albedo (middle panels). Consequently, the planetary

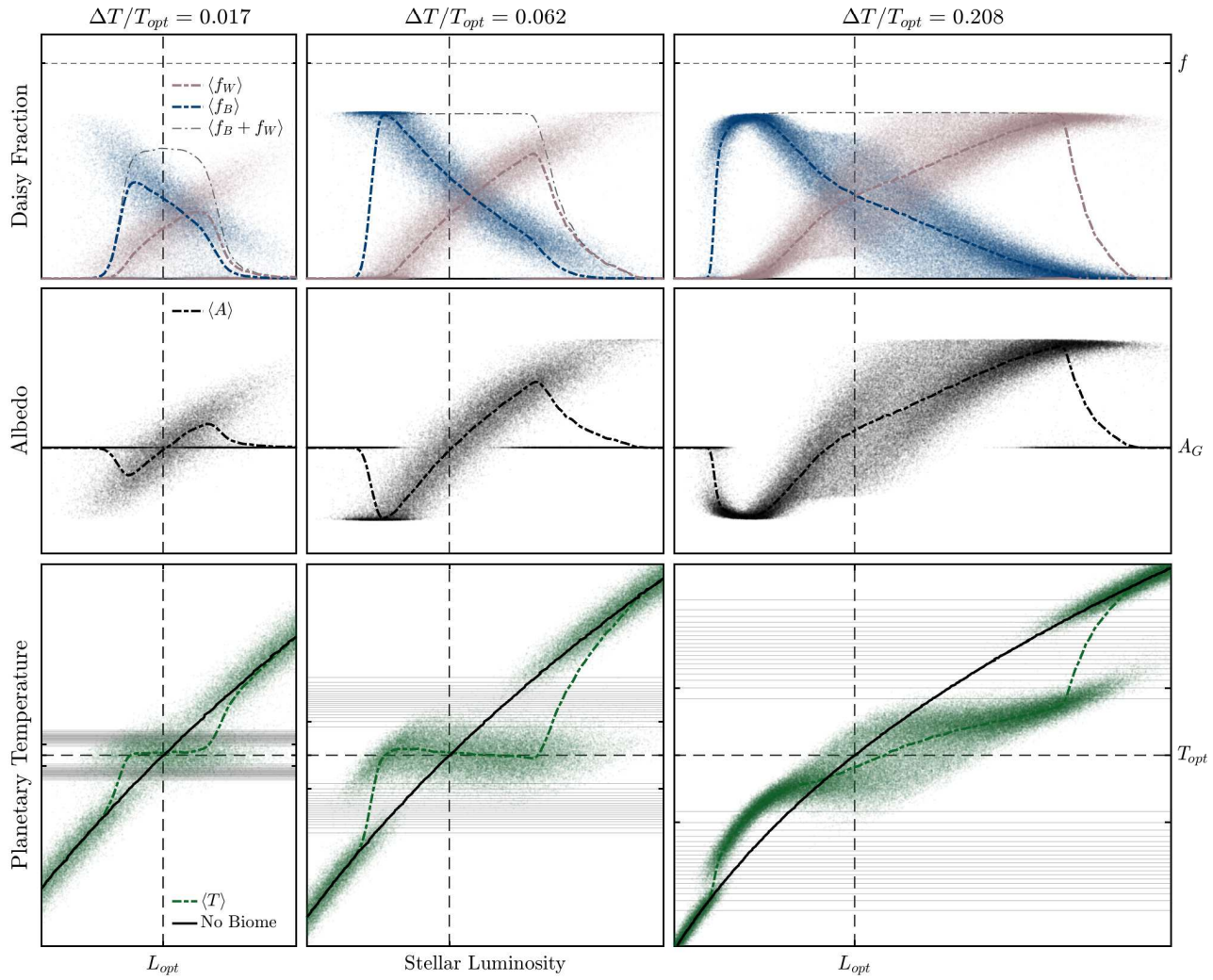

Figure 1. Qualitative behavior of the model; each column is an ensemble generated at a fixed daisy growth rate bandwidth as a fraction of the optimal temperature, $\Delta T/T_{\rm opt}$. In all plots, the scatter points are independent realizations from their column's ensemble, and the dotted–dashed lines are mean values. (Top row) Population sizes of the two species of daisy at equilibrium over a range of stellar luminosities. The horizontal dashed line is the habitable fraction of the planet, f. (Middle row) Daisies alter the planetary albedo, A_G, reining in (bottom row) the mean temperature of the planetary surface into the strip of temperatures that maximize daisy growth. The horizontal dashed line is the optimal temperature, $T_{\rm opt}$, with light gray contours drawn at 5% differences above and below. The vertical dashed line in all plots is the stellar luminosity that gives the optimal temperature on a biome-free planet.

temperature converges according to the Stefan–Boltzmann law, as represented by the thick black curves in the bottom panels.

At the lower threshold of the viable temperature range, the black daisy populations (blue curves) predominate, leading to decreased planetary albedo, which in turn drives a thermoregulatory feedback loop that elevates the geosphere's temperature toward the optimal range. Conversely, at the upper thermal boundary, the white daisy populations (pink curves) become dominant, increasing planetary albedo, thereby inducing a cooling feedback mechanism that lowers the planetary temperature. This exemplifies the well-known *rein control* mechanism, where the biosphere exerts regulatory influence on planetary conditions.

While the eDW behaves similarly to the classical Daisy World model on average, its inherent stochasticity introduces far richer dynamics, particularly at the boundaries of the habitable temperature and luminosity range. The bottom panels of Figure 1 illustrate the planetary temperature exhibiting fluctuations across all stellar luminosities. For lower bandwidths (left and middle columns), these fluctuations remain confined within 5% of the optimal temperature across the entire habitable luminosity spectrum. However, for higher bandwidths (right column), significant edge effects emerge at both the cooler and warmer extremes of the viable temperature range. These edge effects are also evident in the daisy fraction and albedo (middle panels), stemming from the nonlinearity of the Stefan–Boltzmann law, which becomes increasingly pronounced as the system explores a broader range of bearable temperatures.

4. Informational Narrative

In advancing our analysis, we adopt an SIT framework to characterize the system's complex behavioral dynamics. This approach requires partitioning the system into two distinct components: one representing the agent and the other representing the environment. In its current formulation, the eDW model is governed by 4 dof, (f_B, f_W, T, L).

Which of these variables represent the agent, and which represent the environment? One possibility is that the agent

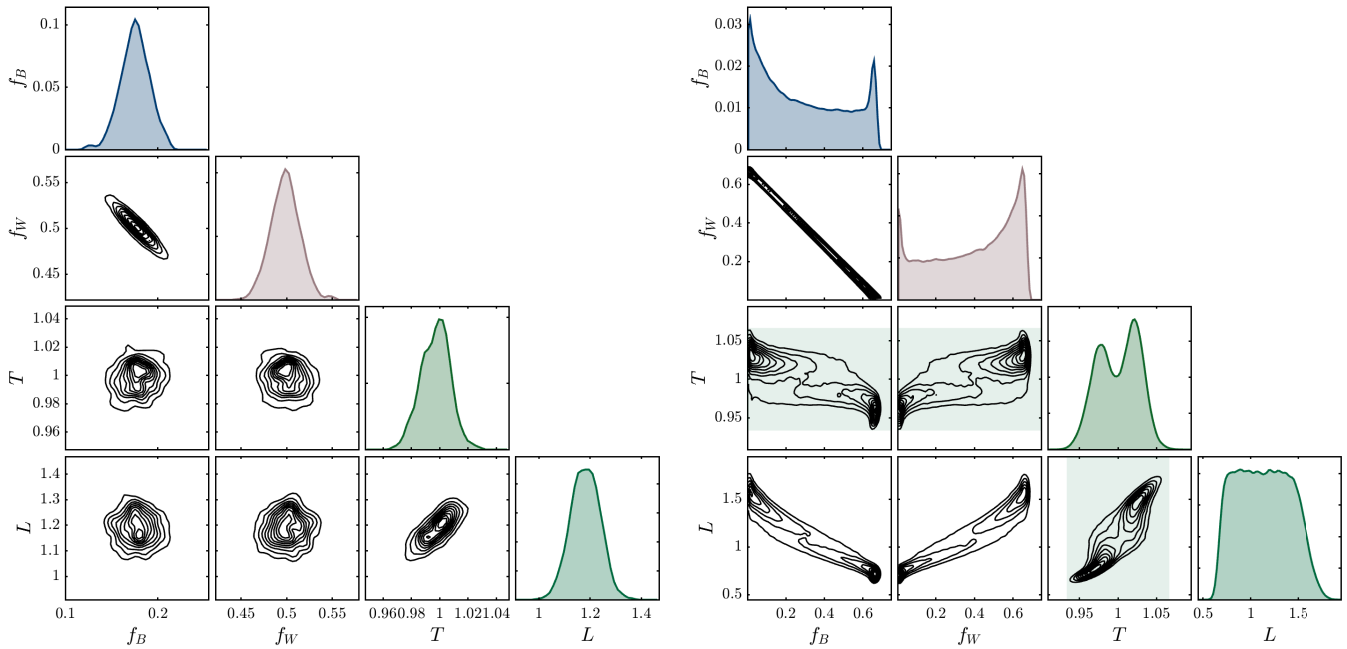

Figure 2. Distributions of dof in the eDW model and joint distributions between pairs of dof. (Left) Conditioned on an average stellar luminosity of $L = 1.1895L_{\rm opt}$ and a bandwidth of $\Delta T = 0.0833T_{\rm opt}$. It is clear that the environmental dof are correlated with one another, as are the agent dof. The pairwise correlations between agent and environment are not as clear. (Right) Not conditioned on any average stellar luminosity, with a bandwidth of $\Delta T = 0.1667T_{\rm opt}$ (visualized by the light green strip in three of the subplots). Over all luminosities, the pairwise correlations between agent and environment dof are much clearer.

corresponds to a single variable and the environment to the remaining three, yielding ${}_4C_1 = 4$ possible partitions. Alternatively, the agent might be described by two or three variables. This leads to a total of $\sum_{n=1}^{3} {}_4C_n = 14$ different ways of partitioning the system into nonempty agent and environment subsystems.

Aligned with the Gaia hypothesis, the most natural partitioning is to designate the planetary biome as the agent, while the stellar and planetary parameters constitute the environment. The daisies constitute the biosphere and are therefore identified as the agent, while the planet and its incident radiation represent the environment. Following this rationale, we define the agent's dof as $\boldsymbol{a} = (f_B, f_W)$ and the environmental dof as $\boldsymbol{e} = (T, L)$.

Next, we turn to the identification of viability. Since this is intrinsic to the agent, it must depend explicitly on the agent dof. A natural choice for viability is the expected proportion of the habitable fraction of the planet occupied by the biome, given by

$$V = \mathbb{E}^A\left[\frac{f_B + f_W}{f}\right], \tag{9}$$

where $\mathbb{E}^A$ represents the expectation value in the presence of the agent. The total biome area is normalized to the maximal habitable area for ease of interpretation, i.e., $0 \leqslant V \leqslant 1$, with $V = 1$ indicating that the biome occupies all of the habitable space on the planet.

We now focus on the correlations between the agent and environment, which can be extracted from the joint distribution $p_{\rm AE} = p_{\rm AE}(\boldsymbol{a}, \boldsymbol{e}|\boldsymbol{\theta})$. Here, $\boldsymbol{\theta}$ represents the full set of model parameters, as outlined in Appendix B, and includes both the average stellar luminosity, $\langle L\rangle$, and biome temperature bandwidth, ΔT. It is important to note that the agent's viability, as defined in Equation (9), is implicitly dependent on both of these parameters. Causal relationships between the agent and environment can also be determined by intervening in the system to obtain the agent-free distribution $p_0(\boldsymbol{e}) = p_{\rm AE}(\boldsymbol{e}|f_B = f_W = 0, \boldsymbol{\theta})$ (J. Pearl 2009). From both $p_{\rm AE}$ and p_0, the next step is to systematically enumerate and estimate the joint entropies for all possible subsets of dof. Information measures are then derived from specific combinations of these entropies, which will be discussed in detail later, to capture the informational architecture that characterizes eDW's dynamics. To avoid overwhelming the reader with extensive numerical details, the procedure for estimating both distributions is presented in Appendix B.

The left panel of Figure 2 shows all the univariate and bivariate marginals of the estimated $p_{\rm AE}$ from simulations with the average stellar luminosity fixed at $\langle L\rangle = 1.1895L_{\rm opt}$. We chose this value in order to motivate the need for information measures when analyzing correlations. In the panel, both intra-agent (f_B versus f_W) and intra-environment (T versus L) dof exhibit clear covariance in their bivariate marginals. However, the agent–environment dof do not display as obvious a relationship; the bivariate distributions appear circular. Two-point measures such as covariance have a hard time extracting intersubsystem correlations in this setting.

On the right, we marginalize $p_{\rm AE}$ over the history of the star's average stellar luminosity. The luminosity evolution is modeled by a linear increase from $0.3L_{\rm opt}$ to $2.4L_{\rm opt}$; such increases in luminosity over main-sequence evolution are typical of low-mass stars (M. Salaris & S. Cassisi 2005). Incorporating nonlinearities in the analysis will shift probability in the marginalized $p_{\rm AE}$ toward luminosities at which stellar evolution is slower but will not affect the qualitative features discussed. On an evolutionary timescale, the agent–environment dof display clear covariance. Taken together with

the left panel, the highly nonlinear interactions between biome and environment generate three-point and higher correlations between the two on short timescales; two-point correlations become prominent only after integrating the system across evolutionary timescales.

Extracting correlations between more than two dof on these short timescales necessitates the use of more advanced analytical tools, with information theory offering the required level of sophistication. The foundational tool within this framework is Shannon entropy. For a discrete random variable X sampled from the distribution p_X, the entropy is defined as

$$H(X) = -\sum_x p_X(x)\log_2 p_X(x) \tag{10}$$

and is measured in bits.

Since X is not restricted to being a scalar, entropy can be defined for collections of random variables. For a set of n random variables, each of the $n!$ possible subsets has an associated entropy. These entropies can be combined in various ways to construct more complex information-theoretic measures, each offering deeper interpretative insights. One of the most fundamental of these measures is mutual information. Consider two random variables, X and Y, and let their Cartesian product be denoted by XY. The mutual information between the two variables is defined as

$$\begin{aligned} I(X\colon Y) &= H(X) + H(Y) - H(XY) \\ &= \sum_{xy} p_{XY}(x, y)\log_2 \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)}. \end{aligned} \tag{11}$$

In the first expression in Equation (11), we have demonstrated how the measure is constructed from the entropies discussed earlier. In the second, these entropies are reorganized in terms of the joint distribution and its marginals. The former provides an intuitive interpretation: $H(XY)$ is the number of bits needed to represent x and y together, while $H(X) + H(Y)$ is the number of bits needed to represent x and y independently. If fewer bits are required to represent x and y together than independently, this implies that knowledge of one variable reduces the uncertainty in the other—mutual information quantifies exactly how much information is shared between the two. The latter expresses mutual information as a Kullback–Leibler divergence. Given that the Kullback–Leibler divergence is necessarily positive definite and measures the deviation from complete independence (as given by the product distribution in the denominator), it reinforces the interpretation of mutual information as a measure of shared information (S. Kullback & R. A. Leibler 1951). Since mutual information is sensitive to nonlinear correlations between entire sets of dof, it serves as an effective tool for capturing agent–environment dof correlations, denoted $I(A : E)$. Additionally, one can also quantify intra-agent correlations $I(a\_1 : a\_2)$ and intra-environment correlations $I(e_1 : e_2)$. All three measures are interpreted as quantifying the degree of dependency between the respective subsets of dof.

A biome, by restructuring planetary energy flows to suit its own needs, exerts a causal effect on planetary–stellar correlations. The question arises: does the addition of a biome to a planet strengthen or weaken these preexisting correlations? While the answer may vary depending on the specific case, the causal relationship between intra-environmental dependencies and the presence of an agent can be explored by computing the mutual information from the agent-free distribution, p_0. The quantity $I_0(e_1 : e_2)$ quantifies the strength of the correlation between stellar luminosity and planetary temperature. In the limit where $\tau_E \ll \tau_s$, the planetary temperature responds almost instantaneously to stellar fluctuations, leading to a high level of dependency between the two, as knowledge of one almost completely determines the other. Conversely, in the limit $\tau_E \gg \tau_s$, the slow response of the planet ensures that its temperature remains highly correlated with the average stellar luminosity but not with short-term luminosity fluctuations.

We focus on the dynamically interesting scenario where these two timescales, along with the agent's timescale, are comparable. To quantify the agent's effect, we take the difference between $I_0(e_1 : e_2)$ and $I(e_1 : e_2)$,

$$\Delta I_{\varnothing\to A} = I(e_1\colon e_2) - I_0(e_1\colon e_2), \tag{12}$$

which measures the production or destruction of intra-environmental correlations caused by the agent.

Unfortunately, we cannot conduct a similar causal analysis with the agent, as the question of what happens to a biome when its planet's star is removed has a simple and catastrophic answer. However, we propose a method to examine how the two daisy species cooperate with each other and the environment. Letting X and Y be as before, and introducing a third random variable Z, the conditional mutual information is given by

$$I(X\colon Y|Z) = I(X\colon YZ) - I(X\colon Z), \tag{13}$$

which represents the information shared between X and YZ, minus the information shared between X and Z alone. It measures the information shared between X and Y that cannot be attributed to Z alone. We utilize this to define intra-agent cooperation as influenced by the environment thus,

$$\begin{aligned} C(a_1\colon a_2||E) &= I(a_1\colon a_2) - I(a_1\colon a_2|E), \\ &= H(E) - H(a_1E) - H(a_2E) - H(A) + H(AE), \end{aligned} \tag{14}$$

with the first line offering the seemingly straightforward interpretation of the information shared between the biome dof that can be solely attributed to the environment. Seemingly, this definition can be reformulated into the manifestly symmetric form shown in the second line, which is known as interaction information (P. L. Williams & R. D. Beer 2010; R. G. James et al. 2011). Interaction information is known to take both positive and negative values. In the former case, it represents shared information across all random variables, interpreted as information redundantly distributed. Negative values of cooperation suggest a synergistic effect between the dof, where the presence of the third variable enhances the correlation between the other two.

5. Results

The information-theoretic measures introduced in the previous section were not the only ones we analyzed; however, they were the ones that produced the most compelling results for constructing an informational narrative of eDW. Our first significant finding concerns the relationship between the change in intra-environmental correlation caused by the

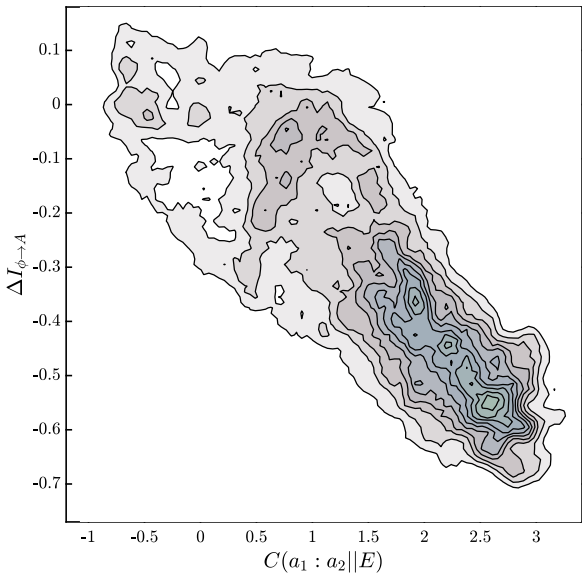

Figure 3. Kernel density of the distribution of cooperation between the biosphere dof (Equation (12)) and the change in intra-environmental correlations (Equation (14)) caused by the presence of the biosphere at high viability, $V > 0.7$. The drop in environmental correlation is strongly related to the robustness of the biosphere.

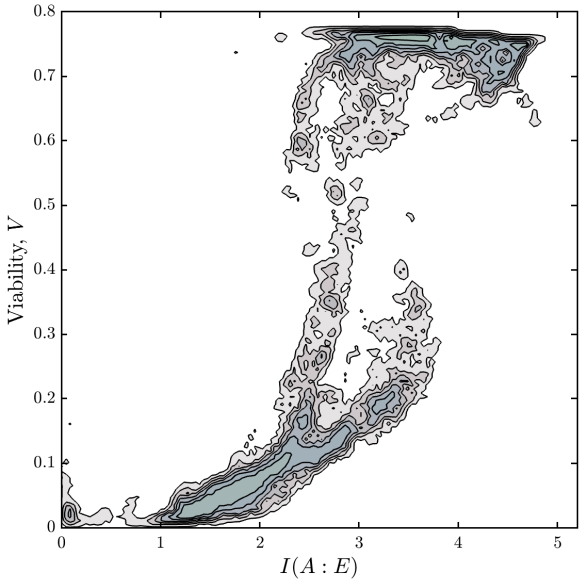

Figure 4. Kernel density of the distribution of viability and total agent–environment dof correlations. Most of the data points are in the two clusters observed, so in order to accentuate structure, the contours are logarithmically spaced.

presence of the agent, Equation (12), and the cooperation between the individual agent dof and all environmental dof, Equation (14). Figure 3 presents the kernel density computed across all simulations with high viability, specifically those with $V > 0.7$. The values on the y-axis are predominantly negative, indicating that the introduction of the biome tends to decorrelate the environmental dof from each other. In contrast, the x-axis values are mainly positive, suggesting that the cooperation between the daisy fractions creates redundant correlations with the environment. Taken together, the figure reveals how correlations are managed by the biome, i.e., the informational architecture of eDW. Maintaining high viability is closely tied to the biome's ability to convert environmental correlations into redundant correlations with itself. The more the planetary temperature becomes decoupled from the stellar luminosity, the more constrained the agent's dof become. Homeostatic regulation is the rearrangement of correlations due to the agents wrangling for control of environmental conditions: the larger the implemented changes are required to be, the stronger the redundancy needed to implement them. This finding elegantly explains the edge effects observed in Figure 1, where planetary temperature fluctuations become much more constrained at the cooler and warmer boundaries of the bearable temperature range.

Our second finding, highlighted in the Introduction, is the canonical SIT viability curve shown in Figure 4. The horizontal axis represents the mutual information, Equation (11), between the full set of agent dof and environmental dof, while the vertical axis displays the viability, Equation (9). The kernel density reveals distinct clustering in the upper right and lower left regions of the parameter space. These clusters align with the viability curve predicted by SIT (A. Kolchinsky & D. H. Wolpert 2018). The cluster in the upper right forms a viability plateau, indicating a set of agent phase space trajectories that achieve maximal viability, unperturbed by additional correlations with the environment. As we move toward the left, the sharp decline in viability reflects an information threshold, analogous to the minimal amount of agent–environment correlation required to sustain viability, as described in SIT.

We use the term analogous intentionally here, as there appear to be two sharp increases in viability, occurring at around 2.5 and 3.5 bits. It is important to note that the precise numerical values are not significant, as they depend on the binning used to estimate the joint distribution. In this case, the maximum mutual information is approximately at $\log_2 25 \approx 4.64$ bits, so the observed thresholds occur at roughly $\sim$0.5 and $\sim$0.75 of the maximum. It is possible that these thresholds represent a form of hysteresis, where the higher threshold corresponds to the minimum correlation required to establish a viable biome, and the lower threshold corresponds to the minimum correlation needed to sustain that biome. While we do not assert the presence of a strict semantic threshold due to the absence of a full counterfactual analysis, we recognize both the thresholds and the viability plateau as important features of eDW's informational architecture.

6. Discussion

Daisy World is a simplified model illustrating how a biosphere exerts homeostatic control over its host planet in response to changes in an external astrophysical driver, such as increasing stellar luminosity. As such, it has played a critical role in debates surrounding Gaia theory and the extent to which biospheres can coevolve with their planets to provide regulatory feedback that maintains habitable conditions. Initially developed to study Earth's history, Gaia theory's implications have gained renewed relevance in astrobiology. Astrobiologically oriented ExoGaia models (A. Chopra & C. H. Lineweaver 2016; A. E. Nicholson et al. 2018) propose that long-term habitability necessitates sustained inhabitation, meaning that Gaian feedback mechanisms must be established for biospheres to persist on geological timescales.

This longevity is crucial for astronomers, as it increases the likelihood of detecting biosignatures across interstellar

distances. Indeed, more recent work using the ExoGaia framework has provided potentially important planetary context measures for assessing biosignature viability. These include the development of a "Gaia Habitable Zone," where appropriate feedbacks can emerge to support long-term inhabitation, as well as new probabilistic methods based on planetary feedback for evaluating claimed biosignature detections (R. Arthur & A. Nicholson 2023; A. E. Nicholson & N. J. Mayne 2023; R. Arthur et al. 2025). Thus, Daisy World models, as fundamental tools for exploring Gaian feedback, remain relevant to both Earth science and astrobiology decades after their introduction.

In this work, we introduce a novel approach to Daisy World models using SIT, a framework that allows us to explore what we refer to as *informational narratives* in living systems. SIT focuses on correlations between an agent and its environment that contribute to the agent's viability. In this paper, we introduce a simplified version of the SIT framework to develop a Daisy World model capable of incorporating information-theoretic measures, such as mutual information. Our eDW model assumes that stellar luminosity fluctuates stochastically at the few percent level over timescales comparable to the growth rates of daisies. While this assumption situates the model within parameter regimes reminiscent of M dwarf habitable zone planets, this association is not essential to the conclusions of our study.

The stochastic eDW model allowed us to investigate the information, or correlational structure, within and between the biosphere and planetary/stellar dof. The biosphere's dof are represented by the two daisy populations, while the latter dof include temperature and stellar luminosity. In simulations of planets lacking biospheres, the joint dof distributions revealed strong correlations between temperature and luminosity, consistent with the Stefan–Boltzmann law. However, with the inclusion of a biosphere, these correlations between planetary dof begin to break down as stellar luminosity increases. As daisy species proliferate, the biosphere effectively *steals* correlations from the environmental dof by exerting rein controls over them. These changes reflect the flow of information between the biosphere and planetary system. We defined interaction information as a measure of cooperation between daisy species. This allowed us to identify a semantic threshold in the biosphere–environment dynamics, which marks the total mutual information load necessary for the biosphere to maintain viability over timescales corresponding to the (average) increase in stellar luminosity. We note again that we explicitly do not carry out the final step of the full SIT program, which involves generating an ensemble of counterfactual trajectories by scrambling the initial mutual information between the agent and environment.

While homeostatic feedback in Daisy World is typically described in terms of physical quantities (e.g., radiation fluxes, albedos, and coverage fractions), our work offers a complementary perspective. Alongside the physical narrative, our study demonstrates how Daisy World evolution can be recast as an informational narrative. Information flows between the environment and the biosphere, with the biosphere "using" that information to establish correlations that enhance its viability by regulating the environment. Crucially, not all of this information is meaningful for maintaining control. By analyzing the structure of information-theoretic quantities (e.g., joint distributions, mutual information, and interaction information), we can identify where, how, and how much of this information contributes to the biosphere's viability and resilience.

Given that Daisy World is fundamentally a toy model, we do not claim that our analysis reveals new dynamic features. Rather, our work highlights the potential of applying information-theoretic approaches like SIT to models of coevolution between biospheres and their planets. More realistic models would need to account for the complex networks of interactions between the living and nonliving planetary subsystems. These include the atmosphere, hydrosphere, cryosphere, lithosphere, and, most critically, biosphere, all of which exert intricate webs of feedback on the planetary system. Because the biosphere, as a living system, processes information in ways nonliving systems do not, information-theoretic approaches can capture aspects of coupled system dynamics that might otherwise remain invisible to purely physical or chemical models.

With regard to future enhancements and additions to the model, a first step would be to reframe it in terms of top-of-the-atmosphere conditions, where outgoing radiation is released. This would allow explicit connections to observables in exoplanet studies. Remaining within the framework of the Daisy World model, it would also be of interest to explore versions in which individual daisies possess agential properties such as memory, control systems, or sensory-motor loops.

In terms of astrobiology, our ultimate interest is the detection of robust biosignatures. There are several ways in which this study may serve as a foundation for fruitfully integrating information narratives into astrobiological research. First, it may help in developing an understanding of the mechanisms through which biospheres (and potentially technospheres) maintain a planet's *inhabitance* over gigayear timescales. As noted above, the ExoGaia thesis asserts that without biospheric feedbacks on the geospheres, planets would drift out of habitability over time (A. Chopra & C. H. Lineweaver 2016; A. E. Nicholson et al. 2018). Applying informational narratives to more complex models of biospheric evolution could clarify both the mechanisms and dynamics underlying such regulatory feedback. Another promising direction is the eventual identification of novel information-theoretic metrics that could inform biosignature detection.

Thus, similar to the application of *information-theoretic methods*—notably Jensen–Shannon divergence—within a network-structured setting (S. Vannah et al. 2023), if informational narratives prove essential in describing biosphere–planetary control (i.e., Gaian or semi-Gaian regulation), they could lead to the identification of new classes of agnostic biosignatures.

As a result, the next step in our research program will involve applying SIT and other information-theoretic approaches to more complex models of coupled planetary systems. There are a number of ways such a research strategy might be carried forward. Biospheric evolution on Earth typically unfolds over timescales far longer than a century, making it unlikely that the information-theoretic methods used here could characterize such evolution directly on exoplanets. However, there are other, shorter timescales over which biospheres may exert strong influence on their geospheres. In particular, seasonal variations might carry imprints of biospheric activity that information-theoretic measures could help detect. The seasonal variation of CO_2 on Earth—driven by

changes in photosynthetic and respiration activity—is suggestive in this regard. One could imagine obtaining spectral observations of exoplanetary atmospheres across different seasons, characterizing variations in multiple chemical species. The mutual information carried by these variations, in conjunction with other planetary characteristics (e.g., albedo), could potentially reveal correlations—or even biospheric controls—analogous to those we have explored in this work.

Pursuing this approach would likely require comparative studies of solar system bodies such as Venus, Earth, Mars, and Titan to distinguish between information architectures associated with purely abiotic feedback mechanisms (e.g., carbonate–silicate cycles) and those arising from biospheric feedbacks (A. Kleidon 2010). This may also entail coupling information-theoretic and thermodynamic measures, as dissipation rates across planetary systems might serve as additional indicators of biospheres—or even technospheres (A. Frank et al. 2022).

We note that this work touches on broader foundational issues related to the use of *information narratives* in emerging studies on the physics of life. While our principal interest lies in understanding how biospheres exert control over the geospheres in which they are embedded, our work also raises questions about the extent to which biospheres exhibit agential characteristics. We thus recognize biospheres as complex adaptive systems that are thermodynamically driven into metastable, far-from-equilibrium steady states. Their ability to harness free energy gradients—initially delivered via stellar radiation—links them to broader information–thermodynamic theories of living systems (e.g., functional information; S. I. Walker & P. C. Davies 2013), including those aiming to describe agential cognition, such as Friston's free energy principle (K. Friston 2010). If the information narrative perspective developed here proves useful as we explore more realistic models of inhabited planetary evolution, we expect it will motivate further study of how such evolution might be interpreted as an increase in what has recently been termed "planetary intelligence."

Acknowledgments

D.S. would like to thank Stuart Bartlett, Matthew Egbert, Alice Quillen, Dominique Segura-Cox, and Michael Wong for stimulating discussions. The authors thank the Center for Integrated Research Computing (CIRC) at the University of Rochester for providing computational resources and technical support. This project was partly made possible through the support of grant 62417 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

Appendix A
Analytical Derivations

A.1. Temperature Dynamics

Here we derive Equation (6), which describes the dynamics of the planetary surface temperature, and the equilibration timescale. Imagine a star with luminosity L that harbors an exoplanet at an orbital radius of r. The exoplanet has a thin atmosphere of scale height h much smaller than its radius, R_p. This shell of gas, whose specific heat is c_V, absorbs the star's light and heats up. Simultaneously, it emits photons into the void, causing it to cool down. The surface temperature equation follows from accounting for these fluxes.

It is safe to assume that the exoplanet is spherical, with a cross section of πR_p^2 facing the star. Since the orbital radius is r, and the star radiates isotropically, the fraction of the star's power output impingent on the exoplanet is

$$P_{\rm in} = \frac{\pi R_p^2}{4\pi r^2}(1 - A)L, \tag{A1}$$

where A is the albedo of the exoplanet: the percentage of impinging starlight that reflects away instead of being absorbed.

The instantaneous temperature at time t of the atmosphere is $T = T(t)$. Assume that the exoplanet radiates away energy isotropically across its entire surface area like a blackbody. The power output of the exoplanet follows from the Stefan–Boltzmann law,

$$P_{\rm out} = \sigma 4\pi R_p^2 T^4, \tag{A2}$$

where σ is the eponymous constant. Let E be the internal energy of the atmosphere. In a short time interval, Δt, any change in this energy is the difference between the energy flowing onto and out of the exoplanet,

$$\Delta E = P_{\rm in}\Delta t - P_{\rm out}\Delta t. \tag{A3}$$

Meanwhile, the first law of thermodynamics tells us that changes in internal energy are due to a difference in heat flowing into the atmosphere and the work done by the atmosphere, $\Delta E = \Delta Q - P\Delta V$. If we assume that the atmosphere does no work and undergoes no phase changes, then its temperature changes according to $\Delta Q = mc_V\Delta T$. Here m is the mass of the atmosphere, which can be approximated by multiplying the average atmospheric density, ρ, by the atmosphere's approximate volume, $4\pi R_p^2 h$. Putting all this together, we have

$$4\pi R_p^2 h\rho c_V \Delta T = \frac{\pi R_p^2}{4\pi r^2}(1 - A)L\Delta t - \sigma 4\pi R_p^2 T^4 \Delta t \tag{A4}$$

$$\Downarrow \qquad \text{Dividing through by } 4\pi R_p^2 h\rho c_V \Delta t \tag{A5}$$

$$\frac{\Delta T}{\Delta t} = \frac{1 - A}{16\pi r^2 h\rho c_V}L - \frac{\sigma}{h\rho c_V}T^4. \tag{A6}$$

Finally, taking the limit $\Delta t \to 0$, we arrive, as promised, at Equation (6).

A.2. Environmental Timescale

In this section, we derive the environmental timescale through a fixed-point analysis. The fixed point of Equation (6), which defines the equilibrium temperature, $T_{\rm eq}$, is nothing more than the temperature constraint, Equation (3),

$$T_{\rm eq}^4 = \frac{1 - A}{16\pi r^2\sigma}L. \tag{A7}$$

Assume that the stellar luminosity L remains constant. Consider an exoplanet that is at $T_{\rm eq}$ and imagine that its temperature fluctuates to $T(t) = T_{\rm eq} + \delta T(t)$, where $\delta T/T_{\rm eq} \ll 1$ will serve as our expansion parameter. Plug the temperature

into Equation (6),

$$\frac{d\delta T}{dt} = \frac{1-A}{16\pi r^2 h\rho c_V}L - \frac{\sigma}{h\rho c_V}T_{\rm eq}^4\left(1+\frac{\delta T}{T_{\rm eq}}\right)^4. \tag{A8}$$

Next expand the binomial, keeping only the term linear in the expansion parameter:

$$\frac{d\delta T}{dt} = \frac{1-A}{16\pi r^2 h\rho c_V}L - \frac{\sigma}{h\rho c_V}T_{\rm eq}^4(1+4\frac{\delta T}{T_{\rm eq}}) \tag{A9}$$

$$=\left(\frac{1-A}{16\pi r^2 h\rho c_V}L - \frac{\sigma}{h\rho c_V}T_{\rm eq}^4\right) - \frac{4\sigma}{h\rho c_V}T_{\rm eq}^3\delta T \tag{A10}$$

$$=-\frac{4\sigma T_{\rm eq}^3}{h\rho c_V}\delta T. \tag{A11}$$

Since the fraction in the last line is positive definite, the minus sign tells us that the fixed point is stable: the fluctuation decays. Furthermore, it has units of inverse time, motivating the definition of the environmental timescale:

$$\tau_E = \frac{h\rho c_V}{4\sigma T_{\rm eq}^3}. \tag{A12}$$

This in hand, the solution for the dynamics of the fluctuation is

$$\delta T(t) = \delta T(0)e^{-t/\tau_E}, \tag{A13}$$

from which we explicitly see the fluctuation decaying away.

Appendix B
Simulation Methodology

B.1. Dimensionalization

eDW is a dynamical system of 4 dof, $\{f_B, f_W, T, L\}$, whose evolution is described by

$$\frac{df_B}{dt} = \beta(T_W)(f-f_B-f_W)f_B - \gamma_D f_B, \tag{B1}$$

$$\frac{df_W}{dt} = \beta(T_B)(f-f_B-f_W)f_W - \gamma_D f_W, \tag{B2}$$

$$\frac{dT}{dt} = \frac{1}{16\pi r^2 h\rho c_V}L(1-A_G-(A_B-A_G)f_B-(A_W-A_G)f_W) - \frac{\sigma}{h\rho c_V}T^4, \tag{B3}$$

$$\frac{dL}{dt} = \frac{1}{\tau_s}(\langle L\rangle - L) + \sqrt{\frac{2}{\tau_s}}\,\delta\langle L\rangle\,\eta, \tag{B4}$$

$$\text{where}\quad \beta(T) = \gamma_G e^{-\left(\frac{T-T_{\rm opt}}{\Delta T}\right)^4} \tag{B5}$$

$$\text{and}\ T_\alpha^4 = T^4 + q(A_B-A_G)f_B + q(A_W-A_G)f_W - q(A_\alpha-A_G). \tag{B6}$$

The 16 parameters of varying dimensions, $\{f, A_G, A_B, A_W, q, \gamma_G, \gamma_D, \tau_*, r, h, \rho, c_V, T_{\rm opt}, \Delta T, \delta, \langle L\rangle\}$, fully specify the model. Dimensionalizing reduces this parameter space, greatly simplifies the equations of motion, and makes the system amenable to numerical simulation, which this Appendix covers.

To this effect, we scale all temperatures to $T_{\rm opt}$, then define the optimal luminosity $L_{\rm opt} = 16\pi r^2\sigma T_{\rm opt}^4/(1-A_G)$ and use it to scale the luminosity. This lets us parameterize the average stellar luminosities as $\langle L\rangle = (1+\lambda)L_{\rm opt}$, where $\lambda \in (-1, \infty)$. The environmental timescale is $\tau_E = c_V\rho h/\sigma T_{\rm opt}^3$. Time is scaled to the biosphere timescale γ_G^{-1}. Lastly, we introduce the dimensionless growth window,

$$w(x) = e^{-\alpha x^4}, \tag{B7}$$

where $\alpha = 8T_{\rm opt}^4/\Delta T^4$, the scaled temperature deviation $Q = q/T_{\rm opt}^4$, and the shorthand $\delta A_\alpha = A_\alpha - A_G$.

The dimensionalized equations of motion are now

$$\frac{df_B}{dt} = w(T_W-1)(f-f_B-f_W)f_B - \frac{\gamma_D}{\gamma_G}f_B, \tag{B8}$$

$$\frac{df_W}{dt} = w(T_B-1)(f-f_B-f_W)f_W - \frac{\gamma_D}{\gamma_G}f_W, \tag{B9}$$

$$\frac{dT}{dt} = \frac{1}{\gamma_G\tau_E}\left(1-\frac{\delta A_B f_B + \delta A_W f_W}{1-A_G}\right)L - \frac{1}{\gamma_G\tau_E}T^4, \tag{B10}$$

$$\frac{dL}{dt} = \frac{1}{\gamma_G\tau_s}(1+\lambda-L) + \sqrt{\frac{2}{\gamma_G\tau_s}}\,\delta(1+\lambda)\eta, \tag{B11}$$

$$\text{where}\quad T_\alpha^4 = T^4 + Q\delta A_B f_B + Q\delta A_W f_W - Q\delta A_\alpha. \tag{B12}$$

We collect the remaining 11 dimensionless parameters and write them as the parameter vector

$$\boldsymbol{\theta}^T = (f, A_G, \delta A_B, \delta A_W, Q, \frac{\gamma_D}{\gamma_G}, \gamma_G\tau_E, \gamma_G\tau_s, \alpha, \delta, \lambda) \tag{B13}$$

and the dimensionless dof as the vector

$$\boldsymbol{x}^T = (f_1, f_2, T, L). \tag{B14}$$

This prescription makes it apparent that eDW is a vector process described by the stochastic differential equation

$$d\boldsymbol{x} = \boldsymbol{a}(\boldsymbol{x}|\boldsymbol{\theta})dt + \boldsymbol{b}(\boldsymbol{x}|\boldsymbol{\theta})dW, \tag{B15}$$

where W is a one-dimensional Wiener process.

B.2. Numerical Integration

Standard numerical integration methods such as Euler or Runge–Kutta must be modified nontrivially for stochastic differential equations (S. Särkkä & A. Solin 2014). These modifications can make a stochastic RK2 method significantly more complicated, and tedious, to implement than a much higher nonstochastic Runge–Kutta method (A. Rößler 2009). Fortunately for us, Equation (B15) does not have the general form of a multivariate stochastic differential equations—the Wiener process is scalar—allowing us to use a strong RK1 method (A. Roberts 2012).

Algorithm 1. Stochastic Runge–Kutta of Strong Order 1

Require $\boldsymbol{X}(t)$, $\boldsymbol{\theta}$, Δt
Ensure $\boldsymbol{X}(t+\Delta t)$
  $\boldsymbol{X} \leftarrow \boldsymbol{X}(t)$
  Draw $Z \sim \mathcal{N}(0, 1)$
  Draw $S \sim \mathcal{U}\{\pm 1\}$
  $\boldsymbol{k}_1 \leftarrow \boldsymbol{a}(\boldsymbol{X}, \boldsymbol{\theta})\Delta t + (Z - S)\boldsymbol{b}(\boldsymbol{X}, \boldsymbol{\theta})\sqrt{\Delta t}$
$\boldsymbol{k}_2 \leftarrow \boldsymbol{a}(\boldsymbol{X} + \boldsymbol{k}_1, \boldsymbol{\theta})\Delta t + (Z + S)\boldsymbol{b}(\boldsymbol{X} + \boldsymbol{k}_1, \boldsymbol{\theta})\sqrt{\Delta t}$
  $\boldsymbol{X}(t+\Delta t) \leftarrow \boldsymbol{X} + \frac{1}{2}(\boldsymbol{k}_1 + \boldsymbol{k}_2)$

B.3. Statistics

The dimensionless parameters chosen for all the simulations are in the upper part of Table 1. The parameters that differ for each simulation, the average stellar luminosity, $(1 + \lambda)L_{\rm opt}$, and growth rate bandwidth, ΔT, are in the lower half of the table. The stellar luminosities span an order of magnitude from $0.3L_{\rm opt}$ to $2.4L_{\rm opt}$, while the bandwidths span 2 orders of magnitude from $0.002T_{\rm opt}$ to $0.267T_{\rm opt}$. The former range is sampled at 400 evenly spaced luminosities and the latter at 128 evenly spaced bandwidths, resulting in 51 and 200 simulations, each at a fixed average stellar luminosity and growth rate bandwidth. An additional 400 simulations are run, one at each fixed stellar luminosity with the agent-free model.

Each simulation samples the joint distribution, $p_{\rm AE}$ or p_0, by running 500 independent instances of eDW with the following initial conditions. For the former distributions, f_B is chosen uniformly at random in the range $[0, f]$ for each instance, and then $f_W = f - f_B$ is set. For the latter distributions, both $f_B = f_W = 0$ are set. The luminosity is initialized to the average stellar luminosity, and the temperature is initialized to the corresponding equilibrium temperature, Equation (3). A dimensionless time step of $dt = 0.1$ is used to integrate each instance 2000 times, using the strong stochastic Runge–Kutta method of order 1 described earlier, then recording the values of the $4/2$ dof. This procedure ensures that any transients have decayed away and that the individual recordings making up the ensembles are not correlated.

Each ensemble is then used to estimate the corresponding joint distribution. Since the dof take on real values, we discretize their ranges using a modified square root rule: given N samples, we subdivide the corresponding range into $1 + \lceil\sqrt{N}\rceil$ equally sized bins. For the agent-free ensembles, this means that the temperature and luminosity ranges, $[\min(T), \max(T)]$ and $[\min(L), \max L]$, respectively, are divided into $1 + \lceil\sqrt{500}\rceil = 25$ bins. For $p_{\rm AE}$, the number of instances, N, with $f_B > 0$ and $f_W > 0$ is used to subdivide the four analogous ranges (which now include the pair $[\min(f_\alpha), \max(f_\alpha)]$) into $1 + \lceil\sqrt{N}\rceil$ bins. The empirical distributions are then constructed by counting the number of instances in each bin and normalizing. These are our estimates of the joint distributions, which are used to compute the expectation value needed to estimate the viability, Equation (9), as well as our corpus of information measures, namely, $I(A\colon E)$, $\Delta I_{\phi\to A}$, and $C(a_1\colon a_2||E)$. Every marginal of the joint distributions is found, and the corresponding naive estimator of the marginal entropy is computed. These are then used to construct estimates of all the information measures used in the analysis.

Table 1
Simulation Dimensionless Parameters

Parameter	Symbol	Value
Habitable land fraction	f	0.88
Ground albedo	A_G	0.3
Black daisy albedo difference	δA_B	−0.2
White daisy albedo difference	δA_W	0.3
Temp. shift heat constant	Q	0.1
Decay–to–growth rate ratio	γ_D/γ_G	0.2
Environmental timescale	$\gamma_G\tau_E$	5.0
Stellar fluctuation timescale	$\gamma_G\tau_s$	3.0
Stellar fluctuation scale	δ	0.05
Average stellar luminosity	λ	−0.7 to 1.4
Growth rate bandwidth	$\Delta T/T_{\rm opt}$	0.002–0.267

All simulations were run in parallel on the Blue Hive cluster at the Center for Integrated Computational Research at the University of Rochester. Since the instances within the ensembles were run in series, the total computation time per simulation ranged from approximately 10 minutes to 8 hr. Analysis was performed on a 2020 Macbook Pro with Apple's M1 processor, using MATLAB (T. M. Inc.* 2022) and Mathematica (Wolfram Research, Inc. 2024).

ORCID iDs

Damian R Sowinski https://orcid.org/0000-0003-3136-7449

References

Acebrón, J. A., Bonilla, L. L., Pérez Vicente, C. J., Ritort, F., & Spigler, R. 2005, RvMP, 77, 137
Arthur, R., & Nicholson, A. 2023, MNRAS, 521, 690
Arthur, R., Nicholson, A. E., & Mayne, N. J. 2025, arXiv:2504.18431
Barandiaran, X. E., & Egbert, M. D. 2014, Artificial Life, 20, 5
Barham, J. 1996, BiSys, 38, 235
Bartlett, S., Eckford, A. W., Egbert, M., et al. 2025, arXiv:2501.08683
Blain, J. C., & Szostak, J. W. 2014, Annu. Rev. Biochem., 83, 615
Chen, Z., Kelty, S., Evsukoff, A. G., et al. 2022, NatCo, 13, 1922
Chopra, A., & Lineweaver, C. H. 2016, AsBio, 16, 7
Cockell, C. S., Simons, M., Castillo-Rogez, J., et al. 2024, NatAs, 8, 30
Cohen, J. E. 1977, Natur, 270, 165
Cooper, F., Ghoshal, G., Pawling, A., & Pérez-Mercader, J. 2013, PhRvL, 111, 044101
Corning, P. A. 2007, Syst. Res. Behav. Sci., 24, 297
Cover, T. M., & Thomas, J. A. 2006, Elements of Information Theory (Hoboken, NJ: Wiley-Interscience)
Dawkins, R. 1982, The Extended Phenotype: The Long Reach of the Gene (Oxford: Oxford Univ. Press)
Deacon, T. W. 2007, Cognitive Semiotics, 1, 123
Doolittle, W. F. 2017, JThBi, 434, 11
Egbert, M., Hanczyc, M. M., Harvey, I., et al. 2023, OLEB, 53, 87
Egbert, M. D., & Pérez-Mercader, J. 2018, Artif. Life, 24, 106
Frank, A., Grinspoon, D., & Walker, S. 2022, IJAsB, 21, 47

Friston, K. 2010, Nat. Rev. Neurosci., 11, 127
Garcìa-Selfa, D., Ghoshal, G., Bick, C., Pérez-Mercader, J., & Muñuzuri, A. P. 2021, CSF, 145, 110809
Gleiser, M., & Sowinski, D. 2018, The Map and the Territory: Exploring the Foundations of Science, Thought and Reality (Berlin: Springer), 141
Henry, T. J., & Jao, W.-C. 2024, ARA&A, 62, 593
Ibánez, M., Dieball, C., Lasanta, A., Godec, A., & Rica, R. A. 2024, NatPh, 20, 135
Inc.*, T. M. 2022, MATLAB, 9.13.0 (R2022b) (Natick, MA: The MathWorks Inc.) https://www.mathworks.com
James, R. G., Ellison, C. J., & Crutchfield, J. P. 2011, Chaos, 21, 037109
Kleidon, A. 2010, PhLRv, 7, 424
Kolchinsky, A., & Wolpert, D. H. 2018, Interface Focus, 8, 20180041
Krakauer, D. 2024, The Complex World: An Introduction to the Foundations of Complexity Science (Santa Fe, NM: Santa Fe Institute Press)
Kullback, S., & Leibler, R. A. 1951, Ann. Math. Stat., 22, 79
Levins, R. 1969, Bulletin of the Entomological Society of America, 15, 237
Lingam, M., & Balbi, A. 2024, From Stars to Life: A Quantitative Approach to Astrobiology (Cambridge: Cambridge Univ. Press)
Lingam, M., & Loeb, A. 2021, Life in the Cosmos: From Biosignatures to Technosignatures (Cambridge, MA: Harvard Univ. Press)
Lovelock, J. E., & Margulis, L. 1974, Tell, 26, 2
Lovelock, J. E., & Watson, A. J. 1982, P&SS, 30, 795
Margulis, L., & Lovelock, J. E. 1974, Icar, 21, 471
Martínez, R. R., Lopez, L. A., Shappee, B. J., et al. 2020, ApJ, 892, 144
Meadows, V. S., Arney, G. N., Schmidt, B. E., & Des Marais, D. J. 2020, Planetary Astrobiology (Tucson, AZ: Univ. Arizona Press)
Mimar, S., Juane, M. M., Mira, J., et al. 2021, PhRvR, 3, 023241
Mimar, S., Juane, M. M., Park, J., Muñuzuri, A. P., & Ghoshal, G. 2019, PhRvE, 99, 062303
Mossio, M., Saborido, C., & Moreno, A. 2009, Br. J. Philos. Sci., 60, 813
Nehaniv, C. L., Polani, D., Dautenhahn, K., te Beokhorst, R., & Cañamero, L. 2002, in Proc. Eighth Int. Conf. Artificial Life, ICAL 2003 (Cambridge, MA: MIT Press), 345
Nicholson, A. E., & Mayne, N. J. 2023, MNRAS, 521, 5139
Nicholson, A. E., Wilkinson, D. M., Williams, H. T. P., & Lenton, T. M. 2018, MNRAS, 477, 727
Onori, L., & Visconti, G. 2012, Rendiconti Lincei, 23, 375
Pearl, J. 2009, Causality (Cambridge: Cambridge Univ. Press)
Pemartín, I. G.-A., Mompó, E., Lasanta, A., Martín-Mayor, V., & Salas, J. 2024, PhRvL, 132, 117102
Polani, D., Martinez, T., & Kim, J. T. 2001, An Information-theoretic Approach for the Quantification of Relevance (Berlin: Springer)
Roberts, A. 2012, arXiv:1210.0933
Rooney, N., McCann, K., Gellner, G., & Moore, J. C. 2006, Natur, 442, 265
Rößler, A. 2009, SJNA, 47, 1713
Salaris, M., & Cassisi, S. 2005, Evolution of Stars and Stellar Populations (New York: Wiley)
Särkkä, S., & Solin, A. 2014, Applied Stochastic Differential Equations, Vol. 10 (Cambridge: Cambridge Univ. Press)
Savi, M. A., & Viola, F. M. 2023, Fractal and Fractional, 7, 190
Schlosser, G. 1998, Synthese, 116, 303
Schreiber, T. 2000, PhRvL, 85, 461
Shannon, C. E. 1948, BSTJ, 27, 379
Sowinski, D. R. 2016, Complexity and Stability for Epistemic Agents: The Foundations and Phenomenology of Configurational Entropy (Hanover, NH: Dartmouth College)
Sowinski, D. R., Carroll-Nellenback, J., Markwick, R. N., et al. 2023, PRXL, 1, 023003
Sowinski, D. R., Frank, A., & Ghoshal, G. 2024, PhRvR, 6, 043188
Steffen, W., Richardson, K., Rockström, J., et al. 2020, NRvEE, 1, 54
Tejero, Á., Sánchez, R., Kaoutit, L. E., Manzano, D., & Lasanta, A. 2025, PhRvR, 7, 023020
Thompson, E., & Stapleton, M. 2009, Topoi, 28, 23
Tuomi, M., Jones, H. R., Butler, R., et al. 2019, arXiv:1906.04644
Tyrrell, T. 2013, On Gaia: A Critical Investigation of the Relationship Between Life and Earth (Princeton, NJ: Princeton Univ. Press)
Vannah, S., Gleiser, M., & Kaltenegger, L. 2023, MNRAS: Lett., 528, L4
Walker, S. I., & Davies, P. C. 2013, JRSI, 10, 20120869
Walker, S. I., Kim, H., & Davies, P. C. 2016, RSPTA, 374, 20150057
Watson, A. J., & Lovelock, J. E. 1983, TellB, 35, 284
Williams, P. L., & Beer, R. D. 2010, arXiv:1004.2515
Wolfram Research, Inc. 2024, Mathematica, v14.1 (Champaign, IL: Wolfram Research, Inc.) https://www.wolfram.com/mathematica
Wood, A. J., Ackland, G. J., Dyke, J. G., Williams, H. T., & Lenton, T. M. 2008, RvGeo, 46, RG1001
Xavier, J. C., Hordijk, W., Kauffman, S., Steel, M., & Martin, W. F. 2020, RSPSB, 287, 20192377
Zalasiewicz, J., Williams, M., Waters, C. N., et al. 2017, AntRv, 4, 9