The Circular Logic of Gaia: Fragility and Fallacies, Regulation and Proofs

Inman Harvey

Evolutionary and Adaptive Systems Group, University of Sussex, Brighton, UK
inmanh@gmail.com

Abstract

Gaia Theory makes controversial systems-level claims about how environmental factors on a planet tend to be regulated towards conditions favourable to biota. Daisyworld models show how such phenomena may arise, but their subtleties are often misunderstood. Here the core concept of Gaian regulation is shown in an ultra-minimalist model, and defined in terms of fragile hysteresis loops. Insights thus gained explain many fallacies and misunderstandings held by critics and advocates alike. Gaia Theory and Darwinian evolution are shown to be complementary not antagonistic. General results are proved for the inevitability of Gaian regulation at steady state in systems of arbitrary size and complexity under very broad conditions.

Introduction

Gaia Theory makes controversial claims about system-level properties of interactions between biota and environment in a bounded world. In some sense, it is claimed, environmental factors tend to be regulated favourably to the biota – but in what sense, and why? What is being regulated? Is something (what?) being optimised? Is it compatible with evolution? Does it depend on a lucky (or deliberate) choice of bio-environmental interactions? We build on Daisyworld (DW: Lovelock, 1983; Watson and Lovelock, 1983), a simple Artificial Life model, to answer these questions. The unfamiliar Gaian circular logic is easy to misinterpret, by critics and advocates alike; confusion has bred mistrust on both sides, not least on the relationship between Gaia Theory and evolution. We aim to clarify how such confusion arose.

The strategy of this paper is to first present the core of this circular logic (Fig. 1) in the context of the simplest model possible: a 2-variable, 1-parameter toy, abstract mathematical model. Gaian regulation (GR), defined in terms of Viability-Zones in Perturbation Space, will be demonstrated. This corresponds to a specific form of hysteresis loop, a zone of 'fragile' bistability. The circular causation $B \rightleftarrows E$ upsets many prior expectations, and is used to illustrate common fallacies and misunderstandings. Thus it may act as an intuition pump when considering the real world of biology and environment, or synthetic worlds of man-made systems.

The second part of the paper generalises from this 2-variable model to n-variables and interactions of any complexity, subject only to very broad conditions. GR, thus defined, extends to arbitrary systems of any size: this is 'inevitable Gaian regulation', perturbations are automatically countered so as to widen, not lessen, the viability range.

It is shown that this viability-based GR is indeed in accord with the core of the original DW of W&L, though even there it is necessary to distinguish the core from added complexity that is not essential to GR. Clarification of the core concept

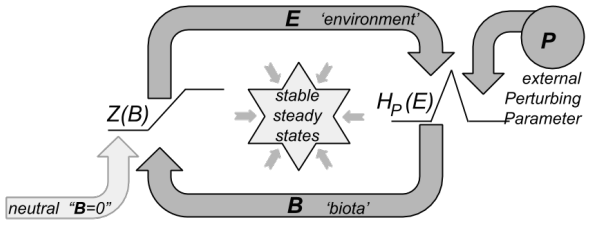

Figure 1. Circular causation: perturbations P regulate the stable steady states of the B⇄E circuit. B is viable (>0) for some range of P values. It will be proven that if the circuit is broken by a hypothetical neutral 'switch' pictured at left, B is viable for a decreased range of P (or at best, the same range). I.e. the unbroken circuit typically increases B-viability.

both makes possible the very general theorem proved here and indicates where translation to the real world may be fruitful. The results pertain to equilibria and not to transient responses.

Principles and Anecdotes. Most studies of DW (Wood, 2008) make extensions to the core so as to show new phenomena or to fit the model better to real data. This paper has the opposite motivation: in dismissing any such extensions as 'merely anecdotal' insofar as they are not generalisable, it aims to find universal principles. Even the original DW has much removed here. This may mean that GR as defined here is narrower than others assume; but by clarifying a line between the universal and the specific it may help illuminate in other DW studies just what is core (and generalisable) and what may not be.

The minimal 2-variable model

DW is an Alife-style model first presented by Lovelock in 1982 (published as Lovelock, 1983). The best-known early citation is here referred to as W&L (Watson and Lovelock, 1983). The minimal version here, based on Harvey (2004), radically simplifies the original DW of W&L. We start with a single 'bio-variable' B, and a single 'enviro-variable' E. They take values within finite ranges that we choose to scale as [0.0,1.0]. Together they form a dynamical system, where each affects the rate of change of the other via functions $H(\,)$ and $Z(\,)$, parameterised by P as described further below:

$$\mu \, dB/dt = H_P(E) - B = H(E+P) - B \quad (1)$$

$$\nu \, dE/dt = Z(B) - E \quad (2)$$

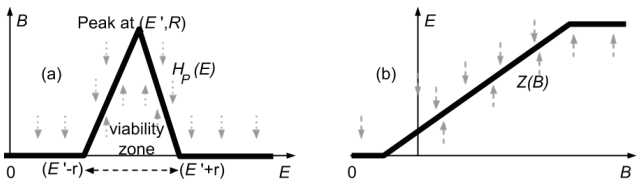

Figure 2. Example nullclines (a) $B=H_P(E)$ and (b) $E=Z(B)$. Directions of dB/dt and dE/dt are indicated by arrows.

μ and v (that we shall typically set to 1) moderate the rates of change of B and E towards the 'nullcline' endstates of $B=H_P(E)$ and $E=Z(B)$. If and when both endstates are reached simultaneously, in the absence of noise further change would cease; we shall be considering the system in the presence of noise that will shift the system from unstable to stable equilibria. Such equilibria points can be seen graphically as where the nullclines intersect; the form of these nullclines is crucial to the model, determining which equilibria are stable and which unstable.

We choose $H_P(E)=H(E+P)$ as a parameterised 'hat-shaped' function that is zero outside some constrained 'E-viability-zone' within which it rises to a peak (Fig. 2a); parameter P shifts the hat left or right. Here we show a piecewise linear 'witches hat'; most of the results that concern us are insensitive to the precise form of the hat. With the peak value R of the hat at $E=E'$, and viability 'radius' r, we have:

$$H(E)=R.max(1-abs(E-E')/r,0) \quad (3)$$

We choose for illustration (Fig. 2b) $Z(B)$ as, in the region where it affects matters, a linear relationship based on $(Q + sB)$; where Q is a fixed default value for E when B=0, and s is the gradient of slope (positive or negative). However this linear form is truncated below and above at E = 0 and 1. This results in a 'zigmoid' or piecewise-linear sigmoid:

$$Z(B)=min(max(Q+s.B,0),1) \quad (4)$$

We may flip Fig. 2b, and overlay on 2a so that the B-E axes now coincide, as shown in Fig. 3, $Z(B)$ now shown dashed. These nullclines intersect at steady states, stable or unstable.

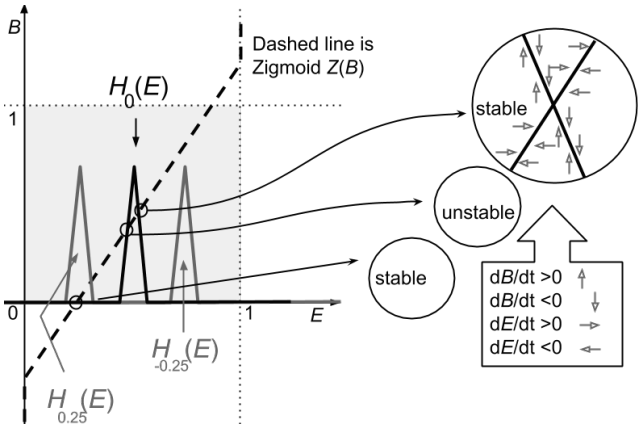

Figure 3. As the hat-function $H_P(E)$ shifts left and right with different P values, there are between 1 and 3 steady states. $H_0(E)$ has 3, circled; examination of the arrows representing dB/dt, dE/dt, indicate they are stable/unstable/stable.

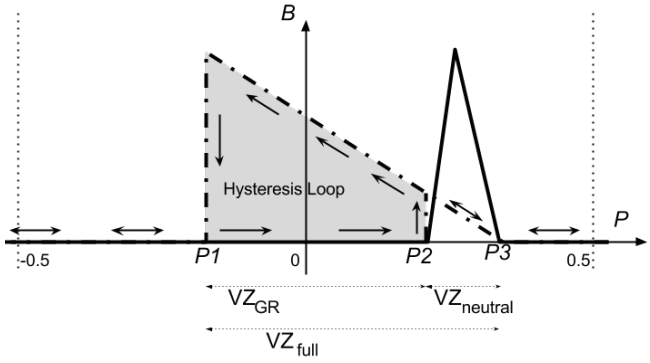

Figure 4. Values of B for the stable steady states of the BE system as P varies. Note that the axes are now B against P.

When parameter $P=0$, examination of the signs of dB/dt and dE/dt (Fig. 3, for $H_0(P)$), shows that the rightmost of 3 steady states is stable; likewise the leftmost is stable. The steady state between these on the nullcline is necessarily unstable. Hence if we assume there is noise in the system such that it leaves the unstable one, after transients the only (stable) steady states to be seen are as shown: one with B zero, one with B positive.

We plot these stable steady states in Fig. 4, as P is varied. For $P1<P<P2$ there are 2 possible values for B, providing a hysteresis loop: as P is increased from low values, the lower part of the loop will be followed, jumping up at $P2$; as P decreases from high values, the upper part of the loop is followed until a jump down at $P1$ (corresponding to where the nullclines are tangent to each other). We define Viability Zone VZ_{full} (in Fig. 4 from $P1$ to $P3$) as the region of P-space for which there is at least one stable steady state with B>0; i.e. here we consider the *upper* limb of the hysteresis loop.

The Neutral Comparison

The hat-function shown in Fig. 4 represents the values taken by B for the (single) steady state of a 'neutral' version of the system; here we follow W&L. This is what would happen if all the effects of B on E were nullified, i.e. if $Z(B)$ was replaced by $Z(0)$ as indicated by a hypothetical neutral 'switch' in Fig. 1, thus breaking the $B \rightleftarrows E$ circuit. This results in a viability zone $VZ_{neutral}$, where B>0, between $P2$ and $P3$. This width is $2r$, the same as the basic width of $H(P+E)$, see eqn 3; but here it is P that is varying rather than E. We may note that $VZ_{neutral}$ corresponds exactly to VZ_{full} except that it is based on the *lower* (B=0) limb of the hysteresis loop rather than the *upper* limb.

In this example, we can see that VZ_{full} covers the zone $VZ_{neutral}$ and extends further. $VZ_{full} = VZ_{neutral} + VZ_{GR}$, where the latter 'Gaia-Regulated' VZ corresponds exactly to the hysteresis region $P1<P<P2$. We use the existence of such a non-empty VZ_{GR} as the basis for defining Gaian Regulation.

Gaian Regulation

We define GR as occurring whenever there is a zone of Perturbation-space within which a biotic variable B *can* be viable (B>0), despite a neutral version of B (where its effects are nullified) being non-viable (B=0); in other words, where there is this 'fragile' form of bistability with both upper and lower (zero) limbs of a hysteresis loop. We have demonstrated

this in the simple example above, where the functions and parameters were chosen so as to highlight this. We chose a narrow hat-function (small *r* in eqn 3), and a high value for *s*, the slope of the zigmoid (eqn 4); such choices tends to make VZ_{GR} more prominent for illustrative purposes. We note that this definition broadly corresponds to the implied definition in W&L, where *B* refers to daisies. *E* to temperature and *P* to Solar Luminosity. W&L recognised the hysteresis from the start; their Figure 1a clearly shows what is here called $VZ_{neutral}$ for a 'neutral daisy' (that has the same albedo as the ground, i.e. its effects are nullified). Elsewhere in W&L's Figure 1 they show what is called here VZ_{full} for various combinations of daisies, expanding the viability range beyond $VZ_{neutral}$.

Sceptics will of course suspect that these simple examples are cherrypicked to show the phenomenon of GR, and that different examples may show some reverse anti-Gaian effect. In the second part of the paper we show otherwise; within a very broad range of dynamical systems of any size, we show that for any arbitrary variable *B,* $VZ_{neutral} \subseteq VZ_{full}$. In other words if the effects of *B* on its environment are such as to influence its own Viability Zone in any direction, this influence can only be in a positive direction, expanding the range of viability and thus displaying GR. Not lucky Gaia (Kirchner, 2002), but inevitable Gaia.

Various Fallacies and Confusions

Before we get there, we can use the present very simple model, loosely equivalent to black daisies in W&L's DW, to illustrate some commonly held fallacies about Gaia Theory.

The Misattributed-Feedback Fallacy. This is the all-too-easy error of calling, for instance, the effect of E on B a 'negative feedback' rather than a 'negative effect' (or positive, as the case may be). Feedback, -ve or +ve, is a property of a complete ***feedback circuit*** (-fc or +fc) of effects $A \rightarrow B \rightarrow C... \rightarrow A$ and so cannot be attributed to any single effect $A \rightarrow B$. This mistake is made time and again, even in W&L where 'changing the direction of effects' is confused with 'changing the direction of feedbacks'. The present author has erred likewise (e.g. in Harvey, 2004). Although it is strictly true that changing the direction of a single effect in a feedback circuit will – *if* nothing else changes – reverse the sign of that circuit, in the current context there is *always* a consequent further change to counteract this. This sloppiness in language hence leads directly to …

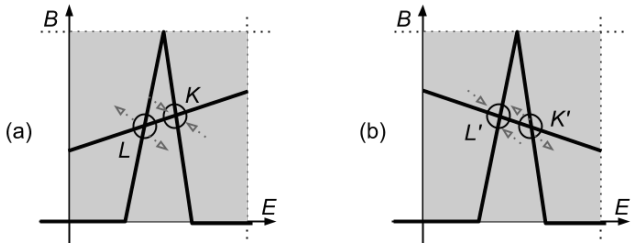

Figure 5. (a) Corresponds to Fig. 3, with -fc at K; this relates to 'black daisies' with a +ve effect. (b) If the slope is reversed ('white daisies'), K becomes K' (+fc) but also L becomes L' (-fc). The stable steady state does not 'disappear' but rather 'shifts elsewhere'.

The Missing-the-point Equilibrium Fallacy (aka Lucky Gaia). Confusing effects with feedbacks misleads the unwary into believing that since, e.g., a +ve effect is associated with +fc in one part of phase space, it will be associated with +fc in other parts of phase space. In Fig. 5 we have a counter-example: *B* has a +ve effect on *E* that contributes to a +fc at unstable equilibrium *L*, yet to a -fc at stable equilibrium *K*.

As indicated in Fig. 5, although swapping the sign of an effect does turn -fc into +fc, it simultaneously turns any neighouring +fc into a new -fc. So state of the dynamical system, in the presence of any noise, will automatically shift through phase space towards another point. The 'Lucky Gaia' criticism takes it as a matter of luck (dependent on directions of +ve or -ve effects) whether a specific steady state is stable or unstable; but one such point becoming 'unlucky' means the next one will become 'lucky'.

The Optimising Gaia Fallacy. It should be clear from the simple example that nothing there can be identified as being 'optimised'. Under Gaian regulation, the peak value of *B* (seen at *P=P1* in Fig. 4) is the same as it was without GR. Further, in the simple example shown, this peak value is right at the tipping-point where regulation is about to be lost. Within $VZ_{neutral}$ the value of *B* is mostly decreased. The Viability Zone is typically *increased* under GR, but that does not translate to its being *optimised* – optimised with respect to what? Suggestions to the contrary, such as:

> We have since defined Gaia as a complex entity involving the Earth's biosphere, atmosphere, oceans, and soil; the totality constituting a feedback or cybernetic system which seeks an **optimal** physical and chemical environment for life on this planet. (Lovelock, 1979)

are better replaced by formulations such as:

> The Gaia hypothesis ... the atmosphere, the oceans, the climate and the crust of the Earth are regulated at a state **comfortable** for life… (Lovelock, 1979) [stresses added]

and **habitable** is more appropriate still (Kirchner, 2003).

The Setpoint Fallacy. One probable motive for the Optimising Fallacy is that conventional Regulators use a predetermined setpoint, with -ve feedback bringing a perturbed system back to that point. This immediately provides something to be optimised – 'distance from setpoint'. But GR is interestingly different from this, and has no fixed setpoint (c.f. 'rein control' below); as Perturbing Parameter *P* changes, so do the positions of equilibria in phase space.

The Beneficial/Harmful Confusion. At a stable steady state, as at *K* in Fig. 5, the -fc is *beneficial* in this sense: a change in the external Perturbation *P* will be compensated for by the feedback circuit tending to reduce the effect of that change. This is no more or less than the Le Chatelier Principle (Le Châtelier and Boudouard, 1898), well known and accepted in chemistry and economics. Nevertheless, around steady state *K* the effect that B has on E (+ve in this example), is actually *harmful* to B, in that an increase in E leads to a consequent decrease in B. The +ve effect of B on E (in this example 'black daisies'; the argument takes the same course with a -ve effect with 'white daisies') can be considered to be simultaneously (a) apparently 'G-beneficial' in promoting GR

by forming part of the -fc that allows this stable state with nonzero B to exist and (b) L-harmful, i.e. locally harmful to small variations in the B-effect on E. This is not the contradiction it first seems to be, since the G-beneficial describes the feedback rather than the effect (see Misattributed Feedback Fallacy). Nevertheless this has led to much confusion. E.g. the subtleties of beneficial/harmful in a Gaian context are not noticed in these quotations from a critic of Gaia theory:

> Coupling between the biosphere and the physical environment can potentially give rise to either negative (stabilizing) feedback, or positive (destabilizing) feedback, and the consequences of this feedback can potentially be either beneficial or detrimental for any given group of organisms. (Kirchner, 2002)

> Which brings us to Daisyworld. The Daisyworld model assumes that traits that benefit the environment also give an individual a reproductive advantage over its neighbors. (Kirchner, 2002)

The Evolution and Gaia Fallacies. This same confusion is also displayed by advocates of Gaia theory, as for example:

> In Daisyworld, natural selection is directly linked to environmental effects such that what is selected for at the individual level is beneficial to the global environment. (Lenton and Lovelock, 2001)

Natural selection refers to Darwinian evolution, and this statement is, as a general principle, entirely wrong. Darwinists know how easy it is to show counter-examples; this feeds their distrust of Gaia Theory. Later in the same paper a possible cause for this confusion is suggested where "what is selected for at the individual level is also beneficial at the global level" is offered as a rephrasing of "organisms alter their immediate (local) environment and the global environment in the same way". There is a sense of 'selection' in which black daisies within DW may be said to be 'selected' under GR when the sun is 'too cold', and white daisies 'selected' when it is 'too hot'; different organisms flourish in different environments. But emphatically this is not *natural selection* in any Darwinian sense, of variant organisms being preferentially selected in the same environment. As a general principle we may observe the opposite: what is G-beneficial to the global environment will be naturally selected *against* if Darwinian evolution occurs within the timescales of our model.

Above we noted that the effect of B on E was locally L-harmful. Any positive variation in this effect – for instance a mutation that caused a black daisy to have a slightly stronger tendency to absorb heat from the sun, thereby locally increasing temperature E – would actually (around a stable equilibrium such as K) *decrease* the amount of B. Such variations are broadly equivalent to changing the slope s in eqn 4. Hence, if the consequences were felt by the mutant B more than by its neighbours, that mutation would be selected against. Such selection pressure on albedo-changing mutations in black daisies tends (in the absence of constraints) to drive them towards neutral; likewise in white daisies. So evolution on such biotic effects ($B \rightarrow E$) can, under plausible circumstances, be expected to decrease and ultimately eliminate (the need for) GR.

Many critics of Gaia Theory assume that GR must have arisen through evolution, since it appears to be in some sense adaptive and hence seems to need some explanation for its origins. The previous Kirchner (2002) quote, discussing 'reproductive advantage', buys into this idea, and of course Dawkins' many criticisms of Gaia (Dawkins, 1982) address the same issue. In fact evolution is neither required for the display of GR – the present simple model makes no reference to evolution – nor, since as we argue that it does not depend on luck, does it require an evolutionary explanation for its origins. We return to further discussion of evolution below.

The Stability-Unlikely Fallacy. It is fallaciously believed by many that as a dynamical system becomes more complex, it is increasingly unlikely to have any stable steady states at all (May, 1972). The errors in May's analysis, as applied to nonlinear systems, have been pointed out elsewhere (Harvey, 2011); one error relates directly to the Missing-the-Point Fallacy. We demonstrate the contrary in the next section, in particular the inevitably of stable equilibria in our systems.

Two Reins and More

The minimal example used so far to display core GR has had just one type of biota B. This may puzzle people who think that DW relies on there being (at least) two daisies, black and white; though from the start W&L observed such regulation with a single type of daisy (e.g. see their Figure 1b).

Black daisies can G-regulate temperature when otherwise it would be too cold; it needs white daisies, or their equivalent, to G-regulate when it is too hot. Though a single environmental variable is being regulated, it needs two separate pathways for regulation in both directions. This matches exactly with Clynes' (1969) observations of 'rein control' (or 'unidirectional rate sensitivity') in biological homeostasis. Each rein of a horse can only pull in one direction, not push; hence for control in both directions two separate reins are needed. When doing so, the circumstances under which they may tend to cancel each other out, rather than complement each other, are discussed in Harvey (2004).

Saunders et al. (1998) were the first to relate rein control to DW. However they were largely focussed on circumstances where zero steady-state error may be achieved – that necessarily implies error with respect to some setpoint. To that end they added a version of integral control (that employs a signal related to time integral of error) to produce what they call 'Integral Rein Control'. To clarify, the original Clynes (1969) version of rein control, as used here and in Harvey (2004), is just plain rein control with no 'integral' element. As such, it has no need for the concept of a setpoint or of error.

The concept of rein control explains why two bio-variables are needed to control an enviro-variable in two directions. Further, it explains why, with many bio-variables but a single enviro-variable (McDonald-Gibson et al., 2008), at the core there is a dynamic equilibrium between those bio-variables that are 'pulling one way' and those 'pulling the other way'.

Extending to several enviro-variables is interesting and challenging (Dyke and Weaver, 2013); a start has been made there on analysing phase-portraits of two- and three-enviro-

variable systems. But here we now prove a theorem valid for arbitrary numbers of variables in a very general model.

Gaian Regulation Theorem

The minimal model presented above is a simple example of the general case, and so may be useful in guiding interpretation. We consider m bio-variables (*bvars*), $\mathbf{B}=(b_1,..,b_m)$; n enviro-variables (*evars*), $\mathbf{E}=(e_1,..,e_n)$; and a set of n parameters, or external perturbations, associated with the n *evars*, $\mathbf{P}=(p_1,..,p_n)$. **B** and **E** form a coupled dynamical system, parameterised by **P**. As before, we shall be examining the stable steady states of this system for different **P**, and in particular which regions of **P**-space allow one or more *bvars* from **B** to be viable. All variables are finite and bounded. This means we can rescale each variable in **B** and **E** to lie within the range [0.0,1.0] at all times.

For many purposes *bvars* and *evars* will be treated identically from a mathematical perspective. There are just two differences. Firstly only **E** is directly influenced by exogenous influence from **P**; indirect influence on the *bvars* is only via the *evars*. Secondly, for the term 'viable' to make any sense when describing *bvars*, there must be a contrast with 'non-viable' (i.e. stable steady state zero); for each *bvar* there must be both viable (>0) and non-viable (=0) regions.

It will be useful to introduce the notation $\mathbf{B_{-i}}$ to refer to all the *bvars* excluding the ith; similarly $\mathbf{E_{-j}}$ excludes the jth *evar*. Our dynamical system is then defined by $m+n$ equations of this form, using any continuous functions $h_i()$ and $f_j()$ at all that obey the bounding constraints below:

$$\mu_i \, db_i/dt = h_i(\mathbf{B}_{-i},\mathbf{E}) - b_i \qquad \text{for } i=1 \text{ to } m \quad (5)$$

$$\nu_j \, de_j/dt = f_j(\mathbf{B},\mathbf{E}_{-j},\mathbf{g}_j(\mathbf{P})) - e_j \qquad \text{for } j=1 \text{ to } n \quad (6)$$

The μ and ν moderate the rates of change, and we shall typically set these to 1. The $g_j(\mathbf{P})$ specify differently weighted subsets of **P**; any weighting is permissible. In turn these generate $m+n$ nullclines of the form:

$$b_i = h_i(\mathbf{B}_{-i},\mathbf{E}) \qquad \text{for } i=1 \text{ to } m \quad (7)$$

$$e_j = f_j(\mathbf{B},\mathbf{E}_{-j},\mathbf{g}_j(\mathbf{P})) \qquad \text{for } j=1 \text{ to } n \quad (8)$$

Bounding Constraints. The important constraints that we put on each $h_i()$ and $f_j()$ are that they must be continuous and lying within the range [0.0,1.0]. On translation to the real world, this reflects the fact that we only consider variables that are bounded below and above (e.g. for a species, it is bounded below at zero and above somewhere before the biomass exceeds the mass of the planet); and we may rescale all variables to lie within [0.0,1.0]. The consequence is that, although such functions may be defined for arguments outside that interval, they can only return values within it. Hence all the nullclines so defined intersect each other within the unit hypercube (defined so as to include its boundary). It follows that, from any starting position in phase space within the unit hypercube, there can be no trajectory leading out of it. When counting steady states, we need only search within this; no variables will 'shoot off to infinity'.

Counting Steady States

For this preliminary stage, we just consider the $m+n$ nullclines, with no need to distinguish between *bvars* and *evars*; we emphasise the generality of this result this by relabelling both here as $\mathbf{X}=(x_1, x_2, \ldots x_{m+n})$. We shall prove by induction Hypothesis 1 that: regardless of the number of variables, there are $2q$-1 steady states, for some $q \geq 1$, of which q are stable and q-1 unstable. We start by proving the result holds for 2 variables, and then show that the result still holds as we add one extra variable; by iterating we can reach any number of variables.

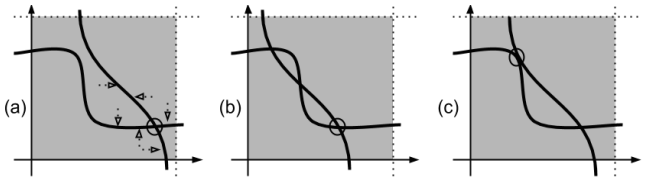

Figure 6. (a) The N-S nullcline must cross the W-E nullcline at least once; (b) if more, an odd number, provided (c) tangents are treated as a coincident pair of intersections.

Fig. 6 shows the 2 variable case (x_1, x_2); we consider each nullcline as starting and ending outside the unit box. They cannot intersect outside the box (because they are enclosed by the box boundaries), and the only intersections we need to count are within the box (including its boundaries). The nullcline heading north from the south edge of the box starts *below* the west-east nullcline and finishes, outside the north edge, *above* the west-east nullcline. Since each crossing toggles between *above* and *below,* there must be $2q$-1 steady states, where the nullclines intersect, for some $q \geq 1$. In Fig. 6a, the arrows around the circled intersection indicate the directions of dx_1/dt and dx_2/dt associated with the nullclines, demonstrating this steady state is stable.

In Fig. 6b, the circled intersection is clearly locally similar to that in 6a, hence will also be stable. A similar argument shows that the intersection in 6b furthest from that circled must also be stable; the central intersection can only be an unstable steady state. Since intersections must always alternate between stable and unstable along any nullcline, in the general 2-variable case we must indeed have $2q$-1 steady states, for some $q \geq 1$, of which q are stable and q-1 unstable. This is our hypothesis satisfied for just 2 variables. We note that since the nullclines are continuous functions of their variables and parameters, any smooth shift in these will result in a smooth movement through phase space of these steady states; except that where (Fig. 6c) tangents are created or lost, pairs of steady states (stable + unstable) appear or disappear; tangent intersections are counted double (Fig. 6c). In bifurcation theory this counts as a pitchfork bifurcation.

How do we extend Hypothesis 1 to 3 variables, where we are looking at 2-D nullclines intersecting in a 3-D unit cube? We note that if we consider a planar slice across the 'bottom' of the cube, where the new third variable x_3=0, the desired property holds true in x_1-x_2 space: there will be $2q$-1 points on that plane representing steady states. As we move the planar slice 'upwards', i.e. smoothly increasing x_3, these points indicating steady states will likewise shift smoothly, with possibly pairs of new points being created (and then

separating) or coming together (and then vanishing). The result will be the intersection of the x_1-nulline and x_2-nullcline in the form of a set of lines (minimum 1) that span the gap between the x_3=0 'bottom' plane and the x_3=1 'top' plane. The steady states of the 3-variable system will be those points where this set of lines cross the x_3-nullcline, that lies somewhere between the top and bottom of the cube.

If such lines cross the x_3-nullcline just once, we necessarily have the required $2q$-1 intersections; if, via any folds in such lines and/or in the x_3-nullcline, any such line crosses more than once, it must add an even number of crossings. Hence the required $2q$-1 intersections still applies. We note that, through reasoning similar to that applied in the 2-D case, any such intersection that is adjacent to the corner of the unit cube is necessarily stable. Likewise, successive steady states alternate between stable and unstable. Hence we can extrapolate from the 2-D case to the 3-D case, where Hypothesis 1 still holds: there are $2q_3$-1 steady states, for some $q_3 \geq 1$, of which q_3 are stable and q_3-1 unstable.

Though it is more difficult to visualise in higher dimensions, we can make exactly the same arguments in progressing from 3 variables to 4 variables, and thus iterate further to any arbitrary number of variables. Since in considering n-dim space we already have the result for (n-1)-dim space, we are each time considering a number of threads ($2q_{(n-1)}$-1) starting outside the 'bottom' (n-1)-hyperplane of the n-hypercube, and progressing continuously until exiting at the 'top'. These threads correspond to intersections of (n-1) nullclines (that are hyperplanes) and must at some stage pass through, on their way from bottom to top, the intervening nth nullcline-hyperplane. The only way any continuous thread may terminate as one travels up, other than exiting at the top, is for the equivalent of tangent collisions to be made or broken. As we have already seen, these add (or subtract) pairs of intersections that separate (or come together). Thus any one thread is ultimately continuous from bottom to top and crosses the intervening nullcline an odd number of times, minimum once. The pattern of the Hypothesis is carried over to the next dimension, now with $2q_n$-1 steady states, q_n of these stable.

This preliminary stage means we can guarantee the existence of stable steady states, typically along with unstable ones, regardless of the complexity of our system; the bounding box is in its entirety a basin of attraction for stable steady states. We need this result to continue our proof.

Comparing Viability Zones. We extend the VZ definitions from this minimal context to the more general domain.

We select any arbitrary *bvar* b_i to be considered as the focus of interest, and define the VZs associated with it. Each of these zones will be a zone in **P** space or a subset of **P** space.

For each possible value of **P** we can find in principle all the steady states of the full set of equations; we have proved above that such steady states exist within the unit hypercube. If (for a specific value of **P**) there is any stable steady state with $x_i > 0$ then by definition this specific value of **P** is within VZ_{full}. We next generate $VZ_{neutral}$ by doing a similar exercise but with x_i 'neutralised'; within all the other n-1 equations, x_i is replaced by zero, thus having no effects, +ve or -ve, on any of the other variables. As with the minimal 2-variable case, we have $VZ_{neutral}$ and VZ_{full} each defined as zones within Parameter space **P**. Can we prove anything about their relationship? Surprisingly, yes – and easily.

Reductio ad Absurdum Proof of Theorem

The Hypothesis 2 that we now wish to prove is $VZ_{neutral} \subseteq VZ_{full}$. Suppose this was false. Then there is at least one point in **P** space that lies within $VZ_{neutral}$ but outside VZ_{full}. We consider such a parameter set.

Since it is outside VZ_{full}, it follows that there all stable steady states of the system require x_i to be nonviable, i.e. x_i=0. But this is exactly the condition we impose when we neutralise x_i so as to define $VZ_{neutral}$; hence we have shown that this point in **P** space must lie outside $VZ_{neutral}$. Assuming the Hypothesis to be false has produced a contradiction. QED.

A less rigorous, but perhaps more comprehensible, explanation would be: if this point in parameter space produces a non-viable x_i in the full (un-neutralised) system, then it will definitely behave the same way in the neutralised version (where x_i is constrained to be zero). We have proved $VZ_{neutral} \subseteq VZ_{full}$.

This means that the relationship between these VZs may be equality, if x_i has no effects on its own VZ; this will certainly be the case if $VZ_{neutral}$ covers all of parameter space. But *if* it has any effect at all on its VZ, this can only be to increase its size. This means we can define a 'Gaia Regulated' VZ_{GR}, such that $VZ_{full} = VZ_{neutral} + VZ_{GR}$. We note again, as above, that ideal conditions for VZ_{GR} to be significant in size include $VZ_{neutral}$ being small and constrained; e.g. if x_i refers to some complex organism with tight constraints on its viability.

Corollaries. The original minimal DW example had the various VZs as continuous intervals in the space of a single parameter. This generalised version has no such constraints; not only can any number of variables be included in the viability conditions, but also disjoint VZs are equally valid.

Because the result $VZ_{neutral} \subseteq VZ_{full}$., for any variable x_i is derived from such a general system, we need not assume that x_i necessarily refers to a single biotic (or indeed environmental) variable. It could for instance apply to a variable defined as 'black AND white daisies', likewise 'black OR white daisies', using logical AND and OR. Indeed the Gaian Regulation Theorem will apply to any entity, including a complete ecosystem, that (a) conforms to any version of equations (5) and (6), (b) can be assessed for viability as some exogenous parameter set influencing the environment varies, (c) has some effect on that same environment, thus (d) forming feedback circuits with stable steady states.

Why does the Theorem Work?

The theorem is very general in application. It is instructive to see just which minimal but necessary constraints provide the interesting results.

Firstly, the form of all the equations used, both for bio-variables (eqn 5) and enviro-variables (eqn 6), is such as to be naturally interpretable in terms of continuous nullclines (eqns 7 and 8). This follows the precedent set by most previous analyses. We consider the steady states of such systems.

Secondly, there is an implicit division of timescales into 3 ranges. Explicitly, the only ones expressed in the equations are μ and ν which provide the timescale within which the system finds its way to stable steady states. Implicitly, there is the faster timescale of small 'thermal' noise that we assume will dislodge the system from any unstable steady state; and the

slower timescale of any changes in the exogenous parameters **P**. We assume such parameters stay fixed long enough for stable steady states to be reached. We return to this below.

Thirdly, finite bounds are put on the nullclines, and we choose to rescale the variables such that nullclines lie within [0.0,1.0]. This immediately means that they can only intersect within the unit hypercube, and guarantees us stable equilibria (that the system has time to reach), as well as any unstable ones. As shown above, we can relate numbers of stable states to unstable ones, and observe that they alternate along nullclines. This is the trick that was missed in previous studies of similar complex systems that considered only linear ones (Gardner and Ashby, 1970), concluding that instability became inevitable as numbers of variables increased; and in studies that claimed inaccurately to extend such linear results to the nonlinear case (May, 1972). Errors in the latter studies have been previously pointed out (Harvey, 2011); it is a classic example of the Missing-the-Point Fallacy.

Fourthly, the concept of viability, states of affairs where a bio-variable can be nonzero, is central. This comes directly from the cybernetic era that provided a context for the origins of DW. Ashby (1952) introduced the similar idea of 'essential variables' to his analysis of homeostasis and homeostats; as an aside, his Chapter 20 on 'Stability' provides the explicit motivation for Gardner and Ashby (1970) – but misses out what nonlinearity brings to the stability table. The key aspect of viability here is its potential fragility, where it may be found and lost. As we saw earlier (Fig. 4), VZ_{GR} relates to zones of hysteresis where B may have either +ve or zero values. The hysteresis can be interpreted as a symptom of fragility, it may be a 'struggle' to recover from a loss of viability. The boundaries to viability are crucial to GR.

Fifthly, GR is defined in terms of zones of Parameter-space and not zones of enviro-variable space. The original DW model was clear about this; for instance W&L, in their Figure 1, illustrate viability as their parameter of Solar Insolation varies.

Sixthly, also explicit in W&L (e.g. their Figure 1a), is the comparison made with a 'neutral' version (in their case a neutral colour of daisy) that generates $VZ_{neutral}$. Basically, VZ_{full} identifies the P-zone where B is viable ***including*** any hysteresis loop (i.e. counting the upper, viable part of such a loop), whereas $VZ_{neutral}$ identifies the same but excluding any hysteresis loop (by in effect counting the bottom, zero-valued part). VZ_{GR} is defined as the difference, the 'fragile zone' of the hysteresis loop. This allows the remarkably simple yet powerful Reductio ad Absurdum proof used above. Indeed with this insight it becomes clearer why there is nothing that could count as the inverse of VZ_{GR}, there is no possibility of 'anti-Gaian-regulation' as assumed by lucky-Gaia proponents.

Tipping Points. The boundaries of VZ_{GR} are associated with discontinuous jumps in the hysteresis loop, as seen at P1 and P2 in Fig. 4.; one will be experienced as P decreases, the other as P increases. Whereas the P2 boundary is shared between VZ_{GR} and $VZ_{neutral}$, the other P1 boundary derives from where the enviro-variable nullcline (here a zigmoid) is tangent to the bio-nullcline (here a hat function). Given the use of a witches hat, it so happens that this coincides with the peak value of bio-variable *B*. More generally, any form of hat may be used, and as seen in Fig. 7, the tangent associated with the tipping point need not be anywhere near the peak.

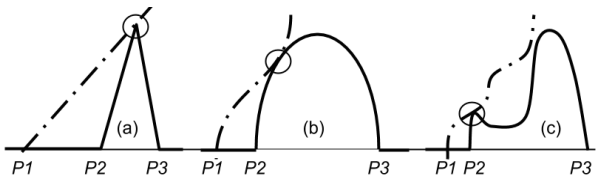

Figure 7. (a) P1 is associated with the tangent (circled) at the peak of a witches hat. More generally, (b) and (c), such tangents need not be at the peak of the hat.

Gaia and Evolution

The basic DW model contains no element of Darwinian evolution, yet displays GR; so GR does not require evolution for *its ongoing operation*. Further, given the fallacious reasoning behind 'lucky Gaia', GR is not some improbable phenomenon that needs adaptive explanations like evolution to explain *its origins*. It arises naturally and inevitably whenever there is some 'fragile' system with the appropriate kind of hysteresis loop. In so far as natural evolution generates living systems that are indeed fragile in that sense, viable only within some ecological niche, it generates the raw material for GR – which then arises by definition, rather than through some further mystery that needs explaining.

Evolving the effects of *B* on *E*. We have seen above (in discussion of the Beneficial/Harmful Confusion, the Evolution and Gaia Fallacies) that there is an opposition between evolution and GR in this sense: when a biotic trait has an environmental effect that contributes to a GR feedback circuit, it is L-harmful and selection would tend (in the absence of constraints) to reduce or eliminate it. A simple interpretation of this is: it is in an organism's interests to evolve so as to *decrease* its fragility within its ecological niche, and in so doing so it inevitably *reduces* the scope (and need) for GR.

Evolving the effects of *E* on *B*. We may also consider the possibility of evolution evolving biota so as to alter their susceptibility to environmental conditions. It can readily be shown (Robertson and Robinson, 1998) that if daisies in DW have unconstrained capacity to evolve so that their optimal growth temperatures (the 'peak' of the hat-function) match the prevailing temperature, then likewise this will *reduce* the scope for GR and ultimately eliminate it. A response (Lenton and Lovelock, 2000) broadly agrees, whilst asserting that in real systems where physical constraints set bounds to the limits of such adaptation, GR can take over when Darwinian evolution runs up against such constraints. Again we see GR and Darwinian evolution as complementary not antagonistic.

Incorporating Evolution inside this model. It is of course possible to incorporate the dynamics of evolution directly within this model, provided the 'terms and conditions' (t&cs) are respected. This typically requires the evolutionary dynamics to run to steady state before any parameter *P* is changed. Traits that are evolvable will typically (a) not be directly affected by parameters and (b) be viable or non-viable at steady state; hence they fit the requirements to be labelled as *bvars*. If eqns 5 and 6 can be tailored to reflect the evolutionary dynamics, then the GR Theorem will apply to such traits just as with any other *bvar*.

A Vicious Circle in Gaia/Evolution debates. DW was invented without reference to evolution (there was no need). Darwinians misunderstood, assumed GR could only be evolved yet it was susceptible to 'cheats'; hence they declared it impossible. DW modellers knew this was not so, and created DW extensions to demonstrate this. Unfortunately they took 'anecdotal' specific examples and elevated them into general principles without justification. Darwinians easily found 'anecdotal' counter-examples to these general principles, and asserted the opposite. The cycle of misunderstanding and suspicion continued.

For example, some studies combining evolution and DW (e.g. Lenton, 1998) mix the timescales by having parameters changing simultaneously with the evolutionary dynamics. The results are valid for those specific choices, but are in this context 'anecdotal' since we have no principles to decide how far we can generalise them. The GR Theorem presented here will not extend to such results, but is fully generalisable within its own carefully stated constraints, its t&cs.

Discussion

This paper is deliberately fact-free, it is mathematics rather than science. The analysis has focused on abstract models that come with t&cs. Any real world lessons depend whether real phenomena do indeed match the t&cs, and this is outside the scope of this paper. Nevertheless, it is hoped that some of the ideas and intuitions here may be useful tools for scientists. The broad generality of the results make it tempting to try and fit this model-template to the world; however we may warn that one of the most challenging t&cs to observe may be the strict separation of timescales. The results here depend on the parameters P being maintained fixed long enough for the $B \rightleftarrows E$ dynamics to reach steady state.

Going out on a limb. An appropriate metaphor for the fragile nature of GR is that of going out on a limb – on the top limb of a hysteresis loop, from the safe tree trunk of $VZ_{neutral}$. The mathematical results here should be safe, but we may speculate what future directions may be promising. One such is going beyond steady state phenomena to metastable states.

Slower timescale changes may alter, even eliminate the safe zone of $VZ_{neutral}$; to rephrase Wittgenstein and his ladder, throwing the tree trunk away after one has climbed up it. Where you are on a hysteresis loop depends on historical contingencies of how you got there, hence changes at multiple timescales may eliminate possibilities of going back to safety. Darwinian evolution of course introduces new timescales.

This may present a picture of Life – and on different scales e.g. tornadoes, and planetary viability for biota – as potentially fragile existence on a limb of multidimensional hysteresis loops with no going back. There are strong echoes here of an autopoietic definition of Life (Varela et al., 1974); this is unsurprising, since autopoiesis and DW have shared origins in many ideas from cybernetics. Such fragility is of course balanced against the powerful forces of Gaian regulation providing the supporting limbs; "Gaia is a tough bitch", as Lynn Margulis commented.

Conclusions. This study shows Gaian regulation merely in a model, not the real world. Nevertheless we can debunk many widely held fallacies; critics of Gaia Theory are not justified in their appeals to these specific misunderstandings. GR is not 'lucky', it is inevitable (subject to the t&cs). The relationship between GR and Darwinian evolution is more subtle and complex than it is often misrepresented to be; their different roles may be seen as more complementary than antagonistic. Bounded physical variables imply the existence of stable steady states, and hence inevitable GR. However this GR is not 'optimising', not even really a 'comfortable' Gaia; perhaps best called 'habitable' Gaia, it fits nearest to what Kirchner (2003) calls 'biological feedback at the limits of habitability'.

The main mathematical lesson is that the curious circular logic of Gaia is full of surprises and challenges our intuitions.

References

Ashby, W. R. (1952). Design for a brain. Chapman and Hall.

Clynes, M. (1969). Cybernetic implications of rein control in perceptual and conceptual organization. *Ann. NY Acad. Sci.* 156:629-670

Dawkins, R. (1982). The Extended Phenotype. W. H. Freeman, Oxford.

Dyke, J. G. and Weaver, I. S. (2013). The emergence of environmental homeostasis in complex ecosystems. *PLOS Computational Biology* 9(5):1-9.

Gardner, M. R. and Ashby, W. R. (1970). Connectance of large dynamical (cybernetic) systems: critical values for stability. *Nature,* 228:784.

Harvey, I. (2004). Homeostasis and rein control: from Daisyworld to active perception. In Pollack, J. et al., editors, *Proc. 9th Intl. Conf. on Simulation and Synthesis of Living Systems, ALIFE9,* pages 309-314. MIT Press.

Harvey, I. (2011). Opening stable doors: complexity and stability in nonlinear systems. In Lenaerts, T. et al., editors, *Advances in Artificial Life, ECAL 2011,* pages 805-812. MIT Press.

Kirchner, J. W. (2002). The Gaia hypothesis: fact, theory and wishful thinking. *Climatic Change* 52:391-408.

Kirchner, J. W. (2003). The Gaia hypothesis: conjectures and refutations. *Climatic Change* 58:21-45.

Le Châtelier, H. and Boudouard O. (1898). Limits of Flammability of Gaseous Mixtures. *Bulletin de la Société Chimique de France (Paris)*, 19:483-488.

Lenton, T. M. (1998). Gaia and natural selection. *Nature* 394:439-447.

Lenton, T. M. and Lovelock, J. E. (2000). Daisyworld is Darwinian: constraints on adaptation are important for planetary self-regulation. *J. Theor. Biol. 206*:109-114.

Lenton, T. M. and Lovelock, J. E. (2001). Daisyworld revisited: quantifying biological effects on planetary self-regulation. *Tellus* 53B:288-305.

Lovelock, J. E. (1979). Gaia: a new look at life on Earth. Oxford University Press.

Lovelock, J. E. (1983). Gaia as seen through the atmosphere. In Westbroek, P. and de Jong, E. W., eds, *Biomineralization and biological metal accumulation,* pages 15-25. D. Reidel Publishing Company, Dordrecht.

May, R. M. (1972). Will a large complex system be stable? *Nature,* 238:413-414.

McDonald-Gibson, J., Dyke, J. G., Di Paolo, E. and Harvey, I. (2008). Environmental regulation can arise under minimal assumptions. *J. Theor. Biol.* 251(4):653-666.

Robertson, D. and Robinson, J. (1998). Darwinian Daisyworld. *J. Theor. Biol.* 195:129-134.

Saunders, P. T., Koeslag, J. H., and Wessels, J. A. (1998). Integral rein control in physiology. *J. Theor. Biol.* 194:163-173.

Varela, F. J., Maturana, H, R., and Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems* 5:187–196.

Watson, A. J. and Lovelock, J. E. (1983). Biological homeostasis of the global environment: the parable of Daisyworld. *Tellus* 35B:284-289.

Wood, A. J., Ackland, G., Dyke, J., Williams, H. and Lenton, T. (2008). Daisyworld: a review. *Reviews of Geophysics* 46(1):RG1001.