Evolutionary Psychology, Complex Systems, and Social Theory

Bruce MacLennan
Department of Electrical Engineering & Computer Science
University of Tennessee, Knoxville
MacLennan@utk.edu

Introduction

Stephen Turner's work on practice theory makes important progress toward constructing a consistent narrative incorporating contemporary social theory, cognitive science, and neuroscience. The purpose in this article is not primarily to criticize this work, but rather to build on it by discussing results from related disciplines that may further enrich social theory. My article is structured as follows. From evolutionary psychology, we know the importance of comparative studies in understanding human behavior, and therefore I will argue that social theory can benefit from the study of complex adaptive biological systems, such as social insect colonies. Next I will argue that practices and other social phenomena can be understood as emergent phenomena in complex systems, and therefore that complex systems theory can illuminate the ways in which social structures emerge and evolve. In particular it helps us to understand how practices can seem like "collective objects" in spite of existing only in individuals. Turner has argued for the relevance of connectionist cognitive science to social theory, and I will explore some of the consequent implications of connectionism for social theory. From there I will turn to cognitive neuroscience, which has much to offer social theory, but I will argue that there remains an enormous "explanatory gap" between the information we get from brain imaging and our theories of information representation and processing in neural networks. Theoretical as well as empirical research will be needed to close this gap. Finally, I will discuss some of the ethical implications of our gradually improving understanding of evolutionary psychology, neuroscience, and complex adaptive systems theory.

Comparative Studies of Emergent Behavior

Karl Popper said, "The main task of the theory of human knowledge is to understand it as continuous with animal knowledge; and to understand also its discontinuity—if any—from animal knowledge" (117). The same statement may be made about social theory. While there are many important differences between human social structures and processes and those of other animals, there are also many similarities and differences of degree, and therefore social theory has much to learn from nonhuman social behavior. In many cases the social structures are simpler than those of humans, and it is generally a good strategy in science to begin by trying to understand simpler systems. Furthermore, comparative studies across species are better able to distinguish between properties that are inherent in all social structures and those that are peculiar to humans.

Evolutionary psychology seeks to understand human psychology in terms of its adaptive role in our species' evolution and by comparison with the evolution of other species' behavior. Fundamentally, it is based on the scientifically manifest observation that humans are animals and are therefore evolutionarily adapted to the historical

environment in which they evolved. This is the *environment of evolutionary adaptedness* of *Homo sapiens*, which provides a background and context for understanding the phylogenetic basis of human behavior and social structure. Although our understanding is subject to the limitations of archeology, it is supplemented by studies of contemporary hunter-gatherers and of related nonhuman species. We expect, of course, the psychosocial characteristics of other primate species to be relevant to human sociology, but the social adaptations of our more distant relatives may reveal deeper structures and processes.

For example, social insects exhibit complex forms of adaptive collective behavior. Indeed it is claimed that an ant colony exhibits intelligent behavior comparable to a vertebrate animal (Camazine et al. 245). The colony as a whole can be considered an organism in the sense that its parts—the individual ants—exist for the sake of the whole, but that the whole provides a context for the behavior of the individuals. Thus the behavior of the whole and its parts are defined relative to each other.

Although the colony as a whole has characteristic forms of behavior (and, indeed, a characteristic life cycle, for older colonies behave differently to younger ones; see Johnson 81), and collective behavior may be organized by shared physical structures, such as nests and pheromone trails, the constituent behaviors themselves are executed by individual insects. Of course, in the case of social insects these behaviors are innate, whereas most human social practices are learned; therefore we should look more closely at this difference.

An important process in evolution, which is also relevant to social theory, is the *Baldwin effect,* which provides a mechanism for genetic assimilation of acquired characteristics (Baldwin; Milner 32; Turner, "Soc. Th. Cog. Neuro."). Briefly and roughly, the process operates as follows. Suppose that there is some learned behavior that has selective advantage for a population. Members of that population that have genetic characteristics that improve their ability to exhibit that behavior (e.g., to learn it more quickly or execute it more skillfully) will have an advantage relative to other members. Other things being equal, natural selection will favor these individuals, and therefore the population will evolve a genetic predisposition to acquire and exhibit this behavior; indeed, the behavior itself may come to be encoded in the genome. Thus there is a process by which advantageous learned behaviors may eventually become innate behaviors (i.e., instincts). Of course, this is balanced by a loss of behavioral flexibility and adaptability, which may be disadvantageous.

The Baldwin effect is a special case of a process that ethologists call *niche construction*, which is also relevant to social theory. Niche construction refers to a feedback process by which a population modifies its environment to its own benefit, but then adapts to the modified environment, leading to further modifications, and so forth. The result is a complex (and largely unpredictable) coevolution of the population and its environment. Human genetic predispositions for complex social organization and certain practices, and in particular the genetic basis for language, are likely the result of niche construction. Therefore comparative studies of niche construction in many species can help us to understand the evolution of the genetic foundations of human social structures.

A similar process, with a potentially important role in social theory, is *epigenesis*, which refers to modifications of gene expression that are maintained by a kind of "cell

memory" over cell divisions; it has the effect of genetic change but without alteration in the underlying DNA sequence (Bird). Epigenesis is essential in the development of multicellular organisms, for it permits embryonic stem cells to differentiate into a variety of cell types. By means of it some organisms adapt their genetic structure, in effect, to environmental conditions. Most interestingly there is evidence for *transgenerational inheritance* of some epigenetic states; that is, environmentally conditioned epigenetic features can be passed from parents to offspring, but the role of this mechanism in evolution is unclear (e.g., Jablonka and Lamb). Some epigenetic changes are a result of social interactions, and of course they affect future interactions (e.g., Fox, Hane, and Pine). Thus *social epigenetics* may be an important contributor to future social theory.

Before leaving the topic of evolutionary psychology, it is important to mention a characteristic limitation. Because of its reliance on comparisons with nonhuman species, evolutionary psychology most easily addresses externally observable behavior and is less suited to addressing subjective experience. Nevertheless, subjective experience is fundamental to human social interactions, and so it is important that behavioral approaches be supplemented by phenomenological investigation. For corresponding to phylogenetic behavioral patterns are archetypal psychological structures coordinating perception, affect, motivation, and social interaction (e.g., Jung, *Str. & Dyn.* 114–138; MacLennan; Stevens, *Arch*.). Thus our goal should be mutually consistent behavioral, phenomenological, and neural descriptions of human social processes and structures.

Complex Adaptive Systems & Emergent Properties

In *Brains/Practices/Relativism* Turner explores the shift in social theory from the view that practices are collective objects, in some way shared by the members of a culture, to the view that "practices, cultures, and so on are ensembles, with no essence, whose elements change over time, but that persist or have continuity by virtue of, and only by virtue of, the persistence of the elements themselves" (14).

This perspective raises important issues about the objectivity of collective objects and processes, which can be illuminated by the parallel shift from *essentialism* to *population thinking* that revolutionized evolutionary biology in the century after Darwin (Mayr). The essentialist view was that a species corresponds to an eternal *essence*, which represents the *ideal type* for members of the species, to which particular individuals conform more or less perfectly. From this perspective the continuity and objectivity of a species is a consequence of the atemporality of its essence, but its evolution is problematic, as is speciation (the emergence of new species).

From the population perspective, however, a species' genome is a kind of statistical average over the genotypes (genetic codes) of all the individuals living at a particular time. Therefore, on one hand, since the genome is dependent on individual genotypes, it can evolve through the birth and death of individuals in the population, and speciation can result when populations divide. On the other hand, the genome has a certain objectivity, for it makes scientific sense to talk about the human genome, the rat genome, etc. and to make objective statements about them. As a consequence, both the persistence and evolution of species are comprehensible.

Analogously, the fact that practices are ensembles does not contradict their reality and causal efficacy as collective objects. These parallel shifts from essentialism to population thinking might seem to be merely analogous, but the connections go deeper and may contribute to the development of social theory (DeVore).

Complex Adaptive Systems

In this section I will consider some insights we may obtain from looking at practices and other social phenomena as *emergent properties of complex systems*, such as studied in nonhuman social animals and even in some nonliving systems.

A *complex system* is composed of a large number of comparatively simple parts interacting with each other so that the emergent behavior of the whole is difficult to predict from the behavior of the parts. Examples of complex systems include the brain, multicellular organisms, social insect colonies, ecosystems, economies, and human societies (Anderson, Arrow & Pines; Johnson; Solé & Goodwin). Many complex systems are *adaptive*, in that they respond to their environments and alter their behavior in such a way that they can maintain or improve their function, or so that they can "survive" (that is, continue to persist as organized systems).

Complex systems manifest *emergent properties*, which cannot be explained in terms of simple, linear interactions among the system's components. For example the foraging trails constructed by an ant colony and the characteristic nests constructed by particular species of wasps or termites are emergent properties of these collectives. Emergent properties may be characterized by *order parameters*, which measure or describe the collective behavior or structure of the system. Often these parameters are statistical in character, but none-the-less objective. For example, *reaction-diffusion equations*, which describe interactions between microscopic elements, such as cells and diffusible molecules, generate patterns similar to animal hair coats and to skin pigmentation patterns. Over time these produce macroscopic stripes and spots of predictable dimensions, although the specifics of a pattern (e.g., the color of a particular small patch) depend on unpredictable microscopic processes (Solé and Goodwin 85–8).

From the perspective of social theory, it is important that these emergent properties and order parameters are objective characteristics of the whole, despite being an effect of interactions among the parts, for social phenomena, such as practices, world-views, and languages, are similarly objective properties of human populations, despite being derivative of individual behavior, learning, and cognition. The order parameters should be predictable by social theory, even if the particulars are not.

One characteristic typical of complex adaptive systems is *circular causality*, or the *macro-micro feedback loop* (Solé and Goodwin 150), which refers to the fact that the large-scale order of the system is created by interaction of its parts, but that the interaction of the parts is governed in turn by the large-scale order. For example, the collective behavior of the ants in a colony creates pheromone trails to food sources, to which individual ants respond, maintaining and adapting the trails, as food sources are discovered or exhausted. On the one hand, of course, the higher order structures exist only as ongoing macroscale phenomena sustained by the actions of individual ants, but on the other they are causally efficacious in that they provide external resistances to

which the ants respond. In an important sense, therefore, they are *real*. Thus it seems that studying circular causality, and the relation between the collective and the individual, in relatively simple complex adaptive systems (such as social insects, bacteria, and even nonliving complex systems) will illuminate the relation of individual people to the macroscopic structures, such as practices, that they collective create.

Coarse Coding

Coarse coding is a process that occurs in complex information-processing systems such as the brain (Rumelhart, McClelland, et al. 91–6; Sanger). It refers to the fact that individual neurons might be quite broadly tuned to stimuli (to frequencies of sound, for example), but that the collective activity of a large group of neurons can represent a stimulus very precisely. That is, the collective response of a population of coarse-coding neurons results in a very fine (precise) representation. This seems to be a fundamental principle of representation in the nervous system, necessary to get precise behavior from neurons, which—from a computational perspective—are very low precision computing devices. Of course, there is nothing mysterious about how it works; it is simply statistics (the law of large numbers) in action.

Similar processes occur in the social systems of humans and other animals. For example, ants encounter other ants and by an exchange of chemical signals form independent estimates of the tasks that need to be done and the number of workers assigned to them (Johnson 74). Because each ant's estimate is based on a very limited number of samples, it is quite inaccurate, but all together the workers have an accurate estimate, and so as each worker decides individually what task to perform, the overall allocation of workers to tasks in the colony is nearly optimal. Similarly in human markets, for example, collective knowledge and intelligence may be much greater than that of any of its participants.

More generally, a collection of individuals can collectively (and distributively) represent an abstraction more accurately than any of the individuals can on its own. Thus, information, skills, practices, etc. may be imperfectly learned and performed by each of the members of a human population, but their collective behavior may appear to be an ideal shared competence, which the individuals have imperfectly acquired (language is a good example). In fact, individual performance is not based on a shared collective object, but the collective object is primarily an emergent property of individual performance. Nevertheless, by circular causality, the emergent collective object exerts a real causal influence on individual behavior. That is, the collective object is simultaneously emergent (and in that sense *descriptive*) and regulating (and in that sense *normative*). By studying these complex, self-organizing processes in simpler cases, such as nonliving systems and simple social organisms, we may gain insight into the more complicated social structures of human populations.

Amplification of Random Fluctuations

Another process that is typical in complex systems, with implications for social theory, is *amplification of random fluctuations*, which results from positive feedback within the system. As a consequence, relatively minor fluctuations can direct complex systems into divergent evolutionary pathways. Therefore the origins of some features of

social systems will be rooted in historical contingency and have no more general explanation (cf. Turner, *Brains/Prac./Rel*. 102–5), although the persistence of these features can be explained by identifying the feedback processes. Sometimes this amplification is useful (adaptive), for it can break symmetries, or balanced forces, that may be blocking further evolution (Buridan's ass is a proverbial example). On the other hand, by stabilizing and overly reinforcing one set of structures and processes, it may effectively block access to others that might be preferable. Understanding the feedback processes from a complex-systems perspective may reveal means for weakening or eliminating them, to permit the formation of new structures, if that is desirable.

We know from complex systems that there can be "phase changes": pervasive and rapid reorganization of a system resulting from a relatively minor change in conditions, either from within the system (e.g., amplification of random fluctuations, "tipping points") or from outside of it (e.g., environmental changes). Therefore, complex systems theory can help us to understand the dynamics of social change, especially large-scale, rapid reorganizations (e.g., political, conceptual, and scientific revolutions; paradigm shifts).

Blind Variation and Selective Retention

Amplification of random fluctuations provides a mechanism for divergence that presupposes no collective "choice," "act of commitment, faith, will," etc. (Turner, *Brains/Prac./Rel*. 102–5). This may be a source of simple diversity or of genuine novelty. In a discussion of evolutionary epistemology, Campbell defines evolution in a general sense as *blind variation and selective retention* (Campbell "Bl. Var." 91–3, "Ev. Epist." 56–7). The concept of *blind variation*, which is not synonymous with *random variation*, is important. It may be defined as variation that is not aimed at some goal, such as fitness, selective advantage, or at some idea of progress or optimality. Therefore it includes random variation (such as random genetic mutation) as a special case. In a social context it includes individual variations resulting from misunderstanding, limitations of learning and experience, contextual understanding, the contingencies of an individual's life, mistakes in communication or action, etc.

In the context of social systems, *selective retention* refers to any process that tends to amplify certain variations and to dampen others. A priori, there is no reason to suppose that such amplification will lead to improvements in the system or to "progress," as judged either by members of the population or by those outside of it; negative or neutral variations can be amplified as well, depending on the feedback mechanisms. Nevertheless, systems that have persisted for a long time often exhibit *adaptive* selection processes, which promote their continued existence ("survival") as systems. An important aspect of social systems is that the feedback processes are not immutable, but can be altered intentionally or by blind variation and (higher order) selective retention. In any case, the study of complex adaptive systems is relevant to understanding the evolution of social systems.

Self-organizing systems in nature illustrate the positive value of error, uncertainty, noise, individual variability, and other sources of blind variation. Ant foraging provides an informative example. As foragers wander about they discover food sources, and when they return to their nests they lay pheromone trails reflecting the quantity and the quality

of the food. Other foragers follow the trail and reinforce it when they return to the nest; this is the amplification of random fluctuations, since the initial discovery of food sources is largely a matter of chance. Thus the creation of foraging trails is an example of blind variation and selective retention.

If this process worked perfectly, the ants would devote all their foraging to the food sources initially discovered and would not discover alternative, possibly superior food sources until those found initially were exhausted. Fortunately, the mechanism does not operate perfectly; ants sometimes deviate from the path and have to wander around until they find it again. In the process they may discover a new food source, and the positive feedback processes will cause the emergence of a foraging trail to it. If the new food source is superior, the trail will become stronger and capture foraging resources from the other trails. Thus the imperfections in the trail-following mechanism cause a certain amount of unbiased exploration, which facilitates the reorganization of macroscopic structures (the foraging trails) from a less advantageous state to a better one.

A general characteristic of adaptive self-organizing systems is the productive use of error, uncertainty, noise, imprecision, error, variability, and other sources of blind variation. This goes against our habits in engineering and many other activities, in which we attempt to eliminate or control these unpredictable factors, but we can see that too perfect a mechanism can lock a complex system into a local optimum and block further adaptation. Indeed, if people were more like computers and could be programmed with identical, precise rules and could follow them with precision, human society would be much less flexible and adaptable than it is. Thus variation among individuals in the acquisition of practices (Turner, "Pr. Th., Cog. Sci., Eth."), contributes to the adaptability of human societies.

Blind variation is also applied in connectionist theories of cognition, such as *harmony theory* (Rumelhart, McClelland, et al. ch. 6). The ideas are easiest to understand in the context of a simpler process, *simulated annealing*, which is used to solve optimization problems by controlling the degree of blind variation, which is measured by a parameter usually called *computational temperature*, by analogy with thermodynamics, in which temperature measures the amount of random motion (Kirkpatrick, Gelatt, and Vecchi). Simulated annealing attempts to improve the state of a system, as measured by some *figure of merit*, by exploring the effect of small perturbations of the state. If the computational temperature is low, then the potential change of state is accepted if it raises the figure of merit and is rejected otherwise. Therefore, at low temperatures, the algorithm makes incremental improvements to the system state. We may say it climbs the "merit landscape," but this behavior runs the risk of becoming trapped in a local optimum (i.e., stuck on the top of a "hill" that is not the highest hill). At "absolute zero" the algorithm is completely deterministic in its hill-climbing behavior. However, higher computational temperatures introduce more randomness into its behavior; at higher temperatures it will sometimes accept a state change even if it decreases the figure of merit. That is, the higher the temperature, the more often locally "bad" decisions will be made. This might seem counterproductive, but it provides an escape route from a local optimum, for there is a non-zero probability that the state will creep down from its hill and find its way to the slope of a higher peak.

The key to simulated annealing, from which it gets its name, is to control the computational temperature, starting at a comparatively high value and slowly (typically in stages) reducing it to zero. Therefore, in the early stages of the processes, the state is varied relatively blindly, so that it conducts unbiased sampling of the space of system states. As the temperature is decreased, the algorithm begins to prefer state changes that increase the merit; the search is biased toward regions where the merit is relatively high, although there is still random exploration. In the later stages the search becomes more directed, eventually approximating deterministic hill climbing, but by then it is likely to be on the slopes of the global optimum. Thus, the decreasing temperature shifts the search priority from *exploration* (gathering information about the state space) to *exploitation* (use of the information). It can be proved that, in rough terms, simulated annealing will almost surely find the global optimum (given a good annealing schedule and enough time). In connectionism similar models are used to account, for example, for a neural system's ability to arrive at a good interpretation of perceptual data subject to context, expectations, etc. (Rumelhart, McClelland, et al. chs. 6, 7).

I am not claiming that something akin to simulated annealing takes place in social systems (for which the notion of a figure of merit is problematic, to say the least). Nevertheless, simulated annealing is a suggestive metaphor, which illustrates the tradeoffs between local changes directed toward some global notion of improvement (changes promoting exploitation), and those that are blind to it (promoting exploration), and especially the consequences of different degrees of blind variation. The simulated annealing process suggests that complex adaptive systems can achieve global optima by starting with a high degree of blind variation and gradually shifting to changes that lead to global improvement.

This argument assumes that the "merit landscape" is constant in shape, as is often true in optimization problems, but which might not be the case in adaptive social systems. If the landscape has changed shape, so that the previous global optimum is no longer the highest peak, then it may be worthwhile to repeat the annealing process by increasing the computational temperature to destabilize ("melt") the existing structures and to allow new, better ones to emerge. Some adaptive systems can detect that they are no longer in an optimal state, due to changed circumstances, and raise the computational temperature (the probability of blind variation) in order to allow a new optimal state to emerge. For example, some organisms respond to environmental stress by epigenetically increasing their genetic variability (e.g., Hernday et al.). The analogies to scientific revolutions and other paradigm shifts are obvious, as are those to other kinds of large-scale social reorganization, but determining whether the similarities are more than superficial will require further investigation.

Of Ants and Humans

Humans are not ants, and so one may reasonably doubt that the social organization of ants and other simple life forms can tell us much about human society. Therefore it is interesting that studies of complex systems have shown that many emergent properties do not depend on the specific structure or behavioral sophistication of the agents constituting the system (Johnson 98–100). A well-known example is traffic flow, which can be described quite accurately without attributing much intelligence to the agents (perhaps

that is not surprising) and without detailed information about individual agents' beliefs, desires, goals, plans, histories, knowledge, etc. In effect, all of this detailed information, which is so important in our personal and interpersonal lives, is averaged out, and can be treated as random noise for the purposes of describing macroscopic properties of traffic (random noise which can, for sure, be amplified into macroscopic behavior). Therefore, ants, wasps, and other relatively simple organisms are relevant to human social theory, especially for comparative purposes. If we observe similar emergent behavior in human and insect societies, then that similarity suggests that the behavior does not depend on the specifics of individual human behavior, which implies that we may be able to discover the causes of some human social structures by studying much simpler complex systems.

Connectionism

Turner discusses the importance to social theory of cognitive science, and in particular of *connectionism*, a new approach to artificial intelligence and cognitive science based on mathematical models of information processing and learning in the brain. So it may be worthwhile to make a few observations about connectionism. Traditionally—and by "traditionally" I mean stretching back in Western intellectual history at least 2300 years to the time of Aristotle—traditionally, knowledge has been understood as a system of language-like propositions, and thinking as a kind of calculation directed by language-like rules. The paradigm is formal deductive logic, especially as systematized in modern symbolic logic. Traditional Artificial Intelligence (AI), often called "symbolic AI" is based on the same assumption, that cognition is a matter of symbol manipulation. Unfortunately, by the early 1980s it had become apparent that symbolic AI, especially as implemented in the programs called *expert systems*, could not achieve levels of performance comparable to human experts.

About this time an old idea was resurrected: that intelligent computers could be based on the same principles by which the brain operates. This approach is called *artificial neural networks* or *connectionism*, because the knowledge is implicit in the connections between the neurons rather than explicit in language-like structures. Knowledge representation is implicit in that each fact is distributed over a large number of connections and each connection participates in the representation of a large number of facts. Furthermore, while symbolic AI systems are programmed by inputting a set of rules or facts expressed in some language, and learning modifies these language-like structures, connectionist networks are trained by giving them examples from which they generalize. Therefore connectionism is better than symbolic AI at dealing with tacit knowledge, which is typically difficult or impossible to put into verbal form.

An objection frequently made against connectionism, especially in its early days, was that many cognitive processes are apparently governed by rules. Language use is the classic example, since it appears to be governed by the syntactic rules (grammar) of the language. The observation that the grammatical rules of natural languages seem to be much too complicated to be learned from the samples to which an infant is exposed (the problem of "the poverty of the stimulus") has led to the conclusion—most commonly associated with Chomsky and his followers—that the human brain must contain a sophisticated "language module" incorporating innate knowledge of certain universal grammar rules common to all natural languages.

Like language, many other practices appear to be governed by rules, and if the hypothesized rules seem to be more complicated or extensive than could be learned in the time available (either explicitly as verbal rules or—more likely—tacitly from the background), then we might be led to hypothesize some innate knowledge of the rules. (The Baldwin effect could provide a theoretical basis for this innate competence.)

Connectionism, however, provides an alternative explanation based on a distinction between *rule-following behavior* and *rule-like behavior* (literally, *regular* behavior). In rule-following behavior, some agent (e.g., a person or a machine) executes a process by following rules that are explicitly and literally expressed in some language (natural or artificial). This is the idea of an *algorithm*, which is fundamental to computer science, but had its origins in human rule-following: the arithmetical algorithms of hand calculation. When consciously executed by humans this is a slow process (because it is fundamentally unnatural), but when people get some practice it becomes less conscious and more automatic. Nevertheless, it has been argued that rules are still being followed, but they have been internalized ("compiled" in computer jargon) and are being executed at a deeper level inaccessible to consciousness.

Connectionist research has shown, however, that an artificial neural network system can exhibit rule-like behavior—that is, *appear* to be following rules—without actually doing so. There are no explicit rules in a connectionist network; nevertheless, the cumulative effect of the many connections and the individual neural computations can exhibit regular behavior that an external observer can describe as following rules (at least to a first approximation, and that is important!). The *locus classicus* is McClelland and Rumelhart's experiment in which a neural net learned to form the past tenses of English verbs (McClelland, Rumelhart, et al. ch. 18). Although this was a very simple model, and can be criticized on a number of fronts (cognitive science, neuroscience, etc.), it nevertheless reveals a different explanation for rule-like behavior. At first the network learns each verb as a special case, learning it in effect by rote; with exposure to only a few, common verbs, every verb is in effect an irregular verb. After it has been exposed to a sufficient number of verbs, however, it *apparently* learns the rule "add -ed," since it over-generalizes and begins applying this rule to irregular verbs to which it doesn't apply (and which it had previously handled correctly). Of course, the network has not learned an explicit rule at all (there is nowhere in a connectionist net to store a rule), but it has adjusted its behavior to act *as though* it were following this rule. Significantly, with continued exposure to both regular and irregular verbs, the network learns the past tenses of both kinds of verbs correctly (effectively learning the irregular verbs as exceptions to the "rule"). Like human English speakers, the net is even able to make good guesses about the past tenses of verbs it hasn't learned, thus demonstrating some internalized inferred knowledge of the phonetic structure of English. The parallels to human language acquisition are, of course, very suggestive (perhaps overly so), but the key point is that a connectionist network can behave *as though* it is obeying rules, and even exceptions to the rules. Many other connectionist experiments tell the same tale.

It is important to understand that a connectionist network is not simply an alternative method of getting the same results as following a set of rules, for connectionist systems are potentially much more flexible than rule-based systems. This is because the response of a connectionist network can be the result of combining many factors, some quite subtle

and meaningful only in the context of other factors. Therefore, connectionist models can exhibit the flexibility and context sensitivity of response characteristic of humans and other animals. Rule-based models, in contrast, are by nature "brittle" (i.e., broken by minor exceptions, context dependencies, etc.) and can achieve flexibility only by having exceptions to rules (expressed, again, in rules or other language-like structures), exceptions to the exceptions, and so forth. Therefore, rule-based attempts to account for the flexibility and fluency of human and nonhuman skillful behavior pile epicycles upon epicycles, and the resulting models are approximate, complicated, implausible, and unnecessarily so, for connectionism explains fluency much more directly. Indeed, our experience with rule following by both humans and computers is that it becomes less efficient with an increase in the number and complexity of the rules, and therefore it is difficult to explain how rule following could produce skillful behavior.

It is sometimes argued that even if much of our knowledge (especially of skills and practices) is not stored in our brains in the form of rules, nevertheless that is our only means for expressing knowledge in tangible form, so that it can be a subject for scientific discourse. That is, it is claimed that cognition should be described *as if* it were following rules, even if it is not in fact doing so. Aside from arguments based on the necessity of a non-verbalizable background and tacit knowledge to provide a context for verbalizable knowledge (Polanyi; Searle 172–4; Turner, *Brains/Prac./Rel*. ch. 1), connectionist research also suggests that often rules can, at best, approximate the flexible behavior of neural networks. (There are formal, historical, and metaphorical connections to the approximation of irrational real numbers by rational numbers.) Therefore, connectionism implies that practices cannot in general be captured by explicit rules.

There is a method for analyzing the matrix of connections between one group of connectionist neurons and another; it is called the *singular value decomposition* of the matrix. Without going into details (but see Appendix), I would like to say a few words about this analysis, because it illustrates the relation between connectionist information processing and rule-directed information processing. Using this method one can extract from the interconnection matrix a series of rule-like relationships of decreasing strength or importance. Each of these implicit rules looks for a pattern in the input, and to the extent it finds it, it generates a characteristic output pattern. If the input matches several of these implicit rules, it will generate a composite output, which is equal to the average of the outputs of the activated rules, weighted both by how well each rule matches the input and by the rule's inherent strength. Thus, in typical situations, not just one rule is selected, but a subtle blending of all the rules, which permits context sensitive and flexible rule application.

Again, I must emphasize that the rules are not literally there, but we can analyze the effect of the connection matrix in terms of a weighted blending of these implicit rules. Furthermore, a large interconnection matrix, which is what we realistically expect in the brain, may require a large number of implicit rules—perhaps many thousands—in order to be completely captured. The rules can be listed in order of decreasing strength (importance), and if we cut off this list before the end, we will get an approximation of the information processing performed by the connection matrix, but some subtlety and sensitivity will be lost. This is one way to understand what happens when we try to capture expert behavior in a small set of rules: we get a crude approximation. In some

cases, a few strong rules will capture most of the behavior of the neural net; in these cases, rule-guided information processing will work fairly well. In other cases, however, the competent behavior of the neural network will have the effect of the combined actions of a large number of weak rules. In these cases, any small number of rules will give crude, inflexible behavior compared to that of the neural network.

The differences between connectionist and rule-directed information processing and control are reinforced by the Dreyfus brothers' research on expert behavior. In brief, they found that while novice behavior may be characterized by the mechanical (that is to say, mindless) application of context-free rules, as people become more skilled, they abandon the strict use of rules and behave in ways that are flexible and sensitive to context. Therefore, as expected, while novice behavior can be captured in small sets of rules, the skillful behavior of experts can only be approximated by increasingly large sets. Experts respond to many subtle factors, which they integrate in a context-sensitive way, to arrive at judgments or otherwise to govern their behavior.

Unfortunately, connectionist research often gives a false impression. Many connectionist experiments involve training a simple network to perform some isolated cognitive task (e.g., learning past tenses of English verbs). In order to demonstrate that nothing has been "preloaded" into the system (in particular, none of the experimenters' knowledge), a simple unstructured net is often used, with unstructured connections and random connection strengths. However the human brain is not like this! It has an enormous number of modules, each with a highly specific structure, and interconnected in very specific ways. Connectionist researchers understand this, but it is difficult to model the brain on this level. Nevertheless, connectionist experiments can give the impression that connectionists believe that the brain begins as an amorphous, unstructured mass of randomly connected neurons, tabula rasa. (As a consequence, connectionism is sometimes misunderstood as a modern variant of simple associationism.) Rather, the brain has an elaborate species-specific structure, which develops in interaction with the environment during an individual's life (especially through young adulthood). Therefore, in bringing neuroscience to bear on social science, we must be cognizant of the species-specific brain structures and processes shared by all humans.

Cognitive Neuroscience

Turner observes that brain-imaging technology provides a new tool for social theory, permitting us to see, for example, "what parts of the brain itself are activated in various 'moral' situations, such as the punishment of free-riders" ("Cog. Sci., Soc. Th. & Ethics," see also "Soc. Th. Cog. Neuro."). These developments are certainly exciting, but I think it is important to be modest in our expectations of what we may learn from brain imaging studies at this time. The current resolution of fMRI and similar imaging technologies is on the order of several square millimeters ("Func. Mag. Res. Img."), which seems quite good, but we must remember that there are at least about 146 000 neurons per square millimeter in human cortex (Changeux 51). Therefore each pixel in one of these images represents the average activity of at least about 150 thousand neurons, which, by comparison, is more than the number of transistors in an early-1980s Intel 80286

computer (used in the IBM PC/AT) ("Intel 80286"); that is, each pixel represents the average activity of an entire PC. This implies there is a lot going on that we cannot see.

Furthermore, fMRI measures blood oxygen level, which reflects metabolic activity in neurons over a duration of several seconds. Since electrochemical activity in neurons takes place on a timescale of milliseconds, the time resolution of fMRI and similar imaging techniques is quite coarse. Within the time interval resolvable by the device, each of these hundreds of thousands of neurons may have fired hundreds or thousands of times; we see the average behavior. Therefore, valuable as these techniques are, they do not come close to telling the whole story.

As further evidence of the importance of sub-millimeter cortical structure, I will mention *computational maps*, which are ubiquitous in the brain and one of the fundamental means for neural information representation (Knudsen et al.). In a computational map, properties of a stimulus, motor plan, etc., are systematically mapped to locations in a patch of cortex. For example, auditory stimuli are systematically represented according to pitch; visual stimuli are mapped according to retinal location, edge orientation, etc.; and the intended destination of an arm movement may be represented in a map corresponding to "reach space." Some of these computational maps are less that a square millimeter in size (Knudsen et al.), and therefore they are below the level of fMRI resolution. Computational maps are critical in sensory and motor neural systems (Morasso and Sanguineti), on which many practices depend, and so the eventual integration of social theory with cognitive neuroscience will depend on an understanding of computational maps and other sub-millimeter neural structures.

Thus, while brain imaging contributes to our understanding of the medium- to large-scale organization of brain activity over relatively slow timescales, this needs to be supplemented by improved understanding of processes taking place in dense networks of hundreds of thousands of neurons on millisecond timescales. In addition to invasive investigations of neural activity, cognitive neuroscience will depend on a deep theory of information representation, processing, and control at the neural level, such as is being developed in connectionist neural network theory. Animal studies are crucial to the development of this theory.

Turner points to the relevance of *mirror neurons* to a neurocognitively grounded social theory ("Pr. Then & Now," "Soc. Th. Cog. Neuro."). By allowing, in effect, the neural activity of one organism to be reflected in the neural activity of another, mirror neurons implement a more direct and efficient means of learning and communication than do conscious observation and imitation. In particular the mirror neuron system may be important in establishing empathy, discerning intentions, and fostering the learning and evolution of language (Arbib and Rizzolatti). Thus, beyond ordinary perception, mirror neurons provide an additional, and much more subtle and efficient, channel for coordinating the behavior of the members of a population (human or nonhuman). Therefore the mirror neuron system is an additional mechanism in the emergence of macroscopic properties and processes, which establishes a closer link between neuroscience and social theory.

Ethics and Human Nature

It remains to say a few words on ethics and, in particular, on relativism. In *Brains/Practices/Relativism* (ch. 4) Turner analyzes the concept of relativism as an explanation of observed differences in cultures, mores, worldviews, standpoints, and so forth. As an explanation it rests on the *premises model*, which accounts for these differences in terms of deductive conclusions deriving from differing *fundamental tacit premises*, which arise from necessarily non-rational acts of commitment, leaps of faith, jumps between worldviews, etc. He argues that the premises model is implausible and unnecessary, and that differences arise from historical contingency, for cultural paths may diverge due to the amplification of accidental differences or of peculiarities of environment, context, individual behavior, learned practices, etc. (see also his "Practice Relativism").

Similarly, the study of complex adaptive systems shows that such systems may have alternative stable states (stabilized by feedback processes) that are equally adaptive. And, as we know, radically different cultures can be equally effective in ensuring the survival and well-being of their members. Which culture arises in a particular situation may depend on historical contingencies, but can also be an effect of amplification of random fluctuations with no adaptive significance. This observation may seem to imply some form of relativism, but it is important to keep in mind that adaptation to the environment is an empirical concept with a basis in evolutionary psychology. The study of complex adaptive systems will help us to understand the dynamical relations that promote the persistence of social organizations in time.

Complex systems theory is especially informative in dealing with systems composed of simple inanimate elements or of animate agents with simple behavior conditioned by their local situations; therefore it might not seem to be useful when the agents are human beings. Its applicability lies in the fact the people often respond unconsciously to social situations (that is, without conscious, explicit judgment), and that when they do make explicit decisions, they are largely based on personal circumstances (local criteria), not global (system-wide) conditions (e.g., driving in traffic, making individual economic decisions). Therefore, to the extent that people behave unconsciously and locally, complex systems theory can account for emergent social structures and processes. Conversely, conscious global judgment (that is, explicit decision making informed by knowledge of the whole system) can result in distinctively human social behavior. Further, since amplification of microdecisions can lead to global phase changes, complex systems theory illustrates the importance of explicit ethical theorizing and individual ethical choice to the self-organization and evolution of social structures.

I suppose it will be granted that any practical ethics—by which I mean any ethics that will promote the long-term survival and well-being of humankind—must take into account the biological nature of *Homo sapiens*, as revealed by evolutionary psychology and neuroscience, and its relation to the rest of the living and nonliving world. This is the raw material with which ethics must work. Ethics that ignores human nature is as futile as carpentry that ignores the nature of wood.

In particular, it is important to understand the *environment of evolutionary adaptedness* (EEA) of our species. Evolutionary biologists use this term to refer to the

environment in which a species has evolved and to which it has become adapted through natural selection. It is by reference to a species' EEA that we may understand its specific adaptations and their functions. Anthony Stevens argues that 99.5% of the evolutionary history of modern humans (*H. sapiens sapiens*) has been spent as hunter-gatherers living in small kinship groups, and that this is the explanation for the "cultural universals" enumerated by George Murdock, Robin Fox, Donald Brown and others (Stevens, *Arch.* 25, *Two Mill.* 15–19, 64–8). Leaving aside the specifics of these claims, which will stand or fall under the investigations of archaeologists, anthropologists, and evolutionary psychologists, it is nevertheless true that understanding our EEA will help us understand human nature and its adaptive function. How, then, can we apply some of the insights that we can obtain from neuroscience, evolutionary psychology, and complex systems theory to our ethical dilemmas?

Although evolutionary psychology is a young science, it is reasonable to suppose that *Homo sapiens* is adapted to living by hunting and gathering in kinship groups a few dozens in size. It is also plausible that many of the "discontents of civilization" and even many neuroses have their root in the divergence of civilized life from our environment of evolutionary adaptedness (Stevens, *Arch*. ch. 9, *Two Mill*. 86). This presents us with a dilemma, for we can neither return to this ancient lifestyle (which would require a sort of regression to an earlier, less consciously organized social structure), nor can we (yet!) alter our genome, which is the foundation on which human nature is built.

We require a *tertium quid*, and it can be found in our ability to bring conscious understanding to this ethical dilemma (Stevens, *Arch*. 276–7). Indeed, conscious understanding and explicit discourse about our behavior is an aspect of human nature, which can follow a self-reinforcing trajectory toward an increasingly conscious and reflective awareness of ourselves, now incorporating insights from connectionist cognitive science, neuroscience, evolutionary psychology, and complex systems theory. In this way we can strive to formulate ethical norms that are compatible with human biology, and that promote the well-being of the human organism (including our psychological well-being, which also has its biological foundation).

This widened consciousness is a valuable goal for both the individual and society. First, a better understanding of our nature, both phylogenetic and ontogenetic, will facilitate our individual well-being. "Know thyself," as the Delphic maxim urges. The solution is to neither repress our biological nature nor to act it out, but to engage in an informed, conscious negotiation with it. Second, a society will be better adapted to its environment if it strives consciously to organize itself consistently with human (and nonhuman) nature.

The foregoing remarks have treated biological human nature as a given, a necessary precondition for any ethics that can be adaptive in the long term. Nevertheless, even without germ-line genetic engineering and neurosurgery, human nature *can* be altered. First, we now know that the brain is much more plastic than previously believed, and that the environment can have important effects on neural processes not only in early childhood, but throughout the human life cycle. Second, epigenetic processes permit the expression of genes to be altered by environmental conditions, effectively altering genetics without a change in the DNA. Finally, as we have seen, the Baldwin Effect

eventually adapts the genome to the niche (e.g., behavioral norms and other social structures) that a population has created for itself.

Yet we must be modest, for there is much that we still do not know, and not base radical changes of direction on premature, inadequately tested, and poorly understood scientific theories. (Social Darwinism illustrates the dangers.) Furthermore, one of the most important lessons of complex systems theory it that it is difficult to predict the effects of changes, due to the complexity of the feedback processes and their amplification of minor influences.

Conclusions

Jung, who was well aware of the instinctual side of humankind; said (*Alch. St.* 184), "Nature *must not* win the game, but she *cannot* lose. And whenever the conscious mind clings to hard and fast concepts and gets caught in its own rules and regulations—as is unavoidable and of the essence of civilized consciousness—nature pops up with her inescapable demands."

This statement is quite precise. "Nature *must not* win"; that is, it is imperative that we not give in to nature, for that would be a regression to an uncivilized state and in fact a unnatural rejection of the human potential for civilization. On the other hand, nature "*cannot* lose"; that is, it is impossible to escape nature because we are part of it and constrained by natural law. Further, human nature permits, and even demands, the formation of concepts and the conscious formulation of behavioral norms, which are fundamental to civilization and have their own self-reinforcing dynamics, but are ultimately constrained by biological necessity.

How can this paradox be resolved? I believe we have to hold the tension of the opposites: on one hand, the phylogenetic or species-specific nature of the human mind, which defines the raw materials we have to work with, and on the other, our equally human ability to consciously and critically understand and govern our perception and behavior—in the context of human nature—and thereby to make our individual contributions to the evolution of our society.

Works Cited

Anderson, Philip W., Kenneth J. Arrow, and David Pines, eds. *The Economy as an Evolving Complex System*. Reading: Addison-Wesley, 1988.

Arbib, Michael A., and Giacomo Rizzolatti. "Neural Expectations: A Possible Evolutionary Path from Manual Skills to Language." *Communication and Cognition* 29 (1997): 393–423.

Baldwin, James Mark. "A New Factor in Evolution." *The American Naturalist* 30 (1896): 441–51, 536–553.

Bird, Adrian. "Perceptions of Epigenetics." *Nature* 447 (2007): 396–8.

Brown, Donald E. *Human Universals*. New York: McGraw-Hill, 1991.

Camazine, Scott, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz, and Eric Bonabeau. *Self-Organization in Biological Systems*. Princeton: Princeton UP, 2001.

Campbell, Donald T. "Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes." *Evolutionary Epistemology, Theory of Rationality, and the Sociology of Knowledge*. Ed. Gerard Radnitzky and W. W. Bartley III. La Salle: Open Court, 1987. 91–114.

Campbell, Donald T. "Evolutionary Epistemology." *Evolutionary Epistemology, Theory of Rationality, and the Sociology of Knowledge*. Ed. Gerard Radnitzky and W. W. Bartley III. La Salle: Open Court, 1987. 47–89.

Changeux, Jean-Pierre. *Neuronal Man: The Biology of Mind*. Trans. L. Garey. Oxford: Oxford UP, 1985.

DeVore, Irven. "Prospects for a Synthesis in the Human Behavioral Sciences." *Emerging Syntheses in Science*. Ed. David Pines. Redwood City: Addison-Wesley. 53–65.

Dreyfus, Hubert L., and Stuart E. Dreyfus. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free P, 1986.

Fox, Nathan A., Amie A. Hane, and Daniel S. Pine. "Plasticity for Affective Neurocircuitry: How the Environment Affects Gene Expression." *Current Directions in Psychological Science* 16.1 (2007): 1–5.

Fox, Robin. *Encounter with Anthropology*. London: Peregrine, 1975.

"Functional Magnetic Resonance Imaging." *Wikipedia: The Free Encyclopedia*. 15 June 2007. 2 July 2007 <http://en.wikipedia.org/wiki/Fmri>.

Hernday, Aaron, Margareta Krabbe, Bruce Braaten, and David Low. "Self-Perpetuating Epigenetic Pili Switches in Bacteria." *Proceedings of the National Academy of Sciences of the United States of America* 99, Suppl. 4: Sackler Colloquium on Self-Perpetuating Structural States in Biology, Disease, and Genetics (Dec. 10, 2002): 16470–6.

"Intel 80286." *Wikipedia: The Free Encyclopedia*. 11 June 2007. 2 July 2007 <http://en.wikipedia.org/wiki/Intel_80286>.

Jablonka, Eva, and Marion J. Lamb. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge: MIT P, 2005.

Johnson, Steven. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York: Scribner, 2001.

Jung, Carl G. *Alchemical Studies (CW 13)*. Trans. R. F. C. Hull. Princeton: Princeton UP, 1967.

Jung, Carl G. *The Structure and Dynamics of the Psyche (CW 8)*, 2nd ed. Trans. R. F. C. Hull. Princeton: Princeton UP, 1960.

Kirkpatrick, Scott, C. D. Gelatt Jr., and Mario P. Vecchi. "Optimization by Simulated Annealing." *Science* 220 (1983): 671–80.

Knudsen, Eric I., Sascha du Lac, and Steven D. Esterly. "Computational Maps in the Brain." *Annual Review of Neuroscience* 10 (1987): 41–65.

MacLennan, Bruce J. "Evolutionary Jungian Psychology." *Psychological Perspectives* 49.1 (2006): 9–28.

McClelland, James L., David E. Rumelhart, and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models*. Cambridge: MIT P, 1986.

Mayr, Ernst. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge & London: Harvard UP, 1982.

Milner, Richard. *The Encyclopedia of Evolution: Humanity's Search for Its Origins*. New York & Oxford: Facts on File, 1990.

Morasso, Pietro G., and Vittorio Sanguineti, eds. *Self-Organization, Computational Maps and Motor Control*. Amsterdam: Elsevier, 1997.

Murdock, George P. "The Common Denominator of Culture." *The Science of Man in the World Crisis*. Ed. R. Linton. New York: Columbia UP, 1945.

Solé, Ricard, and Brian Goodwin. *Signs of Life: How Complexity Pervades Biology*. New York: Basic Books, 2000.

Polanyi, Michael. *The Tacit Dimension*. Garden City: Doubleday, 1966.

Popper, Karl R. "Campbell on the Evolutionary Theory of Knowledge." *Evolutionary Epistemology, Theory of Rationality, and the Sociology of Knowledge*. Ed. Gerard Radnitzky and W. W. Bartley III. La Salle: Open Court, 1987. 115–20.

Rumelhart, David E., James L. McClelland, and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge: MIT P, 1986.

Sanger, T.D. "Probability Density Estimation for the Interpretation of Neural Population Codes." *Journal of Neurophysiology* 76 (1996): 2790–3.

Searle, John R. *Mind: A Brief Introduction*. Oxford: Oxford UP, 2004.

Stevens, Anthony. *Archetype Revisited: An Updated Natural History of the Self*. Toronto: Inner City Books, 2003.

Stevens, Anthony. *The Two Million-Year-Old Self*. College Station: Texas A&M UP, 1993.

Turner, Stephen. *Brains/Practices/Relativism: Social Theory After Cognitive Science*. Chicago & London: U Chicago P, 2002.

Turner, Stephen. "Cognitive Science, Social Theory, and Ethics." *Soundings: An Interdisciplinary Journal*, this issue (2007).

Turner, Stephen. "Practice Relativism." *Crítica: Revista Hispanoamerica de Filosofía* 39.115 (Apr. 2007): 5–29.

Turner, Stephen. "Practices Then and Now." *Human Affairs*, in press.

Turner, Stephen. "Social Theory as a Cognitive Neuroscience." *European Journal of Social Theory* 10 (2007): 359–77.

Appendix

For readers interested in the mathematics of the singular value decomposition (SVD), here it is in brief. Let $\mathbf{M}$ be any $m \times n$ matrix (e.g., a neural connection matrix in which M_{ij} is the strength of the connection to neuron i from neuron j.). Its SVD is the matrix product $\mathbf{U\Sigma V} = \mathbf{M}$, where $\mathbf{U} = \left[\mathbf{u}_1,\ldots,\mathbf{u}_m\right]$ and $\mathbf{V} = \left[\mathbf{v}_1,\ldots,\mathbf{v}_n\right]$ are orthogonal matrices. (The $\mathbf{u}_k$ are the columns of $\mathbf{U}$ and the $\mathbf{v}_k$ the columns of $\mathbf{V}$.) $\Sigma = \mathrm{diag}\left(s_1,\ldots,s_r,0,\ldots,0\right)$ is a diagonal matrix, in which $s_1,\ldots,s_r$ are the singular values, the nonzero diagonal elements of Σ, and r is the rank of $\mathbf{M}$. Everything can be arranged so that $s_1,\ldots,s_r$ are in non-increasing order, $s_1 \geq s_2 \geq \cdots \geq s_{r-1} \geq s_r$. $\mathbf{M} = \mathbf{U\Sigma V}$ can be written in the alternative form $\mathbf{M} = \sum_{k=1}^{r} s_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}$.

If a (column) vector of information $\mathbf{x}$ (representing a pattern of activities in a set of neurons) is fed into this connection matrix, the result will be the matrix product $\mathbf{y} = \mathbf{Mx}$, which represents the pattern of activity that is the output of the connection matrix. (This output is a linear combination of the inputs, and some readers will be aware that most neural networks are nonlinear, but the effect of the nonlinearities on the result $\mathbf{Mx}$ is unimportant for our purposes here.) Using the SVD, the output can be written in the alternative form $\mathbf{y} = \sum_{k=1}^{r} s_k \mathbf{u}_k \left(\mathbf{v}_k^{\mathrm{T}} \mathbf{x}\right)$. This expression can be interpreted in terms of a set of r implicit "soft rules," with the kth rule looking for pattern $\mathbf{v}_k$ and generating output

pattern $\mathbf{u}_k$. (Sometimes such a rule is written $\mathbf{v}_k \Rightarrow \mathbf{u}_k$.) The inner product (or scalar product) $\mathbf{v}_k^{\mathrm{T}}\mathbf{x}$ measures the similarity of the input $\mathbf{x}$ to $\mathbf{v}_k$. This number weights the output pattern $\mathbf{u}_k$ so that the more closely the input matches $\mathbf{v}_k$, the more strongly will $\mathbf{u}_k$ be represented in the output. In addition the *k*th implicit rule has an inherent strength or weight given by its singular value s_k. This reflects its overall influence (importance) in the behavior of the neural connection matrix.

A matrix representing the connections among thousands of neurons might have thousands of nonzero singular values, and so the exact representation of this network would require thousands of rules. (Indeed, since they are "soft rules"—i.e., they admit degrees of applicability—they are more expressive than the more familiar "hard rules," which either do or don't apply; an even larger number of hard rules would be required.) We can approximate the matrix with fewer rules by setting some of the singular values to zero, in effect eliminating the corresponding rules. Since we have put the singular values in non-increasing order, we can do this optimally by beginning with the smallest singular value s_r (corresponding to the weakest rule). Next we can eliminate s_{r-1}, and continue until we have reduced the rule set as far as we want. But each such approximation step eliminates some of the flexibility, subtlety, and sensitivity of the original neural network.