REVISTA MEXICANA DE ANÁLISIS DE LA CONDUCTA
MEXICAN JOURNAL OF BEHAVIOR ANALYSIS
26, 273-287
NÚMERO 2 (SEPTIEMBRE)
NUMBER 2 (SEPTEMBER)

WHAT IS THE NET WORTH? SOME THOUGHTS ON NEURAL NETWORKS AND BEHAVIOR

¿CUÁNTO VALE LA RED? ALGUNAS REFLEXIONES EN TORNO A LAS REDES NEURALES Y LA CONDUCTA

M. JACKSON MARR[1]
GEORGIA TECH

ABSTRACT

Neural network models have played major practical roles in engineering as well as theoretical roles in cognitive science, and now are being explored in behavior analysis. What contributions can these kinds of models make to a science of behavior? Neural networks and behavioral processes can show similarities in dynamical properties, dependency on some variation of a contiguity mechanism, instantiation of some sort of memory, and operations in accordance with some kind of delta rule leading to a quasi-stable state. However, because these are basically inherent properties of all network models, they are grossly indeterminate. Conversely, they may be best described as "implementory", as opposed to "explanatory" models in the sense that they, with few exceptions, only simulate what they were specifically designed to simulate. This "curve-fitting" quality sets them apart from predictive and otherwise suggestive quantitative models. One rationale for their exploration is their putative value in simulating neural mechanisms in learning. Not only is it the case that we understand little about these mechanisms to begin with, but network models capture virtually none of the complexities of the nervous system at any level. Despite all these difficulties, these models are worth further development and exploration as potentially powerful quantitative approaches to behavior, independent of any possible relations to real neural systems.

Keywords: neural-network models, behavioral processes, neural systems, dynamical properties, learning mechanisms, implementation, explanation

---

[1]School of Psychology, Georgia Tech, Atlanta, GA 30332-0170, USA. E-mail: mm27@prism.gatech.edu.

RESUMEN

Los modelos de redes neurales han jugado papeles importantes tanto en aplicaciones prácticas de ingenieria como en teorías de la ciencia cognoscitiva, y ahora están siendo explorados en el análisis conductual. ¿Qué contribuciones puede hace este tipo de modelos a una ciencia de la conducta? Las redes neurales y los procesos conductuales pueden mostrar similaridades en sus propiedades dinámicas, dependencias sobre alguna variación de un mecanismo de contigüidad, realización de alguna forma de memoria, y operaciones de acuerdo con cierta regla delta que lleva a un estado cuasi-estable. Sin embargo, puesto que estas son propiedades básicamente inherentes a todos los modelos de redes neurales, están gruesamente indeterminadas. Contrariamente, pueden ser mejor descritas como modelos "implementadoras", en oposición a "explicativos", en el sentido de que, con muy pocas excepciones, simulan sólo lo que fueron diseñados para simular. Esta cualidad de "ajuste de curva" los diferencia de modelos cuantitativos que son más predictivos y sugestivos. Una justificación para explorarlos es su valor putativo para simular mecanismos neurales del aprendizaje. No sólo es el caso que entendemos muy poco estos mecanismos, sino que los modelos de redes neurales capturan virtualmente ninguna de las complejidades del sistema nervioso en nivel alguno. A pesar de estas dificultades, estos modelos merecen seguir siendo desarrollados y explorados como poderosas aproximaciones cuantitativas a la conducta, independientemente de cualesquiera relaciones posibles con sistemas neurales reales.

Palabras clave: modelos de redes neurales, procesos conductuales, sistemas neurales, propiedades dinámicas, mecanismos de aprendizaje, implementación, explicación

---

Neural networks have been of considerable practical and theoretical value in a number of domains in engineering, business, economics, the military, indeed, any area where the application and analysis of learning and adaptive systems is of interest. Behavior analysis came rather late to this parade, but, as This Issue attests, has begun to exploit the quantitative powers of the method. Neural networks have long been a source of controversy within cognitive psychology in challenging the traditional rule-based and representational models of complex performance in humans (see, e.g., Ellis & Humphreys, 1999). The debates still rage among the clever practitioners in cognitive science, but within behavior analysis and beyond to more conventional learning theorists, network models are a natural source of experimentation and exploitation (e.g., Schmajuk, (1997). In this essay, I would like to explore briefly a number of aspects of neural networks and their potential role in the analysis of behavior. First, I will discuss the place of both neural networks and behavioral processes in the larger domain of dynamical systems. Then I would like to address the question of what networks can and

cannot tell us about behavior. In that context, I also want to comment on the issue of what networks might or might not have to do with the neural bases of behavior.

Neural networks have a long and distinguished history reaching back at least to the nineteen forties and were creations of such luminaries as Alan Turing, John von Neumann, and Pitts and McColloch (see, e.g., Haykin, 1999 for a review of the history of neural networks). These earlier theoreticians attempted to develop quantitative abstractions of neural interactions, and their achievements laid the foundations for the entire field of cellular automata of which neural networks are a particular, but now very large outgrowth. Modern developments in the theory and practice of systems like cellular automata had to await the availability of fast, efficient, and large-memory computing devices. Cellular automata are essentially programs involving a number of units that interact locally, that is, with immediately adjacent units according to certain rules. Conway's "Game of Life" is perhaps the best known and can show remarkable patterns of organization (see, for example, Coveney & Highfield, 1995). Neural networks are systems of interacting units that differ from conventional cellular automata in that a given unit's influence can extend well beyond adjacent units. Properly structured and programmed, these systems are capable of patterns of organization that are immensely useful as learning and problem-solving programs.

Dynamic Networking

Neural networks, cellular automata, numerous mechanical, thermodynamic, electrical, and chemical phenomena, ecological, genetic, physiological, and evolutionary processes, as well as other selective and adaptive systems, including operant conditioning, all exemplify *dynamical systems*. Dynamical systems theory and its more recent offspring, complexity theory, provide an overarching approach to these and many other processes that demonstrate organizational complexity and emergent properties (e.g., Bak, 1996; Bar-Yam, 1997; Casti, 1994; Coveney & Highfield, 1995; Jackson, 1989; 1990; Kauffman, 1993; Peak & Frame, 1994; Nicolis & Prigogine, 1989).

Fundamentally, dynamical systems theory addresses *change*. Dynamics is understood commonly as a field of physics dealing with the description of how forces act to produce changes in motion. Modern dynamics emerged from classical mechanics, the study of motion first developed systematically by Newton, then elaborated and refined by such great 18th and 19th century mathematical physicists as Laplace, Lagrange, and Hamilton. Modern dynamics was founded largely by Poincaré who wrote a now-classic three-volume study

inspired by attempts to address the n-body problem in astronomy. The problem is to describe the interaction of three or more astronomical bodies subject to Newton's inverse square law and to predict future motions from initial conditions. Poincaré's contributions were enormous not only in solving particular cases of this problem, but in developing general methods to approach complex non-linear systems through essentially geometric or topological descriptions of local and global properties of motion. Modern non-linear dynamics, including chaotic processes and analyses of complex systems of many kinds can be traced to this work (Diracu & Holmes, 1996).

The dynamics of virtually any system from a billiard ball on an elliptical table to an electrical circuit with feedback to predator-prey interactions to population genetics to neural networks are all subject to essentially a common analysis in terms of attractors, stability, instability, dissipation, feedback, organization, and emergence. I'll briefly review some to these features and subsequently relate them to neural network functioning.

The concept of an attractor is fundamental to a description of any dynamical process. Motion, or any change in a system, can be depicted in a special field called a *phase space*. This space, or manifold, may be multi-dimensional, depending on the complexity of the system of interest. A simple case is the motion of a pendulum where we may plot position versus velocity (i.e., a phase space in two dimensions) under a variety of initial conditions. If the pendulum is not subject to any friction and the displacement is small, the dynamics will be described by a set of concentric circles. Such a geometric picture of the motion in phase space is called a *phase portrait*. At the greatest displacement from rest, the velocity goes to zero; as the bob moves through its lowest position the velocity is greatest, and so on to the opposite maximal displacement, etc. The possible states of motion achieved after transients die away are called *attractors*. In the frictionless pendulum example, the attractor is called a *limit cycle*. If friction or damping were taken into account (a form of *dissipation*), the phase-space plots would cycle in toward a center, the *fixed-point attractor*. Dissipation generally refers to a condition where some form of energy is required to drive the system. In the pendulum case, combinations of driving forces and friction can result in multi-or quasi-periodic or even *chaotic attractors*. In the latter case, even though the attractor is confined within regions of the phase space, the motion becomes essentially unpredictable. Chaotic attractors may be called *strange* or *stochastic*, depending on their particular properties. In general, if driven appropriately, non-linear dissipative systems are capable of very complex behavior (e.g., Moon, 1992).

Stability refers to regions of an attractor field wherein nearby motions stay near or eventually merge with an attractor, as in the case of the damped pendulum. Another way of saying this is that nearby trajectories tend to stay

nearby. Unstable attractors are regions where nearby trajectories can diverge from an attractor, so that slightly different initial conditions result in large differences in subsequent motions. A pencil precariously balanced on its point is an example. Such an unstable attractor is also called a *repellor*.

Phase portraits of complex systems may resemble an uneven or rough landscape with many attractors and repellors of various extents of "depths" or "heights ", that is, local regions of stability or instability, called *basins of attraction*. The overall picture is somewhat like a contour map, or a weather map showing circulations around regions of high and low pressure. A boundary region between basins of attraction is called a *separatrix*. One might think of separatrices as ridgelines in a mountainous landscape. Globally, what emerges from the dynamic analysis is a field of attractors defining an organization or pattern of the system.

Feedback refers to conditions in which some outcome or consequence of a system is fed back into the system. Negative feedback contributes to stability in that an outcome of motion tends to return the motion to a previous state. Gravity acts on a swinging pendulum at its maximum displacement to either maintain the limit cycle or, in the case of dissipation, return the pendulum to its ultimate fixed-point attractor where motion ceases. Positive feedback acting alone will lead to instability. For example, vibrations at resonant frequencies may cause a suspension bridge to collapse. Particular combinations of negative and positive feedback, that is, conditions contributing to both stability and instability can lead to complex, emergent patterns, a process sometimes given the name *self organization* (e.g., Bak, 1996). There are numerous examples in nature: crystallization, tornados, the flocking of birds and the schooling of fish, ant colonies, animal coat colors, embryological development, ecological systems, biochemical pathways, and the acquisition and control of patterns of behavior engendered under contingencies of reinforcement.

How do neural networks exemplify characteristics of complex dynamical systems like those mentioned? I should first note that there are many different architectures and programming approaches to these models (e.g., Anderson, 1995; Bozinovski, 1995; Ellis & Humphreys, 1999; Haykin, 1999; Schmajuk, 1997). Most, however, consist of a number of elements or units, arranged in layers, including input and output layers, with connections in various configurations between the units from layer to layer. Rules of unit interaction, feedback arrangements, initial weights, and other specifications are established and the system tested to see if it accomplishes the intended purpose. What is interesting dynamically about these arrangements? The typical programming procedures are essentially a concatenation of coupled non-linear difference or differential equations whose inputs and outputs drive and reflect the ongoing

variations in activities from unit to unit. Often, stochastic features may be added, for example, with respect to activation states, making the system not only non-linear, but also probabilistic in its function. The shifting pattern of activities of the elements, either discretely or continuously as a function of their inputs, and the feedback from the system outputs can be described as a field of attractors, stable, unstable, periodic, quasi-periodic, and possibly chaotic. Stochastic features, curiously enough, often contribute to achieving stability by helping the system into a "lower energy state", that is, minimizing discrepancies between the inputs and the desired outputs. The network pattern of organization defining the successful transformations of inputs to outputs displays a fundamental property of complex dynamical systems, namely *emergence*. This means that we cannot understand the input-output transformations by looking only at the component elements of the system. The function of the network depends on *interactions* between and among elements; indeed, in a sense the network acts as a whole, somewhat like a flock of birds soaring and diving as a unit, or better, a full orchestra playing say, a Mahler symphony.

Because dynamical systems can have common properties, one kind of system may simulate another. It has long been known, for example, that a corresponding electrical circuit can simulate virtually any mechanical system (the correspondence is the basis of an analog computer). The possibility of simulation results from the fact that any given set of differential equations (and difference equations) can model a multitude of systems. Presumably, any behavior modeled by a set of such equations has its analogy with some other dynamical system (see Marr, 1992 for examples). With behavior, aspects of antecedents and consequences can have the role of forces in driving or retarding behavior (e.g., Nevin, 1992). Operant behavior has characteristics of a complex system in that in addition to showing a variety of non-linearities, it is dissipative and can sit on the edge between stability and instability.

Removing reinforcement, for example, is equivalent to altering the dynamics such that the basin of attraction shifts and the behavior "goes elsewhere". Consider also the complexity inherent in the action of a contingency of reinforcement. I have previously described this as:

> "The effect of reinforcement is to induce change through selection. Reinforcement effects depend on the initial states of the system, for example, where in time, or what features of responding are occurring. As this continues, the system is changing, so reinforcement acts on a different pattern, and so on. The patterns of behavior emerging and the pattern of reinforcement delivery are in a kind of dynamic dance, a flowing partnership between the effects of patterns of reinforcement on patterns of responding and the counter effects of patterns of responding

on the patterns of reinforcement" (Marr, 1997(a), p. 77).

Network models can also display the interactive pattern of input and output, ultimately (if all is well) achieving the meta-stability characteristic of contingency-controlled behaviors. With behavioral change, it is often very difficult, if not impossible to discern just what aspects of behavior, at what level, are being acted upon. This is the molar –molecular problem and may reflect the emergent aspect of a complex system. Likewise, in the network, the shifting distribution of weights and activations as the system "learns" are virtually impossible to comprehend.

What Can Neural Networks Tell Us About Behavior?

There are at least two dimensions of concern here. First, we may ask to what extent neural networks and their computational cousins have significant correspondence with actual nervous systems and thus may tell us something useful about brain-behavior relations. Second, even if such models were not enlightening about the role of the brain in behavior, then they still might be useful as analytic abstractions of behavioral processes and thus form part of the armamentaria of mathematical theorists of behavior. I would like to explore briefly this latter alternative first and then return to the question of physiological plausibility.

There is now a vast literature attesting to the success of network systems in modeling an astonishing range of activities for both theoretical and practical purposes including pattern association and recognition, signal processing, concept learning, memory processes, Pavlovian and operant conditioning, language function, motion detection, and a host of other applications in engineering and elsewhere (Amit, 1994; Anderson, 1995; Ellis & Humphreys, 1999; Haykin, 1999; Rouder, Ratcliff, & McKoon, 2000; Schmajuk, 1997). The contributions to This Issue are clear illustrations of the range of the method. Neural networks thus share in the general advantages of quantitative models. Minimally, useful quantitative models require careful specification of assumptions, relevant variables and their interrelations, computational rules, and initial and boundary conditions. These efforts serve to sharpen theoretical positions and structures that, in psychology, are usually restricted to verbal description. Good models handle, within the domain of interest, extant results in at least relative detail; they are falsifiable, thus test the limits of their applicability; they link or integrate a seemingly diverse set of phenomena, and they are capable of extending our understanding through non-trivial experimental suggestions and predictions.

There are a number of examples we can point to in the quantitative analysis of behavior that display all these desirable characteristics of good

models: behavioral momentum, melioration, global optimality, linear systems theory, varieties of behavioral economics, and delay reduction, just to name a few. How might these examples differ from a successful network model? It is in the areas of enlightening integration and prediction that network models seem the least useful or effective. I've have found very few predictions of new, not to say surprising, behavioral phenomena or principles through the use of these simulations, but I may be simply showing my ignorance (see Ellis & Humphreys for a few possible examples). One case to consider here is the model discussed by Donahoe, Palmer, and Burgos (1997) simulating basic operant conditioning. A significant aspect of their theory of operant conditioning is that stimuli present when the reinforcer is delivered gain control over responding thereby, so what is strengthened or selected is environmental control as opposed to simply responding. This theory, of course, was in no way dependent on any quantitative model of behavior. Their model simulates a variety of effects consistent with their theory and with some extant data, as it should because it was built with that purpose. One interesting feature is that without the input from a context, conditioning did not occur. Since the success of the model is evaluated on how well the simulations simulate, just what this result tells us is not clear because it is difficult to imagine behavior occurring in a total contextual vacuum. Conditioning also depended on "spontaneous activity" of units engendered by variation in the logistic activation function, in turn producing a stochastic variation in threshold activation. While this characteristic of the model was adopted for reasons of physiological plausibility, as mentioned earlier, stochastic "jiggling" is a common technique in network modeling and has the function of improving the dynamical properties of the system leading to stable attractors related to minimal error.

Biological plausibility aside for the moment, network models of behavior appear to do just what they are designed to do, namely simulate some process, or processes, *and nothing else*. They may be constructed to lend support to an existing theory (e.g., Donahoe, et al., 1997; Rouder, et al., 2000) by simulating the experimental data previously suggested by that theory. In that sense, they are implementers, not explanatory mechanisms. To be certain, mighty portions each of mathematical sophistication and skill, experience, ingenuity, patience, and fiddling are required to accomplish this kind of feat for any interesting theory. There are an astonishing variety of *types* of models from which to choose (see, e.g., Haykin, 1999), or one may construct a hybrid, or even a new type altogether (see Rouder, et al., 2000 for a recent example). Within any given type, numerous parameters have to be adjusted and rules generated specifying interactions and weights, initial input states and other unit weights, stochastic "jiggling", activation functions, output criteria, feedback conditions and distributions, etc., etc. Given all the available options, network

models can be devised for almost any purpose. Thus they are examples of methods Uttal calls "all too powerful" (1999, p.60). His reference is to Fourier methods applied to putative visual processing, but the idea is the same. The generality of networks (and associated techniques such as genetic algorithms) to simulate dynamical processes is enormous.

In the recent paper by Rouder, et al. (2000), the researchers make the following statement about their network model in their General Discussion: "The model successfully *explains* performance in three object identification tasks, quantitatively accounting for naming latencies and their distributions, accuracy rates, and probability correct in forced choice" (p. 18, italics mine). What does "explain" mean here? Presumably, the model has successfully instantiated the assumptions of their theory and shown thereby to simulate extant data. But is the network the explanatory device, or is the theory somehow programmed into the network? We would not say a computer program explained a particular electromagnetic phenomenon by solving Maxwell's equations. Maxwell's theory explained the phenomenon; the computer, properly programmed, only did the calculations. Actually, what I believe Rouder, et al. mean is that because their theory could be successful in a network, alternative theories (e.g., those requiring some "representational" process) were either unnecessary, or inadequate. This situation harks back to the controversy in cognitive psychology I alluded to at the beginning of this essay. What is at issue here are differences in overall theoretical approaches to the same problem. Some behavior analysts might be bemused by this controversy, but the question of just what neural networks can explain, if anything, remains wherever they are used (Ellis & Humphreys, 1999). My sense is that many neural network modelers do see networks as explanatory mechanisms because they believe them to possess physiological relevance (e.g., Donahoe, et al., 1997). In other words, the network is acting like a brain. To stretch a point, if the model is successful, then perhaps that's the way the brain does it.

Why are they so effective simulators, especially in the domain we call "learning"? First, they share common properties with other of dynamical processes as indicated in the first part of this essay. More particularly, behavioral control and change as we understand it through say, conditioning theory and experimentation, is subject to a dynamical systems approach as with any other process subject to change---from mechanical and electrical systems to developmental processes in biology, natural selection, population genetics, biochemical limit cycles, predator-prey interaction, etc. (see, e.g., Killeen, 1992 and Marr, 1992; 1997(a) for more detail here). Second, network models, at a minimum, capture three major (one could almost say axiomatic) aspects of behavioral change: (1) some form of selective association, (2) some form of a memory, and (3) some variation of a delta rule or feedback process that drives

the system toward fields of stable or meta-stable attractors. Given all the options for programming these aspects, no wonder these models are so powerful. Thus, any given model is, to say the least, underdetermined; perhaps an infinity of other networks with different properties might do equally well, if not better. One could counter that all models are underdetermined, as indeed they are. However, few seem as unconstrained as a network system.

But, the retort of many a serious neural network modeler of learning is that *they are constrained---by what we know of the nervous system and its role in learning.* This is a principal assertion of Donahoe, et al. (1997), and current information about neural mechanisms of conditioning clearly guided the design of their model. Generally, from the perspective of psychology most treatments of network design begin with a discussion of real neural systems, and some treatments are quite sophisticated (e.g., Anderson, 1995). Despite this deference to the brain, as one who has taught neurophysiology, I remain deeply unimpressed. As I said in a previous discussion of this issue: "...in comparison with any proposed or known *actual* neural circuits (never mind big chunks of brain with perhaps billions of cells) in the cortex, the retina, the thalamus, the hippocampus, the cerebellum, and so forth, network models are a joke. Just as Woody Allen described *War and Peace* ("It's about Russia"), network models are "about the nervous system" (Marr, 1997(b), p. 234).

On what is this assertion based? The anatomy and physiology of a single neuron is, in itself immensely complex. For example, these cells occur in a bewildering variety of sizes and shapes; in no other organ system is there anything like the variety of cells found in the nervous system (and only about 10% of the total are neurons; most of the rest are varieties of glial cells). In the primate retina alone, there are at least 80 different types of neurons (Sterling, 1998). This variety, in turn, has profound functional significance. Without going into any detail, the anatomy of the cell (including axonic and dendritic branching) affects its electrical properties that, in turn, modify or control the role of the cell in any neural circuit. The circuits themselves defy simple description (see, e.g., Shepherd, 1998 for details). For example, a single pyramidal cell in the cortex may have more than 20,000 synapses on it, each, because of its position, or extent, or particular transmitter, etc., exerts a differential influence on that postsynaptic cell. The known variety of different synaptic interactions alone is huge, for example, serial, reciprocal, rectifying, excitatory, inhibitory, axo-axonic, axo-dendritic, dendro-dendritic, electrical, chemical, ionotropic, metabotropic, and neuro-modulated, to name but a few possibilities. In myriad combinations these arrangements make up what are called *microcircuits* and are probably the major regions undergoing the changes reflecting what we call learning and memory. The possible modifications and the processes bringing these changes about are also enormous and intricate.

(The sorts of Hebbian-like, NMDA-driven mechanisms presently thought to underlie long-term potentiation, for example, are probably only a tiny subset of actual mechanisms of change).

Microcircuits are only a component of a large hierarchy of cellular arrangements. At the next level are the *local circuits* that are perhaps closer to what neural network models are attempting to model. However, real local circuits are far more complex, not only in the connections, but in the number and variation of cell types, transmitter functions, time constants, ionic and metabolic mechanisms, the relative role of passive spread through the circuit versus action-potential-driven, as well as the inputs to, and outputs from the circuit. It has been possible in certain cases (almost exclusively in a selected subset of invertebrate systems) to trace out local circuits and relate their operations to characteristic behaviors of the organism (e.g., Simmons & Young, 1999). Some local circuits have been described in vertebrate systems as well, but how a particular circuit arrangement relates to the behavior of the organism is almost totally unknown. But whether we look at invertebrate or vertebrate local circuits, one fact is clear: Given *only* the circuit, it is impossible to say what it does in the sense of what, if any, behavior relates to it. One is in the position of say, of trying to understand what Don Francisco is saying on the Spanish-language variety show "Sábado Gigante" by examining a circuit board inside the television!

Local circuits either alone or, much more commonly, in interaction with other local circuits make up what are traditionally called "centers", the next stage of the neural hierarchy. Regions of the hypothalamus, cerebellum, hippocampus, etc. are examples. These may comprise many millions of cells with astronomical numbers of connections. The function of "centers" and their interactions seem to be of major interest to most bio-psychologists and cognitive neuroscientists, as the recent ocean of research with computerized axial tomography (CAT) and functional magnetic resonance imaging (fMRI) scans attest. Perhaps neural network models might reflect the operation of "centers" as opposed to single neurons, in other words, the elements making up the architecture of the model would each represent huge numbers of cells acting as a unit. The problem with this is that we have little or no idea how interacting "centers" work in the brain either. For example, the primate visual system is one of the most intensively studied and mapped in any vertebrate. Depending on who is counting and how, some 40 or more regions ("centers"?) have been identified with more than 300 connections between them (e.g., Felleman & Van Essen, 1995). What are we to make of this? Now do we know how we see? Of course not.

The unfathomable complexity of the vertebrate nervous system implies that the reductive program of trying to understand molar behaviors from the

perspective of the operations of real neural cells, or even "centers" is probably extremely limited in its *possible* achievement. Uttal (1999) in his recent book argues this point brilliantly, and asserts that because of inherent limitations on reductive explanations of behavior based on physiology, efforts should be directed toward a "new behaviorism", a remarkable position for one who has throughout his career been a major contributor to the analysis of sensory mechanisms.

The problem of behavior-brain relations takes us back to complex non-linear dynamical systems with potential self-organizational and emergent properties. While the component functional units and their interactions are in some way responsible for the outcomes of the system as a whole, it is impossible to predict those outcomes by looking at the units themselves, in no matter how much detail. An analogy is the n-body problem in astronomy discussed earlier. With as few as three bodies, it is impossible to break the system down by looking at how one body interacts with any other. In a word, the system is irreducible, and here there are only three units! In the domain of neural network models, we could have little or no understanding of what the network was accomplishing by looking at what was happening to any given unit, or even a collection of such units. In that sense, the models probably do share properties with real neurons.

Reductionism is the virtual defining criterion of a science, but it is practiced in at least two ways. Here I am following Nagel's (1979) framework concerning the logic of reductive explanation (see also Marr, 1990 for another application). Nagel distinguishes between *homogeneous* and *heterogeneous* reduction. Homogeneous reduction is essential to any scientific enterprise and consists of delineating, defining, and organizing descriptions of some phenomena of the world into what Ernst Mach might call an "economy of thought". There is a common set of terms as well as a common set of variables entering into functional relations. Classical thermodynamics is one example; modern dynamical systems theory is another. Behavior analysis, as we commonly practice it, is also in this category. Examples include the sort of quantitative theories I mentioned above, various other perspectives on reinforcement, response differentiation, stimulus control, schedule patterns, and the like, all for the stated purpose of the prediction, control, and interpretation of behavior.

Heterogeneous reduction is best defined by a famous quote from Skinner (1950, p. 193): "...any explanation of an observed fact which appeals to events taking place somewhere else, at some other level of observation, described in different terms, and measured, if at all, in different dimensions". Accounting for behavior in terms of neural functioning is an obvious example. Just what "accounting for" means here is uncertain, but, in general, this is a

laudable goal; and without question, studying actual neural systems in conjunction with a behavior analysis has resulted in some major accomplishments in addressing brain-behavior relations. However, as mentioned before, such projects ultimately face severe limitations and insurmountable difficulties. Moreover, I do not see neural networks, or their computational system cousins, even in hybrid combinations, as enlightening us much on the vertebrate brain-behavior problem, certainly in comparison with studying *real* brains.

So, what role can network models and their like play in behavior analysis? At their best, they represent a homogeneous reductive approach in simulating interesting behaviors based upon relatively few principles. One major contribution is that they address *changes* in behavior---conditions of acquisition, extinction, and, in general, transitions. Despite paying lip service to transitional phenomena, most of the experimental efforts in the history of the analysis of behavior have focused on the steady-state (see Marr, 1992 for a discussion of this issue). Happily, this trend is much less dominant than in the past and we now see considerable interest, especially among quantitative modelers, in behavior dynamics.

I have mentioned earlier that when it comes to integration and prediction, network models seem to be deficient. The real question is are they inherently limited in this regard. Ellis and Humphreys (1999), for example, acknowledge this problem, but note a couple of exceptions of models conferring predictions and displaying "behaviors" that, at the time were unexpected, yet were subsequently borne out by actual experiments. Given such exceptions (and they truly are exceptions), then network models can potentially fulfill all the criteria noted earlier for useful quantitative models of behavior. Just what properties are essential for this to occur I cannot say, but addressing this question provides a challenge to those applying the method to behavior analysis. As indicated earlier, behavior itself manifests emergence. Two examples are the patterns of responding under various contingencies of reinforcement and the relational stimulus control exhibited as equivalence (see also Rumbaugh, Washburn, & Hillix, 1996). Given the sort of basic conditions programmed in most learning network models I mentioned earlier (e.g., association, memory, and feedback), could we demonstrate the emergence of such behaviors as stimulus equivalence? Certainly, in this example, the network should be "exposed" to no more training stimulus contingencies than a normal subject in such an experiment. The basic question here is given a particular architecture, what are the minimal network rules to yield the maximum of behaviors? Clearly, much is left for exploration.

The other articles in This Issue attest to the range of interest and application of these models and we should expect more from the behavior

analysis community as researchers become familiar with the techniques. Many students now have considerable computer expertise and sophistication, and the next generation of able students in the experimental analysis of behavior, undergraduate as well as graduate, should be exposed to a variety of computational models applied to behavior analysis, as well those from cognitive science where these kinds of models were developed and remain as basic methods. More generally, we, as behavior analysts, need to establish a climate where quantitative approaches to behavior are an inherent part of the curriculum at all levels. This is a considerable challenge because most students trained in the behavioral sciences are accordingly untrained in quantitative methods, aside perhaps for an elementary course in statistics. Indeed, many students choose fields like psychology precisely because they wish to avoid rigorous science and mathematics courses. We can do little to affect that, but, through proper exposure and instruction, we can encourage those who do have an interest and talent in applying quantitative methods to the analysis of behavior, and thus strengthen our science.

REFERENCES

Amit, D. J. (1994). *Modeling brain function*. Cambridge: Cambridge University Press.

Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.

Bak, P. (1996). *How nature works*. New York: Springer-Verlag.

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Reading, MA: Addison-Wesley.

Bozinovski, S. (1995). *Consequence driven systems*. Bitola, Macedonia: GOCMAR Publishers.

Casti, J. (1994). *Complexification*. New York: Harper-Collins.

Coveney, P., & Highfield, R. (1995). *Frontiers of complexity*. New York: Random House.

Diracu, F. & Holmes, P. (1996). *Celestial encounters*. Princeton: Princeton University Press.

Donahoe, J., Palmer, D., & Burgos, J. (1997). The S-R issue: Its status in behavior analysis and in Donahoe and Palmer's Learning and complex behavior. *Journal of the Experimental Analysis of Behavior, 67*, 193-211.

Ellis, R., & Humphreys, G. (1999). *Connectionist psychology*. Hove, East Sussex, UK: Psychology Press.

Felleman, W. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in primate visual cortex. *Cerebral Cortex, 1*, 1-47.

Haykin, S. (1999). *Neural networks* (2nd Ed.). Upper Saddle River, NJ: Prentice-Hall.

Jackson, E. A. (1989). *Perspectives of nonlinear dynamics 1*. Cambridge: Cambridge University Press.

Jackson, E. A. (1990). *Perspectives of nonlinear dynamics 2*. Cambridge: Cambridge University Press.

Kauffman, S. A. (1993). *The origins of order*. New York: Oxford University Press.

Killeen, P. R. (1992). Mechanics of the animate. *Journal of the Experimental Analysis of Behavior, 57*, 429-463.

Marr, M. J. (1990). Behavioral pharmacology: Issues of reductionism and causality. In J. E. Barrett, T. Thompson, & P. B. Dews (Eds.), *Advances in behavioral pharmacology* (Vol 7), (pp. 1-12). Hillsdale, NJ: Lawrence Erlbaum.

Marr, M. J. (1992). Behavior dynamics: One perspective. *Journal of the Experimental Analysis of Behavior, 57*, 249-266.

Marr, M. J. (1997a). The mechanics of complexity: Dynamical systems span the quick and the dead. In L. J. Hayes, & P. M. Ghezzi (Eds.), *Investigations in behavioral epistemology* (pp. 65-80). Reno, NV: Context Press.

Marr, M. J. (1997b). The eternal antithesis: A commentary on Donahoe, Palmer, and Burgos. *Journal of the Experimental Analysis of Behavior, 67*, 232-235.

Moon, F. C. (1992). *Chaotic and fractal dynamics*. New York: John Wiley & Sons.

Nagel, E. (1979). *The structure of science*. Indianapolis: Hackett Publishing Co.

Nevin, J. A. (1992). An integrative model for the study of behavioral momentum. *Journal of the Experimental Analysis of Behavior, 57*, 301-316.

Nicolis, G., & Prigogine, I. (1989). *Exploring complexity*. New York: Freeman.

Peak, D. & Frame, M. (1994). *Chaos under control*. New York: W.H. Freeman

Rouder, J. N., Ratcliff, R., & McKoon, G. (2000). A neural network model of implicit memory for object recognition. *Psychological Science, 11*, 13-19.

Rumbaugh, D. M., Washburn, D. A. & Hillix, W. A. (1996). Respondents, operants, and emergents: Toward an integrated perspective on behavior. In K. Pribram & J. King (Eds.), *Learning as a self-organizing process* (pp. 57-73). Hillsdale, NJ: Lawrence Erlbaum.

Schmajuk, N. A. (1997). *Animal learning and cognition: A neural network approach*. Cambridge: Cambridge University Press.

Shepherd, G. M. (Ed.) (1998). *The synaptic organization of the brain*. New York: Oxford University Press.

Simons, P., & Young, D. (1999). *Nerve cells and animal behavior*. Cambridge: Cambridge University Press.

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review, 57*, 193-216.

Sterling, P. (1998). Retina. In G.M. Shepherd (Ed.), *The synaptic organization of the brain* (pp. 205-253). New York: Oxford University Press.

Uttal, W. R. (1999). *Toward a new behaviorism*. Mahwah, NJ: Lawrence Erlbaum.