

Opening Stable Doors: Complexity and Stability in Nonlinear Systems

Inman Harvey

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

Abstract

Generic complex systems of many interacting parts can model both natural and artificial systems, and the conditions for their stability are of interest. Two influential papers (Gardner and Ashby, 1970; May 1972) laid down a mathematical framework suggesting that, without some specific constraints on the interactions, such systems are very likely to be unstable as they increase in size and connectance. We draw attention to a programming error and to flaws and omissions in reasoning that discredit such conclusions when applied to nonlinear systems. With nonlinearity the role of ‘ambiguous’ functions, where the sign of an interaction varies between plus and minus according to context, is highlighted as increasing the number and variety of equilibria. Unimodal ‘hat-functions’ are a very basic class of ambiguous function. In Daisyworld models they serve to model a viability range for a species in a biota-environment system that results in an extended range of stable equilibria. It is concluded that the flaws in the traditional framework make it unsuitable for understanding generic complex nonlinear systems, and that opening up to the use of these ambiguous functions increases the prospects for finding stability.

Introduction

An influential early paper (Gardner and Ashby, 1970) used computer simulations to assess the probability that a large system of interacting component parts that has been assembled at random, or has grown haphazardly, will be stable or unstable. They considered systems where the interactions between parts were linear, and looked at how the expectation of stability changed as the number of variables increased. This was a theoretical study, to be motivated by its possible application to both biological and man-made systems: brains (real or artificial), planetary climate systems, social or financial systems, ecosystems. The conclusion was the suggestion that all such large (random or haphazard) complex linear dynamic systems may be expected to show the property of being stable up to some critical, fairly small, level of connections; but above that phase transition they are overwhelmingly likely to be unstable. From this it could be deduced that if one observed large complex linear systems that were indeed stable, there must be something exceptional and non-random about the way that the parts were connected.

The influence of this work stems primarily from its extension and development by Robert May, and the subsequent proliferation of a wide body of research in this area. He replicated the results analytically rather than

computationally (May, 1972), and claimed that their validity extended beyond the linear systems of Gardner and Ashby (hereafter: G&A) to systems “which in general may obey some quite nonlinear set of first-order differential equations”. May’s interest mainly focused on ecological systems, and a subsequent book (May, 1973) largely set the agenda for discussion of the relationship between complexity and stability in ecosystems ever since.

Before this work, there was a common perception that the more diverse was the range of species in an ecosystem, the more robust and resilient to perturbations that system would be; and further, it was often assumed that this may well be due to some underlying law of large numbers that could apply very generally across all sorts of systems with many interacting components. But this work, apparently using very minimal mathematical assumptions, appeared to suggest that the opposite was true – at least, in the absence of further specific constraints. So subsequent argument and analysis have tended to focus on what further constraints, what limitations on the number, sign and size of interspecies interactions, might be necessary in order to make it likely that a complex ecosystem was stable. The mathematics, it has largely been assumed, is relatively simple and correct. Hence if we want to explain the existence of complex stable systems, it looks like we need to add further assumptions.

In this paper, I shall demonstrate that the reasoning within these two primary sources (Gardner and Ashby, 1970, May, 1972) is partially invalidated through omission and errors, and in particular should not be generalised in this way to nonlinear systems. Firstly, I draw attention to a programming error in the G&A paper, which has been noted previously (Solow, Costello and Beet, 1999). Secondly, both papers omit from consideration a large and important class of systems, namely those where the significant component variables have zero direct feedback upon themselves. Thirdly, I point out a flaw in May’s reasoning when he claims to go beyond the purely linear systems of G&A towards a more general set of nonlinear systems. May considers local stability at just a single point in the space of possible values for the system, a point that makes sense when considering linear systems with negative self-interactions. Unfortunately, even in linear systems, that point may not uniquely exist when some self-interactions are zero; and when we move on to nonlinear systems there can be a large (and in some circumstances unlimited) number of points of potential stability to consider. Global instability would require local instability at every one of those points. Hence the probability of global stability will

be underestimated if one just considers local stability at a single point.

These various criticisms are, as far as I am aware, all drawn together here for the first time. The flaws in these two foundational papers suggest that a radical reappraisal is needed in the mathematical foundations of a substantial body of work that has built up over some 40 years. Rather than seeking a route to stability by adding further constraints to these abstract models, we need to open the doors to those possible locations of stability that have until now, through error or omission, been excluded. The significance goes beyond ecosystem theory to the study of all kinds of natural and artificial systems with complex nonlinear interactions, including financial systems (Haldane and May, 2011).

Gardner and Ashby on Linear Systems

Their short paper, a Letter of less than one page in *Nature* (Gardner and Ashby, 1970), was an early example of a computer simulation, using a Monte Carlo approach. They considered a very simplified formal model of any large system of many interacting parts. This could be traffic at an airport, or the neurons in a human brain. They asked the question: supposing one did not know all the details of the interactions between component parts, but modelled these as coming from some random distribution that gave the signs and sizes of these interactions, then what was the chance that such a large system will be stable? Although in the real world most of these large systems, perhaps biological or social, will be grossly nonlinear, they explicitly restricted themselves to considering only systems with linear interactions, as a ‘first step towards a more general treatment’.

The model had n component parts. The intention was to investigate how the generic properties of such systems varied as n increases. The instantaneous state of the system can be expressed by a vector \mathbf{x} , where x_i represents the current value of the i th variable. The rate of change of x_i is a weighted sum of the current values of all the variables:

$$\frac{dx_i}{dt} = \sum_j a_{ji} x_j$$

Because this is a linear system, there is a unique equilibrium point where for all i , $dx_i/dt = 0$. The issue will be: what is the probability that this unique equilibrium is stable, given the distribution from which the weights a_{ji} in the connection matrix A are drawn. A is the Jacobian matrix of the first-order partial derivatives, and in this case of a linear system these terms are all scalars, of fixed size and sign; when, later, we move on to nonlinear systems, these terms will be variable in both size and sign.

G&A chose to make this a partially connected system, with a proportion C of the (fixed) weights being nonzero. These nonzero weights were distributed evenly between -1.0 and +1.0. Further, they ensured that all the weights a_{ii} in the main diagonal of the connection matrix were negative. They distributed these evenly between -1.0 and -0.1 (in May’s version that followed, May set all these to -1.0). G&A had what might be considered a reasonable motive for making all these self-connection weights negative; after all, if any one of these were positive the system could not be stable because any perturbation on such a variable would run away through

positive feedback. But as we shall see shortly, G&A (and May who followed them) excluded an enormous space of possible stable systems – arguably the space within which most complex biological systems lie.

G&A are thus discussing a family of linear feedback systems, parameterised by these two values: n , the number of component parts, and C , the connectance or the proportion of possible interactions between parts that are non-zero. For any given values of n and C , their Monte Carlo approach involved testing many cases of such systems, with the connection weights drawn from the appropriate distributions, and finding out through computation what proportion of the systems were stable at their unique equilibrium point. For low values of the connections, where the interactions are dominated by the stipulated negative values of self-connections, the probability of stability approached 100% for all values of n tested. But as the connectance C increased, the probability of stability fell away. Using the limited computational facilities of their day (Gardner and Ashby, 1970), they were only able to test examples where n equals 4, 7 or 10. Their conclusion, illustrated by a figure, was that as n increases the relationship between connectance and stability changes from a smooth falling away of probability of stability as connectance increases (for any cause for) towards a step function for values of n of 10 or more. Their figure (partly replicated by the thin lines in Figure 1 here) suggests that for $n=10$ this phase transition (from “almost certainly stable” to “almost certainly unstable”) occurs at or around a connected value of 13%, $C=0.13$.

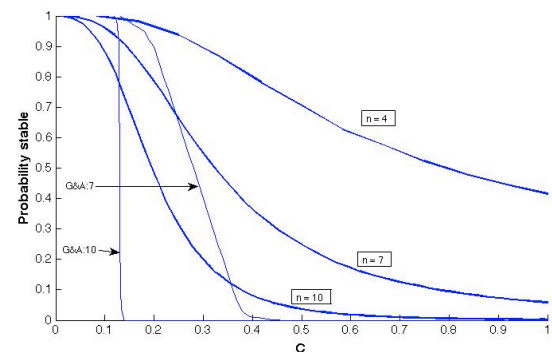


Figure 1: Thick lines give the correct results for G&A’s examples, for $n = 10, 7, 4$ from left to right. Diagonal terms a_{ii} drawn from $[-0.1, -1.0]$; a proportion C of off-diagonal terms a_{ji} ($i \neq j$) drawn from $[-1.0, 1.0]$, with the remainder zero. Thin lines copy the incorrect results that G&A showed for $n = 10, 7$ (Gardner and Ashby, 1970).

The Programming Error

When I replicated their method¹ my results were similar for $n=4$, but noticeably different for $n=7$ or 10. The difference, shown in Figure 1, is striking, and in particular eliminates the sharp nature of the phase transition claimed for $n=10$. This phase transition was their main result, and was commented on by May in his subsequent work (May, 1972), who on the basis

¹ Matlab code at www.informatics.sussex.ac.uk/users/inmanh/stable

of his analysis claimed: “The sharp transition from stability to instability, which was the essential feature of their paper is confirmed”. Having failed to replicate this transition, I was puzzled as to whether I had misinterpreted their methods. I consulted my colleagues and the considerable literature that has arisen in this area since this seminal paper. Eventually a colleague discovered a rarely cited 1999 reference (Solow, Costello and Beet, 1999) pointing out the same problem, with results agreeing with my own. They attributed the problem to some unknown programming error in G&A’s code. Further, they comment that this nullifies one of May’s conclusions where he had assumed that the G&A phase transition was a real phenomenon. Correction of this programming error does not alter the conclusion that as n increases and C increases the probability of stability goes down; it does alter the conclusion that for values of n above some fairly small value the relationship between stability and connectance turns into something close to a step function.

For the purposes of this paper, this programming error is the least important of the errors and omissions to be discussed. Nevertheless, it is of note that it took nearly 30 years until this error was pointed out in print.

Stability through indirect means

G&A have explicitly set up their model with n variables of interest, and considered where that system gave a stable equilibrium for all these variables. Realistically there may, in a real world system, be other variables that are interconnected, say m of these, such that the stability of these m variables is not required; yet the stability of the original n arises through interactions with the others. For ecosystem models, one may be interested in the stability of n species, but these are affected by, and in turn have an effect upon, some m environmental variables that are implicated in a larger system of $(m+n)$ variables.

Here is a concrete example that, in anticipation of a later section introducing Rein Control, is presented in the metaphor of the reins of a horse. Right and left reins R and L each affect the heading H of a horse, and in this model are in turn being adjusted in strength in accordance with H:

$$\begin{matrix} dH/dt = R - L; \\ dR/dt = -H; \\ dL/dt = H; \end{matrix} \begin{pmatrix} 0 & +1 & -1 \\ -1 & 0 & 0 \\ +1 & 0 & 0 \end{pmatrix}$$

We can see that the heading H will be kept at a stable equilibrium around H=0, whilst R and L can take any values as long as they tend to pull equally in opposed directions. The Jacobian matrix has zeros in the main diagonal, yet H is held stable.

Such systems are excluded by design by G&A, and by May who follows them. In the context of, for example, concerns about the stability of an ecosystem of n species, this assumes that the only variables relevant to the system are the species themselves. But one of the insights derived from Gaia hypothesis (Lovelock, 1972; Watson and Lovelock, 1983), that is now incorporated within the field of Earth Systems Science, is that the biota and environmental variable are coupled in multiple overlapping feedback loops. For many

purposes, an analysis that ignores the intermediate role of environmental variables, and focuses on models that solely include the species, may miss out on the underlying dynamics that explain ongoing stability.

We shall return to these issues in later sections. For now, we return to examining the restricted set of models that G&A, and May following in their footsteps, considered.

May’s generalization to nonlinear systems

Whereas G&A explicitly limited themselves to the consideration of linear systems “merely as a first step towards a more general treatment” (Gardner and Ashby 1970), May claims to be considering systems “... which in general may obey some quite nonlinear set of first-order differential equations.” (May 1972). His method is to focus on the behaviour of such nonlinear equations around “the equilibrium point”. Through making a Taylor expansion and ignoring the higher-order terms one can consider this locally as a linear system. Thereafter, May goes on to analyse the same kind of linear system as G&A, while still claiming that it generalises to nonlinear systems.

It is puzzling that this extension to nonlinear systems seems to have gone unchallenged in the literature. For the most part, people referring to May’s work do not mention at all the distinction between nonlinear and linear systems; indeed some authors only mention it as relating to linear systems. It is not clear whether this is a shorthand reference to the fact that May’s method for analysing nonlinear systems required him to first reduce them to linear ones, or whether this was through a lack of recognition of May’s explicit claim to be dealing with the general case of nonlinear systems.

Insofar as May’s analysis is restricted to the linear version, he tackled analytically the same class of systems (to be precise, a slight variant with qualitatively the same behaviour) that G&A had tackled computationally. His results were broadly similar. May claims that the central feature of the results for large systems is “the very sharp transition” from stable to unstable behaviour above a critical value, that “accords with Gardner and Ashby’s conjecture”. As we have pointed out above, in fact the transition is not as sharp as G&A indicated, since they were apparently misled by an artefact of a programming error. However, the analytical results do agree with a correctly coded computational Monte Carlo approach. The influential take-home message from both the computational and analytical results has been: in any such system of many interacting parts, as soon as the connectance parameter (the proportion of nonzero interactions between component parts) rises above some small value, the probability that such a system will be stable drops to near zero. This limitation on stability becomes worse as n , the number of parts, increases. In the context of ecosystems, such a result challenges the commonly held assumption that the more diverse an ecosystem is, the better it is able to remain stable in the face of perturbations.

The abstract systems that May analyses have the (off-diagonal) interaction terms drawn from a random distribution of positive and negative values. Counter-arguments of those who want to understand why, nevertheless, we do observe many examples of biological complex nonlinear systems that

are generally stable, have been to focus upon this distribution. A typical response would be: the mathematical analysis is unassailable, so it must be that in the natural world the interaction weights are biased in some way, and not drawn from such a random distribution. One version of this argument is to suggest that there are biological constraints shaping the interactions so that their values must come from some significantly different distribution. A second approach is to accept that the original random distribution of connection strengths was likely to lead to instability, and follow through the consequences. With some further assumptions as to how such instability could lead to the extinction of various species in succession, it can be shown that after about half of them have disappeared, the remaining ones are now likely to have settled into a stable equilibrium. The connection strengths of this reduced subsystem will have been biased appropriately through this weeding out process.

The response of this current paper is different. Whilst not challenging the main thrust of the case for linear systems, it is to point out errors in the mathematical reasoning that mean that May's results for linear systems do not generalise, as he claims, to nonlinear ones.

Inability to locate Unique Equilibrium

Whereas with linear systems the terms in the connections or Jacobian matrix are scalars and fixed, with nonlinear systems these terms may be any functions of the variables and varying. So, given that May's strategy is to take the linear approximation to the system at 'the equilibrium point', it is clearly of vital importance just how one identifies this unique location in phase space, since this will determine the values fed into the linear approximation.

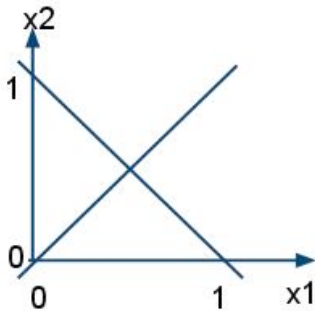


Figure 2: Equilibrium is where the two lines intersect at (0.5,0.5).

This may be a misinterpretation of the intended meaning, but one clue may be given by May's statement: "Following Gardbner and Ashby, suppose that each of the n species would by itself have a density dependent or otherwise stabilized form, so that if disturbed from equilibrium it would return with some characteristic damping time." Since 'by itself' is contrasted with (in a following sentence) "Next the interactions are 'switched on'...", this could provide a method that might identify a unique point: for each x_i identify a value (if such exists) for which it would, in isolation, be in stable equilibrium. What makes me think this could be a misinterpretation is the fact that this quickly leads to absurdities, even in simple cases such as the following.

Consider a linear system of two variables such that:

$$\frac{dx_1}{dt} = 1 - x_1 - x_2; \frac{dx_2}{dt} = x_1 - x_2; \begin{pmatrix} -1 & -1 \\ +1 & -1 \end{pmatrix};$$

The Jacobian matrix is on the right. By inspection we can see that it has an equilibrium ($dx_1/dt = 0 = dx_2/dt$) where the lines intersect in Figure 2, at (0.5,0.5), and this will be a stable equilibrium. But if we consider x_1 on its own in the absence of any interactions with x_2 , it will vary according to the reduced equation $dx_1/dt = 1 - x_1$; with a stable equilibrium at $x_1=1$; similarly, x_2 on its own has a stable equilibrium at $x_2=1$. However, back in the expanded two-variable system the suggested (x_1, x_2) point at (1,1) is not an equilibrium point at all.

So this cannot be a proper method for identifying 'the equilibrium', and indeed this may not have been the intended meaning. But no other method is given, so we are left with a quandary. For with nonlinear systems there are likely to be many equilibria, and the distribution of interaction weights, and hence consequences for stability, will vary from one to another.

Many Possible Equilibria

In the linear case, there will be just one equilibrium point. Global stability of the whole system is thus identical to local stability at this one point. In the more general case of nonlinear systems, however, there may be many equilibria, indeed potentially an unlimited number. Global stability can arise from local stability at just one of these. For example, consider the two-variable system given by these first order partial differential equations, with the associated Jacobian:

$$\frac{dx}{dt} = y - \cos(x); \frac{dy}{dt} = x - 20y; \begin{pmatrix} \sin(x) & 1 \\ 1 & -20 \end{pmatrix}$$

This has equilibria at values where the lines $y=\cos(x)$ and $x=20y$ intersect. As can be seen from Figure 3, this gives 13 equilibria, alternating between stable and unstable.

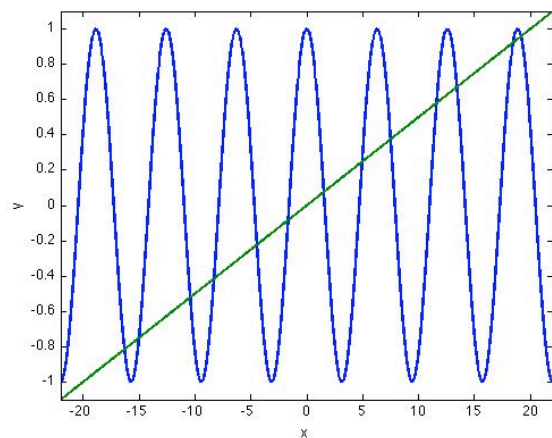


Figure 3: Intersections between the functions $y=\cos(x)$ and $x=20y$ indicate equilibria. With nonlinear functions there is the potential for many equilibria, both stable and unstable.

If there are several different equilibria, then the values of the connection strengths at one equilibrium point will in general be different from the values at another equilibrium. Whereas in the linear case it is relatively straightforward to propose some simple distribution from which all the connection strengths are drawn at random, in the nonlinear case it is an open question just what family or families of randomly parameterised nonlinear functions one might choose to throw into the melting pot. Until such choices made it is not possible to make any assertions about the distribution of varying connection strengths; and certainly not possible to characterise the (possibly biased) sample of these connections strengths at the various equilibria.

It might be argued, in the context of using these systems as models of ecosystems, that in practice the extension from linear systems to nonlinear systems could be highly constrained. For instance, if the nonlinear functions were constrained to be monotonic, then (as with the straight lines of linear systems) there may be only one equilibrium point to consider. In other words, as one moves around phase space the size of the connection strength might vary but its sign would stay the same; for example species P may have a greater or lesser effect upon species Q depending on other circumstances, but the sign of this effect could be expected to stay the same. But realistically, ecosystem models should probably not focus on just the numbers of different species as the only variables remodelled. Typically, many interactions will be mediated through environmental resources. Changes in such resources will have effects on the different species, and likewise changes in species number or concentrations will have effects upon the resources. From a mathematical perspective both species and environmental resources should just be considered as different variables within the model. It then becomes possible to visualise cases where the knock-on effects of an increase in one species could be either positive or negative or on another species, depending on circumstances.

For example, suppose that we have a system with four variables. P and Q are species, R and S are environmental resources. We shall assume that the growth rate of P (to be balanced against its natural death rate) depends on the availability of both resources R and S, whereas the growth rate of Q depends solely on R, and in consuming R Q (unlike P) increases the amount of S available. We can think of Q feeding on R and excreting S. When S is abundant and R is scarce, P and Q are competing for the same limited resource. But when R is abundant and S is scarce, then any increase in the numbers of Q will (through increasing the supply of S) act for the benefit of P. Thus under different environmental circumstances the (indirect) effect of Q on P can be either positive or negative. In the next section we discuss such 'ambiguous' functions, where the sign can change according to circumstances.

Ambiguous Functions

Thomas and colleagues (Thomas and D'Ari, 1990; Thomas and Kaufman, 2001a, 2001b; Kaufman and Thomas, 2002) discuss the roles of positive and negative feedback in nonlinear biological systems. It so happens that their main interest is in the positive feedback circuits that lead to multistationarity, or switching, in genetic regulatory circuits. Nevertheless, much of their analysis can be applied to investigating issues of negative feedback circuits leading to homeostasis or stability. As with May, they are considering a dynamic system of n variables where many (but typically not all) pairwise interactions are present. This leads to the same connectance or Jacobian matrix. But unlike May they explicitly note that in the general nonlinear case the strengths (and indeed possibly the signs) of these interactions will vary throughout phase space.

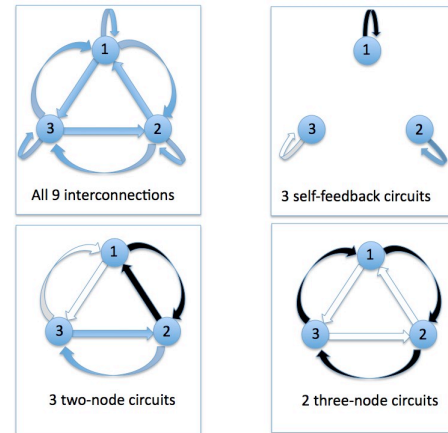


Figure 4: The eight potential circuits, differentiated by shading, within a system of 3 variables fully interconnected

Following their analysis, we note that any connectance matrix A can be considered as composed of multiple overlapping feedback circuits.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

For example the circuit $z \rightarrow x \rightarrow y \rightarrow z$ is characterised by the connection weights $\langle a_{31}a_{12}a_{23} \rangle$. For any such circuit, the indices are circular permutations of each other. The full list of potential circuits, in such a 3-variable systems, as sketched in Figure 4 is: $\langle a_{11} \rangle$, $\langle a_{22} \rangle$, $\langle a_{33} \rangle$, $\langle a_{12}a_{21} \rangle$, $\langle a_{23}a_{32} \rangle$, $\langle a_{31}a_{13} \rangle$, $\langle a_{12}a_{23}a_{31} \rangle$, $\langle a_{21}a_{13}a_{32} \rangle$. If one or more of the connections in such a circuit is zero, that circuit as a whole is non-functional; but otherwise, a count-up of the number of negative connection weights decides whether that individual feedback circuit constitutes a negative feedback (odd number of negatives) or positive feedback (even number). The limiting case of such a circuit is that constituted by self-feedback, given by the term a_{ii} on the main diagonal; that minimal circuit will be non-functional, negative-feedback or positive-

feedback depending on whether its value is zero, or its sign is negative or positive.

Thomas and Kaufman (2001a) defined a full-circuit as those circuits and unions of disjoint circuits that involve all the variables of a system. Hence in this 3-variable system, there are six possible full-circuits:

$$\langle a_{11} \cdot a_{22} \cdot a_{33} \rangle, \langle a_{11} \cdot a_{23} a_{32} \rangle, \langle a_{22} \cdot a_{31} a_{13} \rangle, \langle a_{33} \cdot a_{12} a_{21} \rangle, \langle a_{12} a_{23} a_{31} \rangle, \langle a_{13} a_{32} a_{21} \rangle$$

These correspond to the terms of the determinant of the Jacobian matrix. For any one such full-circuit, considered in isolation, the type of steady state this generates will be determined entirely by the signs, positive or negative, of the various component circuits that comprise this full-circuit. Given that in nonlinear systems any (or all) connection strengths can vary according to position in phase space, and given that when such a connection changes sign it will change the sign of any component circuit of which it is part, we can see that this will alter the type of steady state generated.

This highlights the significance of those connection strengths in a nonlinear system that, varying as one moves through phase space, change in sign: so-called ‘ambiguous’ interactions. These generate ambiguous circuits, and thereby ambiguous full-circuits. Such changes in sign, in one or many such ambiguous connections, carve up the phase space into different regions, and one can expect the properties of steady states to differ from one such region to the next. This gives the richness of possibilities to nonlinear systems that is missing from the linear ones.

Basic Classes of Ambiguous Functions

Linear systems have no ambiguous functions. To move from this to systems “... which in general may obey some quite nonlinear set of first-order differential equations.” (May 1972) is rather too unconstrained to start with. But there are two basic classes of ambiguous functions (to add to linear functions) that might be first considered.

The first class is oscillatory, such as a sine wave, since the slope changes in sign repeatedly and indefinitely. As we have seen above in Figure 3, the combination of a straight line and a sine wave can lead to an arbitrary number of equilibria that will alternate between stable and unstable. Going further, it can be shown (Kaufman and Thomas, 2002) that a system of 3 variables:

$$\frac{dx}{dt} = -bx + \sin(y); \frac{dy}{dt} = -by + \sin(z); \frac{dz}{dt} = -bz + \sin(x);$$

can, depending on the parameter b , move from having a single steady state for $b > 1$, through multiple steady states as b decreases, with the number of steady states tending to infinity as $b \rightarrow 0$. The dynamics change from simple to chaotic, with periodic or multiperiodic windows. It is the unlimited repetition of sign-change, of ambiguity, that produces the richness of possibilities.

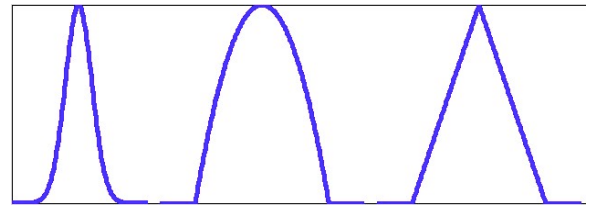


Figure 5: Three broadly similar ‘hat-functions’: gaussian, truncated parabola, and witch’s hat.

Retreating to the simpler end of nonlinear functions, we can choose to limit ourselves to unimodal or ‘hat-shaped’ functions whose slopes have a single change of sign. The examples in Figure 5 share the property of dropping to zero (or approaching zero in the case of a gaussian) each side of a central region. If we take any of these hat-functions as $y = H(x)$, this could represent a viability function of an organism or species y that can only survive (in the case of the gaussian version: survive to any significant level) within some range of values of an environmental variable bounded above and below. These can be considered amongst the most basic of ambiguous functions, and it turns out that they do indeed play a crucial role in giving rise to homeostasis, or a particular form of stable equilibrium, in Daisyworld models. These we shall review briefly.

Lovelock introduced the Daisyworld model (Watson and Lovelock, 1983) as a possible explanation of how organisms coupled in mutual feedback with some environmental variable could form a homeostatic system, biotic-environmental, as is proposed in the Gaia Hypothesis (Lovelock, 1972). The Faint young Sun paradox (Sagan and Mullen, 1972) suggests that despite the heat output of the sun changing significantly over the last few billion years, the planetary climate has maintained itself around the temperatures conducive for life. The Gaia Hypothesis suggests that this arises through homeostatic properties of the interactions between biota and environment. In the Daisyworld model the organisms (Daisies) have a viability whose dependence on temperature is given by such a hat-function; the truncated parabola version is used in (Watson and Lovelock, (1983). In turn, through differential absorption or reflection of sunlight, these Daisies had a positive or negative affect on the same local temperature that influenced their viability. Such systems can be analysed for stability in the context of noise or perturbations at two levels.

In the first instance, any equilibrium state of such a system can be analysed for stability or instability in the presence of small levels of noise; only stable equilibria will persist, and only stable equilibria that have the biota (Daisies) within their viability zone are relevant. But the main interest of Daisyworld models is the extent to which such stable equilibria can persist in the face of major systemic external perturbations, such as major changes in heat output of the sun. It turns out that the Daisyworld temperature is maintained within the viability zone for significantly greater ranges of solar forcing *with* the biotic feedback to the local temperature, as compared to *without* such feedback. This homeostasis arises from the ambiguous nature of the hat-function.

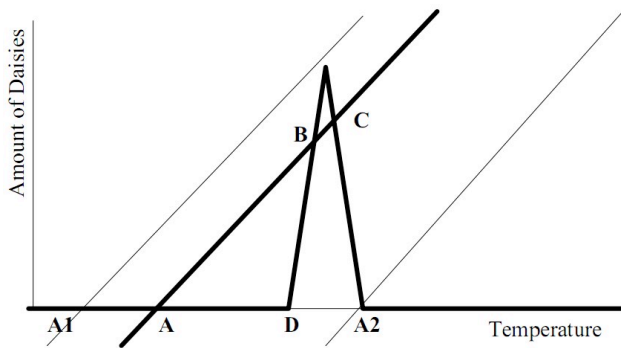


Figure 6: The witches hat function represents the dependency of Black Daisies on local temperature.

Harvey (2004) showed how a simplification of the Daisyworld model produced the same effects, using a witches-hat function. A reduced version of such homeostatis can be shown with just one species of Daisies, e.g. Black ones. With Y black daisies, local temperature T , level of solar forcing S , then for suitable constants k_1 k_2 we have:

$$\frac{dY}{dt} = H(T) - Y; \frac{dT}{dt} = S - k_1 T + k_2 Y;$$

The equilibria are shown where the corresponding lines intersect in Figure 6. The different sloping lines, intersecting the temperature axis at A1, A, A2, correspond to different possible levels of solar forcing. It can be seen that, depending on the level of solar forcing, there is either one equilibrium (e.g. at A1 or lower temperatures, or at A2 and higher temperatures) or three (e.g. A, B, C). This latter case gives us: a possible stable equilibrium with zero Daisies at A; or an unstable equilibrium with Daisies at B, the instability being despite the temperature being viable; or a further stable equilibrium at C with Daisies present within their temperature viability-zone. This latter stable equilibrium is the focus of interest, and we consider the range of solar forcing for which C exists; i.e., for which there is a stable population of Daisies within the local temperature viability zone. From inspection of Figure 6, we can see that the biotic feedback (from black Daisies increasing local temperature) has given rise to viable local temperature over a wider range of solar forcing (corresponding to the range A1↔A2 in the figure) than in the absence of such feedback (corresponding to D↔A2, the viability range of the hat-function).

Thus the presence of black Daisies extends the range of viability towards lower solar forcing (the ‘faint young sun’); conversely, white Daisies (giving rise to a line with the opposite slope in the Figure) would extend the range of viability towards higher solar forcing, a hotter sun. This increased range of homeostasis arises from the ambiguous nature of the hat-function, that gives rise to extra possible equilibria.

Rein Control

It is noted in Harvey (2004) that the different colours of Daisies in the Daisyworld model perform the roles of ‘reins’

in a Rein Control system (Clynes, 1969). Clynes observed that homeostasis in biological systems typically arises through (at least) two channels for sensing and regulating variables: one (or more) in one direction from the norm, and another (or others) in the other direction. The black and white Daisies in the Daisyworld model perform these two roles.

This is in contrast to the usual approach of a control engineer in regulating a variable: to use a single channel to sense the current error, in whichever direction, and apply negative feedback based on such error. The rein metaphor is appropriate for Rein Control since a rein can only pull and not push. In the Daisyworld model black Daisies on their own can only regulate against a sun that is too faint, white Daisies against a sun that is too strong. The necessary and sufficient conditions under which these two opposing reins can interact to produce homeostasis, or stable regulation against solar forcing in both directions, are discussed in (Harvey, 2004).

The conventional control engineer’s regulation will specify a set-point or optimum value, and then regulate towards it by negative feedback. It should be noted that the Rein Control, as demonstrated in the Daisyworld model, is not such an optimizing process; the Daisies are never maintained at their optimum value, the peaks of their viability functions. Instead, the regulation process tends to keep them away from the boundaries of their viability range. Intuitively, this can be compared to two different strategies for walking down a corridor. The optimizing strategy would be to regulate movement towards an optimal line painted in the corridor centre, whereas the Rein Control strategy would be to regulate movement away from whichever wall is more pressing.

It should be mentioned here that there is an alternative interpretation of Rein Control, as in (Saunders et al, 1998). In this version, called ‘Integral Rein Control’ the model is so set up that it does indeed work towards a specified set-point. This is in contrast with the non-optimizing version of Rein Control in (Harvey, 2004).

It can be demonstrated (Harvey, 2004) that an artificial system (a simulated robot) coupled with the environment via an arbitrarily large number of interactions comprising hat-functions (on sensory inputs) and linear functions (on consequent outputs) could find a homeostatic equilibrium. This is so even if the signs of the linear functions are set positively or negatively at random, and the relevant parameters are varied across some two orders of magnitude. Other examples of systems with multiple interacting component parts achieving equilibrium through the use of hat-functions can be found in ((Dyke et al, 2007) and (McDonald-Gibson et al, 2008). In these cases there was a single environmental variable, and numerous biotic variables subject to hat-function viability limits. The Daisystat (Dyke, 2010) extends this approach to multiple environmental variables.

Discussion

The formal conclusions of May (1972), following on from the early computational simulations of Gardner and Ashby (1970), are accurate for complex *linear* systems of many parts, with the specific constraints spelt out there; as the number of parts grows, and as the connectance increases, the probability of stability at the unique equilibrium point will

decrease. The unfortunate programming error of G&A, apparently unnoticed for some three decades, grossly exaggerated the step-nature of the transition and may have misled many. But for those interested in complex systems of many interacting parts as models of many biological, social or artificial – brains, ecosystems, planetary climate systems, social systems, financial systems – their interactions will likely be *nonlinear*, and not subject to those constraints. The lessons can be summarized in three horse-related metaphors.

Clearing out the Stable

Despite the claims in May (1972) to be generalizing to nonlinear systems, the extension is invalid. The method, through linearization around an assumed single equilibrium point, will at best give local stability; there may be many potential equilibria, and global stability can arise through stability at just one of these. There is no sensible method for identifying the location of that so-called ‘unique’ equilibrium point. With nonlinear interactions, the size and sign of the terms in the Jacobian connection matrix varies according to position in phase space, and there is no attempt to characterize what different values these may have at many different possible equilibria. To be rigorous, the probability of global stability would depend on assessing the (different) probabilities of local equilibria, and combining these to calculate the probability that at least one was stable. No attempt at this was presented in (May, 1972).

The flaws in the reasoning are deep-seated and fatal. Despite the established nature of these two foundational papers, and their relevance to linear systems, it is argued here that, as a basis for considering stability in nonlinear systems, they should be rejected and swept away.

Opening the Stable Doors

Those who have accepted the validity of the traditional models, yet sought an explanation of why we do see so many stable natural systems, have typically looked for further constraints to impose on the models. The generic models may have a high probability of being unstable as complexity increases, so the reasoning went – but perhaps some small specific subset of these generic models would have a better chance.

In contrast, what is advocated here is the opening up of the original constraints, not the addition of fresh ones. When different parts of the system can have nonlinear effects on other parts, it is in particular those functions with ‘ambiguous’ signs (Thomas and Kaufman, 2001a, 2001b) that open up possibilities of multiple equilibria. The ‘hat-functions’ associated with viability profiles are plausible for many systems, and immediately offer both positive and negative slopes. Oscillations, such as sine waves, offer the possibility of an unlimited number of potential equilibria. The search for potential equilibria can be opened up through these ambiguities.

Rein Control

The ambiguous hat-functions of Daisyworld models lead to forms of regulation characterized by a balance of opposing forces, rather than regulation about a set-point. This is characteristic of Rein Control, where it can be shown that

such viability functions can extend the range of circumstances in which appropriate stable equilibria can be found.

Acknowledgments

I thank Nathaniel Virgo and David Waxman for discussions, and for checking that my code was working properly when it failed to replicate G&A. NV found the prior reference to this anomaly (Solow et al., 1999).

References

- Clynes, M. (1969). Cybernetic Implications of Rein Control in Perceptual and Conceptual Organization. *Annals N.Y. Acad. Sc.*, 156:629-664.
- Dyke, J., McDonald-Gibson, J., Di Paolo, E. and Harvey, I. (2007). Increasing Complexity can increase Stability in a Self-regulating Ecosystem. In Almeida e Costa, F., Rocha, L. M., Costa, E., Harvey, I. and Coutinho, A., editors, *Eur. Conf. Artificial Life ECAL 2007*, pages 123–132. LNCS 4648, Springer.
- Dyke, J. (2010). The Daisystat: A model to explore multidimensional homeostasis. In H Fellermann, M Dörr, M. Hanczyc, L Laursen, S Maurer, D Merkle, P-A Monnard, K Stoy and S Rasmussen (Eds.): *Artificial Life XII*. MIT Press.
- Gardner, M. R. and Ashby, W. R. (1970). Connectance of Large Dynamic (Cybernetic) Systems: Critical Values for Stability. *Nature*, 228:784.
- Haldane, A. G. and May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, 469:351–355.
- Harvey, I. (2004). Homeostasis and Rein Control: From Daisyworld to Active Perception. In Pollack, J., Bedau, M., Husbands, P., Ikegami, T. and Watson, R. A., editors, *Artificial Life IX*, pages 309–314. MIT Press, Cambridge, MA.
- Kaufman, M. and Thomas, R. (2002). Emergence of complex behaviour from simple circuit structures. *C. R. Biologies*, 326:205–214.
- Lovelock, J. E. (1972). Gaia as seen through the Atmosphere. *Atmospheric Environment*, 6(8):579–580.
- May, R. M. (1972). Will a Large Complex System be Stable?. *Nature*, 238:413–414.
- May, R. M. (1973). *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, New Jersey.
- McDonald-Gibson, J., Dyke, J. G., Di Paolo, E. A. and Harvey, I. R., (2008). Environmental Regulation can arise Under Minimal Assumptions, *J. Th. Biol.*, 251(4):653–666.
- Sagan, C. and Mullen, G. (1972). Earth and Mars: Evolution of Atmospheres and Surface Temperatures, *Science*, 177(4043):52–56.
- Saunders, P. T., Koeslag, J. H. and Wessels, J. A. (1998). Integral Rein Control in Physiology, *J. Th. Biol.*, 194(2):163–173.
- Solow, A. R., Costello, C. and Beet, A. (1999). On an Early Result on Stability and Complexity. *The American Naturalist*, 154(5):587–588.
- Thomas, R. and D’Ari, R. (1990). *Biological Feedback*. CRC Press, Boca Raton, Florida, USA.
- Thomas, R. L. and Kaufman, R. (2001a). Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior. *Chaos*, 11(1):170–179.
- Thomas, R. L. and Kaufman, R. (2001b). Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos*, 11(1):180–195.
- Watson, A. J. and Lovelock, J. E. (1983). Biological Homeostasis of the Global Environment: the Parable of Daisyworld. *Tellus*, 35B:286–289.