1 Introduction to Nonlinear Dynamics and Complexity

STEPHEN J. GUASTELLO AND LARRY S. LIEBOVITCH

Elephants and Horses

Things change. Sometimes they change gradually, sometimes dramatically. The prevailing concept of change in psychology consists of only one form of change, linear change, which is simply undifferentiated, and with the assumption that outcomes are proportional to inputs in a straightforward manner. The overreliance among psychologists and others on the general linear model as a statistical tool for depicting change has only served to reinforce this monochrome conceptualization of change. Perhaps the most significant deviations from the concept of linear change are the concepts of *equilibrium* and *randomness*. For most intents and purposes, the concept of equilibrium has been used to describe places or times when change stops occurring. Randomness suggests that the changes are unpredictable and not explicable by any known concepts or predictors.

Nonlinear dynamical systems (NDS) theory significantly enriches our capability to conceptualize change, and it provides a rich array of constructs that describe many types of change. The concept of equilibrium is no longer specific enough to describe either the change or the events that surround the point where change stops. The new constructs are the attractors, bifurcations, chaos, fractals, self-organization, and catastrophes. As this chapter explains, each of these constructs contains several more, including those associated with the "complexity" of a system. Importantly, change is not proportional to inputs. Large inputs sometimes produce small results, and a small input at the right time can produce a dramatic result.

Psychology is not the first science to break out of the linear rut. According to Stewart (1989), the physical sciences made the transition more than a half-century ago:

> So ingrained became the linear habit, that by the 1940s and 1950s many scientist[s] and engineers knew little else.... [W]e live in a world which for centuries acted as if the only animal in existence was the elephant, which assumed that holes in the skirting-board must be made by tiny elephants, which saw the soaring eagle as a

wing-eared Dumbo, [and] the tiger as an elephant with a rather short trunk and stripes (1989, p. 83–84).

Some nonlinear phenomena, particularly chaos, force us to reconsider what it means for an event to be random. According to Mandelbrot (1983), the word *random* came into English by way of a medieval French idiom meaning "the movements of a horse that the rider cannot predict" (p. 201). Later-day statistics and experimental design have placed a great deal of emphasis on what the rider does not control, resulting in the notion that variability in observations is either "due to the model" or "due to error." NDS places more emphasis, however, on the horse's point of view: There are reasons for the horse's motions, and its rider can get used to them or not. Less metaphorically, simple deterministic equations can produce processes that appear random, but closer scrutiny may indicate that they are not. It follows that a lot of so-called error variance can be accounted for if we can identify the processes that generated the observations; those processes are most likely to be overwhelmingly nonlinear.

In contemporary colloquial English, people speak of "random events" when referring to events that occur without any apparent connection to prior events or to any clues about which the speaker is aware. In NDS, these events are called *emergent phenomena*. Their important features are their disconnection with recent past events and that they occur without the deliberate action of any person or agent. In other words, those phenomena are novel, and they can sometimes be clear examples of nonlinear events and deterministic processes, as will become apparent throughout this book.

General Systems Theories and Paradigms

One of the conceptual foundations of NDS is the general systems theory (GST), an interdisciplinary theory that contains rules, propositions, and constructs that extend beyond the confines of a single academic discipline. Within the realm of GSTs, there is a strong representation of mathematically centered theories; this approach is usually attributed to von Bertalanffy (1968) and Wymore (1967). NDS is also an example of GST that is centered on principles of mathematical origin.

A GST can be regarded as a metatheory and as an overarching methodological approach in some cases. A metatheory is a theory that organizes concepts, objects, or relationships inherent in several local (or specific) theories and places those into an overarching conceptual framework. The development of such theories facilitates interdisciplinary cross-fertilization because at the metatheoretical level, scholars from different disciplines will share a conceptual framework. For instance, a theory that successfully solves a problem in macroeconomics with principles of evolutionary biology would qualify as an example of such cross-fertilization. Thus specific objects can be interchanged from one application to another, but the relationships among those objects could remain approximately the same. Another approach might show interchangeability of objects and relationships, but the "blueprint" that defines the metatheory would tell the scientist where to look for objects and relationships that could be useful. GST and NDS can both be seen as examples of metatheories. In fact, the same concepts that are encountered in this book can also be found in texts that are concerned with the subject matter of other academic disciplines, such as economics and biology.

When viewed as a methodology, a general systems analyst would typically begin with a description of the phenomenon of interest and assess the relevance of particular tenets of the GST to that phenomenon. For example, if one were to observe a nonnormal frequency distribution with a sharp peak near the left and a long tail to the right, one might hypothesize that a power law distribution and the dynamics that usually go with it (see later in this chapter) were involved. The next step would be to create a model to represent the system using the tools and constructs of the GST plus any adaptations that are specific to the application. The model-making process typically draws on the past successes and failures encountered with applications of the general theory. If the GST is truly meritorious, the knowledge gained from a successful application would increase the knowledge about the core principles of the theory and thus facilitate the hunt for further applications. Hence one good application serves as a metaphor for another (Guastello, 1995, pp. 4–6), and NDS applications have indeed found their way into many applications within physics, chemistry, biology, psychology, economics, sociology, and political science. With repeated successes, one can develop one's intuition as to where dynamical events of different types might be found.

Scholars have pondered whether NDS can be considered a paradigm of science itself (Allen & Varga, 2007; Dore & Rosser, 2007; Fleener & Merritt, 2007; Goerner, 1994; Guastello, 1997, 2007; Ibanez, 2007). A new scientific paradigm would represent a new approach to a wide range of problems and ask entirely different classes of questions. It would pursue its answers with its own set of standards and challenges. Thus a new paradigm unearths and explains phenomena that could not have been approached through pre-paradigmatic means. Additionally, the new paradigm could be shown to provide better, more compact, and more accurate explanations for existing questions.

NDS shows other symptoms of paradigmatic behavior beyond its new perspective on randomness, nonlinear structures, and system change. For one thing, the concept of a *dimension* is not what it used to be. Although society at large has assimilated the four basic dimensions of Euclidean space and time, mathematicians have rendered 5 through N dimensions as comfortable abstractions, and social scientists have extensively analyzed complex multidimensional data spaces using factor analysis and multidimensional scaling. NDS offers a new development: fractional dimensions. There are entities between lines and planes, planes and volumes, and so on. Fractional dimensions, or fractals, are described later in this chapter. It is worthwhile to point out early that the fractals serve as good indicators of the complexity of a system.

Our notions about the nature of a system have shifted as well. In Newton's view of a mechanical system, the function of the whole can be understood by understanding the function of each of the parts. A correction of a flaw in a system can be accomplished by removing and replacing one of its parts. In a *complex adaptive system* (CAS), however, the whole is greater than the sum of its parts. This description should surely ring familiar to psychologists because of the Gestalt laws of perception. The parts of the system, perceptual or otherwise, interact with each other, shape each other, and pass information around, and they are not replaceable or removable without fundamentally altering the dynamics of the system as a whole. Furthermore, attempts to correct "flaws," or to change otherwise a part in a CAS, often do not succeed because the parts adapt in such a way as to protect the system from the intrusions of the outside tinkerer. By the same token, a CAS can survive intrusions or assaults by the same means; the tendency for brain tissue to pick up a function that was lost by a nearby tissue area is an example. These self-organizing tendencies are a primary area of interest in the study of NDS, and they are expanded further in this chapter and throughout the remainder of this book.

Has NDS produced better explanations of phenomena? The answer in each chapter that follows is "yes." It has described phenomena that could not be described in any other way, especially when temporal dynamics are concerned. In some cases, they provide an organizing center where several partially relevant theories were needed to describe a phenomenon; work-group coordination is one such example. Although the explanations may appear more complex in one respect, they become more efficient to the extent that the general principles transcend many problems in psychology and also expand beyond the usual confines of psychology.

Does it account for more of the data than conventional linear models? In most cases reported in this book, there was no linear alternative available, so the answer is a simple "yes." In the cases in which it was possible to compare R^2 coefficients for linear and nonlinear models, the cumulative advantage is 2:1 in favor of the nonlinear models (Guastello, 1995, 2002). That is to say that about 50% of the explanation for a phenomenon comes from knowing what dynamics are involved.

Elements of NDS Functions

It is no secret that the central concepts of NDS have mathematical origins, some of which might qualify as exotic even among mathematicians. Fortunately, the majority of the concepts can be represented in pictorial or graphic form, which helps interpretations greatly. In psychological applications, we only need to use the products of the mathematics; we do not need to reperform derivations, solve differential equations, or anything of that nature. The principles are analytic when applied to psychology, nonetheless. Some equations have a great deal of meaning, and they serve to codify the dynamical principles concisely. There are some novel computations that are often used in NDS research, and some of them can be rendered statistically, which is perhaps more compatible with the analytic habits of social scientists.

The equations that we do encounter are all *functional* and sometimes structural as well. Some functions that we encounter state that a dependent measure Y is a nonlinear function of another variable X. Some functions are *iterative*, in which Y at Time 2 (Y_2) is a function of Y_1 and other variables in the system.

In an iterative function, a time series of Y is produced by running Y_1 through the equation to produce Y_2; Y_2 is then run through the equation to produce Y_3, and so on, thereby producing a forecast of future behavioral patterns. Many psychological phenomena that occur over time are iterative in nature, such as the flow of a conversation or series of conversations between two people; see Chapters 9 and 11 through 13 in this volume. If we do not know what equation is the best representation for the phenomenon, it helps to recognize when an iterative function could be implied.

Structural equations are similar to functional equations, but they have the added feature of allowing, if not encouraging, the user to plug complex functions into a position that might have been signified by a single variable. For instance, we might have a function containing a bifurcation effect called c, but c might consist of several psychologically defined variables. Sometimes an element that looks as simple as c could turn out to be a constant, a variable, or an entire complex function.

Control parameters are essentially independent variables, with the important difference that they can act in ways that are often more interesting than the simple additive relationships that are found in conventional research designs. Several distinct types of control parameters are described in a later section of this chapter and in subsequent chapters. *Order parameters* are essentially dependent measures in the social scientist's worldview. There may be more than one order parameter in some complex dynamical systems, however. Order parameters within a system might be completely independent of each other, or they might interact with each other as they evolve over time.

The basic principles of nonlinear dynamics are presented next. In addition to the references that are interspersed in the text, we recommend Abraham and Shaw (1992), Bassingthwaighte, Liebovitch, and West (1994), Kaplan and Glass (1995), Korsch and Jodl (1999), Liebovitch (1998), Nicolis and Prigogine (1989), Puu (2000), Scott (2005), Sprott (2003), and Thompson and Stewart (1986) as broad technical references to the mathematical literature.

The chapter concludes with an overview of the analytic techniques that one might apply to real data to determine which types of dynamics are present. The authors of the subsequent chapters elaborate on their uses of particular techniques.

Attractors and Chaos

An *attractor* can be seen as a box of space in which movement could take place or not. When an object, represented by a point, enters the space, the point does not leave, unless a strong enough force is applied to pull it out. An attractor, like a magnet, has an effective range in which it can draw in objects, known as its *basin*. Some attractors are stronger than others, and stronger attractors have a wider basin.

An attractor is regarded as a stable structure because all the points within it follow the same rules of motion. There are four principal types of attractors: the fixed point, the limit cycle, toroidal attractors, and chaotic attractors. Each type reflects a distinctly different type of movement that occurs within it. Repellors and saddles are closely related structures that are not structurally stable.

Fixed-Point Attractors

As its name suggests, a *fixed-point attractor* is one where the trajectories of points that are pulled within it gravitate toward a fixed point. Figure 1.1A illustrates points that are being pulled in from all directions toward the epicenter. A point might start its path by shooting by the attractor, and might even have a chance of breaking free, but is pulled into the attractor, where it rests. Figure 1.1B is a *spiral* attractor, also known as a *sink*; the point spirals into the epicenter rather than moving in directly.

Attractor behavior that is viewed in a time series would look similar to the trends shown in Figure 1.1D, which would correspond most closely to the case where points are being pulled in from all directions. The structural equation that generates that type of motion is:

$$Z_2 = \theta_1 \exp(\theta_2 Z_1). \tag{1.1}$$

Equation 1.1 shows the function in iterative form, where the order parameter Z_2 is a function of itself at a previous point in time (Z_1). θ_i are empirical weights that can be determined through nonlinear regression. θ_1 is a proportionality constant (this is likely to become important when two or more attractors are involved in the dynamical field); θ_2 is the more important of the two weights because it reflects how quickly the points are converging onto the fixed point. Notice that the trajectories of the points in the figure all started with different initial values but ended up on the same attractor. If θ_2 is a negative number,

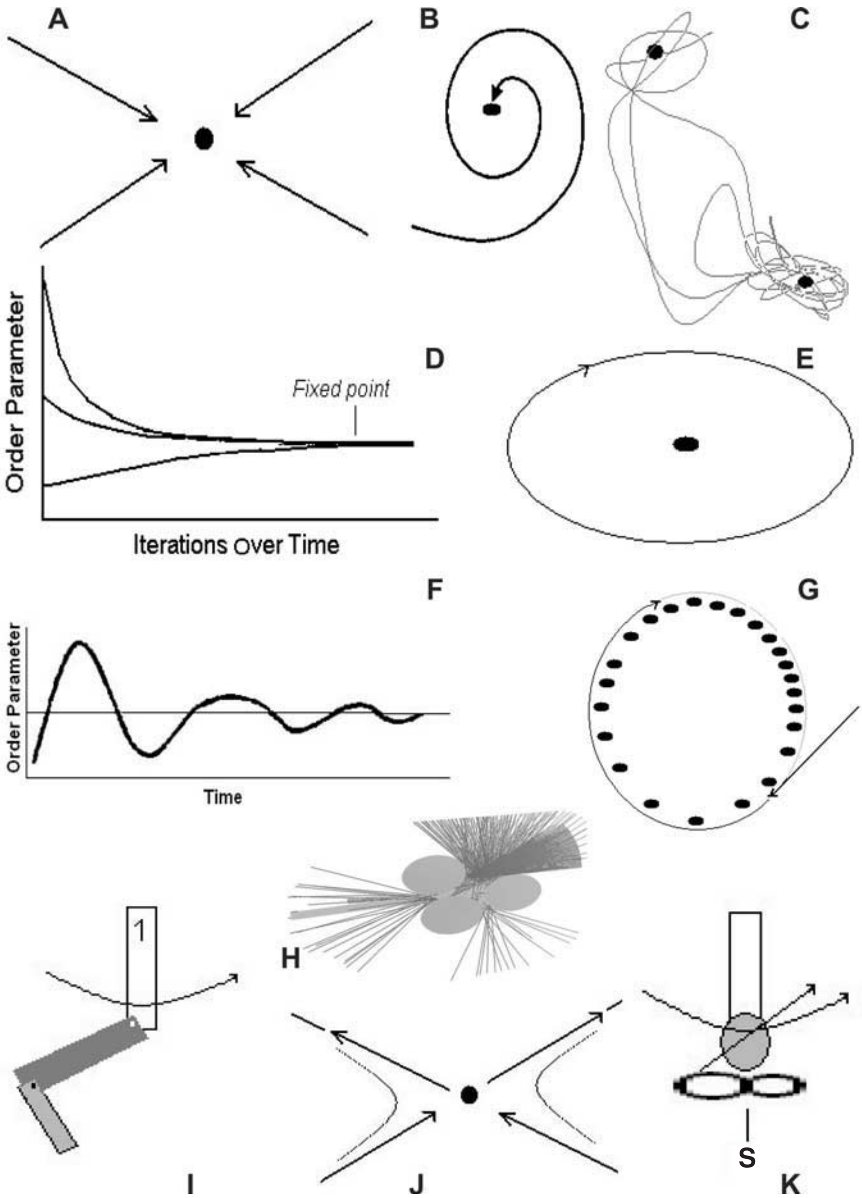

Figure 1.1. Gallery of basic dynamics: A, fixed point attractor (Eqs. 1.1 and 1.2); B, spiral attractor (Eqs. 1.3); C, trajectory of a control point in a field containing two fixed points; D, time series of points gravitating to a fixed point; E, limit cycle (Eqs. 1.4 and 1.5); F, time series for a dampened oscillator (Eq. 1.6); G, limit cycle created by organizing fixed points into a circle; H, system of three repellors; I, three coupled oscillators; J, saddle point; K, saddle created by a perturbed pendulum.

the points are indeed converging; otherwise the points' trajectories would be expanding exponentially.

Another equation for a fixed-point attractor that is used often is:

$$Z_2 = K/[1 + \theta_1 \exp(\theta_2 Z_1)]. \tag{1.2}$$

A particular value of Z in Eq. 1.2 is recognized as the asymptotic point, which is designated as K. The regression weights would indicate how quickly the points were reaching K. A close variation on Eq. 1.2 is one in which the researcher wants to determine whether an exogenous independent variable X is responsible for the movement of Z to K. In that case, one would substitute X_1 for Z_1.

A spiral attractor is a more complicated phenomenon to analyze, but we can describe it nonetheless. Imagine that the trajectory of points on the spiral has the usual X and Y coordinates, and both X and Y are changing over time, t. We would need to write equations for both the X and Y functions (Puu, p. 21):

$$X = \theta_1 \exp[\theta_2 t \cos(\theta_3 t + \theta_4)] \tag{1.3}$$
$$Y = \theta_5 \exp[\theta_2 t \cos(\theta_6 t + \theta_7)].$$

Equations 1.3 allow for the full range of possibilities that the spiral can be perfectly circular or stretched (elliptical) by some amount in the vertical or horizontal direction by changing θ_1, θ_3, θ_5, or θ_6. θ_4 and θ_7 denote the lower boundaries of the cosine functions in real numbers. Notice that the forcing function, θ_2, is common to both axes. If θ_2 were negative, the system would be converging to the fixed point and thus stable. If θ_2 were positive, however, the points would be floating away from the epicenter, and the spiral would be unstable. The spiral attractor functions become more interesting when we consider repellors and saddles a bit later.

Figure 1.1C illustrates what can happen in a dynamic field containing two fixed-point attractors. A point that might have been entered into the dynamic field in the neighborhood of the upper attractor makes a few irregular turns around the attractor and then travels to the other attractor. Nothing prevents the point from revisiting the first attractor before landing on the second one, where it appears to be spending most of its time. The convoluted pathways take different forms, depending on the relative strength of the two attractors, their proximity to each other, and the exact location where the traveling point is entered.

Figure 1.1C is, in essence, the three-body problem that got the entire study of nonlinear dynamics going. Henri Poincaré made a profound contribution in the 1890s by identifying the problem that we did not have the mathematics available to characterize the path of a "speck of dust" in the neighborhood of two celestial bodies, that is, two strong attractors (Stewart, 1989). Those trajectories were eventually characterized as chaotic several decades later.

Limit Cycles

A *limit cycle attractor* is also known as a periodic attractor. Its behavior is cyclic, in the same way that the earth orbits the sun or the moon orbits the earth. The motion of the point is thus depicted in Figure 1.1E. Once again, if a traveling point enters the basin of attraction, it does not leave. Points entering from the outside of the limit cycle are drawn into the attractor. Points that happen to be inside it are pulled outward into the orbit but do not move toward the epicenter in the way that fixed-point or spiral attractors do. Instead, there is an ongoing oscillation around the epicenter value. Oscillating functions are common in biology (May, 2001) and economic business cycles (Puu, 2000), although both types of oscillators tend to become more complex by the presence of other oscillators in the niche or economy.

A classic limit cycle can be generated by a simple function:

$$Z = \sin(X). \tag{1.4}$$

Again, one might substitute time, *t*, for the exogenous variable, *X*. One might characterize the function in both an iterative and statistically tractable form as:

$$Z_2 = \theta_1 \sin(Z_1 - \theta_2) - \theta_3 \cos(Z_1 - \theta_4). \tag{1.5}$$

Fourier analysis or spectral analysis is often employed to separate compound oscillators. This is a common practice in auditory signal processing where the sinusoidal components are additive. The generic structure of a compound oscillator is defined as a sine function, a subtracted cosine function, plus another sine function, and so on. θ_1 and θ_3 are forcing constants that adjust the amplitude of the sinusoidal functions. θ_2 and θ_4 are lower limits of the oscillating variable Z and depend on the scale used to measure Z. Note that if θ_2 or θ_4 becomes a variable, then f(Z) becomes aperiodic.

Another way to obtain a limit cycle is to place a number of point attractors in a circular configuration. If a point is injected into the neighborhood of this configuration, it will exhibit the behavior of a limit cycle (Fig. 1.1G).

We can also have *dampened oscillators*, which exhibit periodic behavior, but the amplitude of the fluctuations gradually becomes smaller until the order parameter gravitates to a fixed point. The temporal signature of a dampened oscillator would look like the time series that is shown in Figure 1.1F. A dampened oscillator might not be easily discerned from a spiral fixed point. One would be looking for a control variable that induced the dampening effect. Equation 1.6 would be a likely place to start:

$$Z_2 = \theta_1 \exp[\theta_2 X \cos(\theta_3 Z_1 + \theta_4)], \tag{1.6}$$

where $\theta_2 X$ is the variable that induces the dampening with its regression coefficient. Note that Eq. 1.6 is structurally the same as Eq. 1.3, with a substitution of X and Z_1 for t.

In light of the potential similarity of spiral points and dampened oscillators, psychologists who might encounter these dynamics should sample a long enough time series to view the full dynamics of the system. The concept of "restriction of range" is probably familiar to most psychologists from linear statistical analysis. A comparable concept here is *restriction of topological range*, meaning that the full range of topological events or movements needs to be captured in the data set to draw reliable conclusions about the dynamics of the model under study (Guastello, 1995). This principle will recur later in conjunction with other dynamics.

We can also couple oscillators so that two or more are linked to each other. Again using Eq. 1.6 as a starting point, X would be a second order parameter, and it would also have an equation that is structured like Eq. 1.6. We can also imagine a third oscillator Y in the system that is responsible for the sinusoidal forcing of X. To close this system of three variables, let Z be the forcing parameter of Y:

$$\begin{aligned} Z_2 &= \theta_1 \exp[\theta_2 X \cos(\theta_3 Z_1 + \theta_4)], \\ X_2 &= \theta_5 \exp[\theta_6 Y \cos(\theta_3 X_1 + \theta_4)], \\ Y_2 &= \theta_1 \exp[\theta_2 Z \cos(\theta_3 Z_1 + \theta_4)]. \end{aligned} \tag{1.7}$$

According to a theorem by Newhouse, Ruelle, and Takens (1978), three coupled oscillators are minimally sufficient to produce chaos. Chaos, for present purposes, is a highly unpredictable time series of numbers or events that appears random but that is actually the result of a deterministic process. Chaos in the form of coupled oscillators can be readily demonstrated by joining one pendulum to another (Fig. 1.1I), so that they all swing freely. Swing the pendulum marked "1," and watch the highly volatile motion of the others.

Not all combinations of three oscillators will produce chaos, however. Puu (1993) examined different combinations that could occur in intertwined economic cycles and concluded that although some combinations produce chaos, some serve to dampen the volatility of the others in the system. It is also possible to take two or more oscillators that are oscillating at different speeds, couple them, and synchronize the whole set (Strogatz, 2003).

Repellors

A *repellor* is also a box of space, but it has the opposite effect on traveling points. Any point that gets too close to it is deflected away from the epicenter. It does not matter where the traveling point goes, so long as it goes away. Thus repellors characterize an unstable pattern of behavior. We already encountered a type of repellor when we considered the spiral with a positive coefficient.

Another type of repellor is more akin to the fixed point. A system containing three such repellors is shown in Figure 1.1H. A beam of points coming in from the left hits the edge of an attractor basin and is deflected outward, as indicated by the thin lines around the repellor regions. In some cases, the repellors deflect