The Mechanisms of Emergence

R. KEITH SAWYER
Washington University

This article focuses on emergence in social systems. The author begins by proposing a new tool to explore the mechanisms of social emergence: multi agent–based computer simulation. He then draws on philosophy of mind to develop an account of social emergence that raises potential problems for the methodological individualism of both social mechanism and of multi agent simulation. He then draws on various complexity concepts to propose a set of criteria whereby one can determine whether a given social mechanism generates emergent properties, in the sense that their explanation cannot be reduced to a mechanistic account of individuals and their interactions. This combined account helps to resolve the competing claims of methodological individualists and social realists. The author's conclusion is that the scope of mechanistic explanation may be limited due to the extreme complexity of many social systems.

Keywords: *emergence; mechanism; computer simulation; methodological individualism; social realism*

In the 1990s, the 19th-century term "mechanism" reappeared in philosophy of biology (Bechtel 2001; Bechtel and Richardson 1993; Craver 2001, 2002; Glennan 1996; Machamer, Darden, and Craver 2000) and in philosophy of social science (Elster 1989; Hedström and Swedberg 1998; Little 1998; Stinchcombe 1991). Certainly by "mechanism," these philosophers do not mean that they believe that their systems were consciously designed and built, like the clockwork mechanisms of the 18th-century; most of what are now called mechanisms are not mechanical (Bunge 1997, 411; Machamer, Darden, and Craver 2000, 2). Philosophers of both biology and social science retain the belief that these systems were not consciously designed but developed through some sort of emergent, undirected process.

Bunge has been advocating systemist and realist approaches for several decades (1967, 1977; see the autobiographical comments in

Received 5 March 2003

Philosophy of the Social Sciences, Vol. 34 No. 2, June 2004 260-282
DOI: 10.1177/0048393103262553

1997), but a more explicit focus on social mechanism emerged in the mid-1990s, with Bunge's presentation at the 1996 "Social Mechanisms" (published as Bunge 1997) conference that resulted in the Hedström and Swedberg book (1998). In his older works, Bunge's usage of the term "mechanism" was consistent with the traditional 19th-century notion of mechanism as atomistic reduction (e.g., 1969/1973, 1977). By 1996, he was using the term "mechanism" in his current manner, as a form of explanation contrasted with empiricism (1996, 137-42; 1997, 414-20). Many philosophers have likewise advocated mechanistic approaches as a way to move beyond the deductive-nomological covering law model associated with logical empiricism (e.g., Hempel 1965). Mechanistic approaches implicitly assume a realist perspective and reject empiricism (Aronson, Harré, and Way 1995; Bhaskar 1975/1997; Layder 1990).

In most aspects, I agree with social mechanist (SM) approaches. I am a scientific realist not only about mechanisms but also about higher-level system properties. I believe that philosophers of science have much to gain from drawing on recent conceptions of mechanism, particularly from complex dynamical systems theory (as I do in this article). However, my dissatisfaction with SM is based on one significant disagreement: most social mechanists deny that social properties are real. This denial is implicit throughout their writings: for example, "A 'class' cannot be a causal agent because it is nothing but a constructed aggregation of occupational titles" (Hedström and Swedberg 1998, 11; also see Abbott 1996, 3). Whereas SM realism focuses on the mechanisms that realize emergent system properties, I argue for a realism about those properties themselves; and I argue that once social properties emerge, they have an ontological status distinct from their realizing mechanisms and may participate in causal relations. My argument for the realism of certain social properties is grounded in an emergentist, systemist, and mechanist approach.

In this article, I elaborate SM so that we can better understand which systems will require a macrosocial explanation rather than a methodologically individualist one. I do this by examining the mechanisms of emergence in complex dynamical systems. I show that some (although not all) emergent system properties cannot be fully explained in terms of mechanism (conceived of as a detailed description of the system's components and their interactions). They cannot be so explained when they are multiple realized in wildly disjunctive complex mechanisms.

MICROSOCIOLOGY AND MECHANISM

Predating this recent philosophical interest in mechanism, many microsociologists have argued that macro social properties can only be explained in terms of the mechanics of the relationships between individuals (Collins 1981; Ellis 1999; Lawler, Ridgeway, and Markovsky 1993; Rawls 1987, 1990; Ritzer and Gindoff 1992). This form of process theory in sociology extends back to Simmel (1950, 13): "[social entities should be explained] by the notion of societal production, according to which all these phenomena emerge in interaction among men, or sometimes, indeed, *are* such interactions" (also see Cederman 2002). In contemporary social theory, process ontologies are prominent (Sawyer 2002c) and have been advocated by theorists such as Abbott (1995), Emirbayer (1997), and Giddens (1984). In process theory, the preference for mechanistic, generative explanation is driven generally by the rejection of covering law models of science but more specifically by the perceived failure of sociology to identify lawful regularities in social systems (e.g., Giddens 1984). Many process theorists emphasize the importance of building theoretical models in which the components are often unobservable and must be hypothesized by the researcher (Barth 1981; Harré 1970; Miller 1987). These approaches, then, combine realism, process, and mechanism in presenting a version of social science that is opposed to empiricist, deductive-nomological approaches.

ARTIFICIAL SOCIETIES: SIMULATING THE MECHANISMS OF EMERGENCE

Since the mid-1990s, process and mechanism have been increasingly central in the study of complexity. Using a complexity methodology known as *multi agent systems*, researchers have begun to model the mechanisms of emergence. A multi agent system contains many autonomous computational agents that negotiate and collaborate with each other, in a distributed, self-organizing fashion. A few sociological theorists have begun to use this technology to develop what are called *artificial societies:* the developer implements models of the individual agents, chooses a formal communication language, establishes network connections between the agents, and then begins the simulation and observes the macro patterns that emerge. Artificial societies thus focus on three distinct sociological phenomena: the model of the individual, the model of the communication language, and the observation of emergent social phenomena. Such simulation tools allow social mechanists to explore the implications of their theories by running "virtual experiments" (Sawyer 2003a). The scientific realist foundations of this approach are strongly implicit among such researchers (Cederman 2002; Epstein 1999).

Artificial societies are mechanistic in the contemporary sense. They are also firmly methodologically individualist. Artificial society models represent a shift from equation-based modeling to agent-based modeling (Macy and Willer 2002; Sawyer 2003a; and this is consistent with Hedström and Swedberg 1998, 15-17). However, philosophers of social science who advocate a mechanistic understanding of social systems rarely ground their arguments in contemporary complexity science.

THE PROBLEM WITH MECHANISTIC EXPLANATION

Like SM, I am a scientific realist, and I believe that this necessitates a position between reductionism and holism. But my problem with SM is that it is methodologically individualist. I agree with Bunge's (1996, 145-50) claim that social explanation will require a combination of bottom-up and top-down explanatory mechanisms; yet in spite of this frequent claim, Bunge's systemism is difficult to distinguish from the bottom-up methodological individualism (MI) advocated by most social mechanists.[1] Rather than attacking MI as advocated and practiced by its defenders, Bunge created a straw man to attack, resulting in ineffective dismissals of MI rather than a sustained critique (e.g., 1996, 129, 243-53). For example, Bunge caricatured individualism by claiming that it considers only individuals in isolation: although Bunge acknowledged in passing that this is "radical" individualism (p. 255), his criticisms are limited to it. Furthermore, Bunge claimed that "individualism is incapable of accounting for the emergence . . . of social systems" (p. 256), when individualists from Menger and Hayek to Homans and Coleman viewed exactly this task as their primary goal. All intellectually interesting variants of MI accept that social bonds, social systems, and emergent properties exist, yet they

1. Also note that Boudon (1999) held that systemism is equivalent to methodological individualism.

argue that scientific explanation must proceed by reducing these to explanations in terms of individuals in networks of relations. Such misrepresentations of MI appear throughout Bunge's work; for example, he stated that individualists do not examine micro-macro relations (p. 256), when Coleman introduced his 1990 book by saying he would do exactly that, and Homans (1964) claimed that his 1961 book did that. (Bunge sidestepped this issue by redefining Coleman as a systemist: 1996, 148; 1997, 454).

The advocates of SM collected in Hedström and Swedberg (1998) are largely well-known methodological individualists, including Cowen, Elster, and Schelling, and the book includes a dedication to Coleman (suggesting that the editors disagreed with Bunge's claim that Coleman was not a methodological individualist). The editors' introductory chapter explicitly noted that "the principle of MI is intimately linked with the core idea of the mechanism approach" (p. 12), one of their four principles was "the general reductionist strategy in science" (p. 25), and both their examples and their typology of social mechanisms were methodologically individualist. Although Hedström and Swedberg included downward causation in their framework, calling it "situational mechanism," it was not seriously considered; as Abbott noted in his paper at the 1996 conference, "they allow it to drop with evident distaste" (p. 3). They closed by implicitly indicating their opposition to social realism: "there exist no such things as 'macro-level mechanisms' " (also see Little 1998, 198; Stinchcombe 1991, 367).

SM denies the causal power of emergent social properties and argues that "to identify a causal relation between two kinds of events or conditions, we need to identify the typical causal mechanisms through which the first kind brings about the second kind" (Little 1998, 202). But if social properties are real, then they can have autonomous causal powers.[2] Thus, SM denies a form of social realism that holds that emergent social properties can have downward causal powers (as advocated by Archer 1995; Bhaskar 1975/1997). Yet this antirealism is assumed rather than argued; SM has not engaged with arguments for social realism.

2. In contrast to most social mechanists, McAdam, Tarrow, and Tilly's (2001) analyses of contentious politics explicitly rejected the methodologically individualist version of social mechanism advocated in Hedström and Swedberg (1998). They accepted that social relations and transactions have "great causal efficacy" (p. 23) and argued that sociologists must examine mechanisms at both the individual and the collective level (p. 25).

When one takes MI seriously, it becomes difficult to distinguish from Bunge's systemism. Bunge (1977, 503) accepted that "every emergent property of a system can be explained in terms of properties of its components and of the couplings amongst these," and this is the version of emergence held by the methodological individualists, including Watkins, Homans, and many others (see Sawyer 2001). Methodological individualists accept the existence of emergent social properties yet claim that such properties can be reduced to explanations in terms of individuals and their relationships. MI's focus on micro-to-macro processes is explicitly considered to be a study of how social properties emerge from individual action (Axelrod 1997, e.g., 4; Coleman 1990; Epstein and Axtell 1996, e.g., 6-20; Homans 1964).

The unresolved sociological debate is about how explanation should proceed. Individualists argue that one should proceed by analyzing the system's components, then their relations and the behaviors of bigger system components, and all the way up until we have the explanation of the macro system property. All reductionist approaches in science are "mechanistic" in a broad sense, because even reductionists believe that scientific explanation of complex systems should be in terms of the system's components and relations, and the processes that occur among the components that give rise to emergent properties.[3]

The problem with systemism and SM more generally is that it leaves us with no principled way to distinguish between reducible and nonreducible systems, and thus we have no principled way to determine which explanatory approach is appropriate for a given emergent property.

A SOLUTION: EMERGENTISM

Laws are generally thought to be compatible with mechanisms; mechanisms "explain" laws (Bunge 2004 [this issue]; Elster 1998). However, it is possible that social laws may exist that are difficult to explain by reduction to micro mechanisms; if so, the scope of MI

3. Bunge has often described his systemism as the position that components must be studied not only in isolation but also to see how their behavior changes when in relations with other parts in complex systems. But even reductionists accept this truism: "If this is emergence, then this is a sort of emergence that the most reductionist and mechanist physicalist will never have dreamed of denying" (Smart 1981, 111; also see Phillips 1976, 35).

would be limited. In this section, I provide an account of emergence that shows how this could be so. I argue that some emergent social properties may be real, and may have autonomous causal powers, just like real properties at any other level of analysis. To the extent that social properties are real, SM explanation may be limited to the explanation of individual cases that do not generalize widely, resulting in an interpretivist or case study approach rather than a science of generalizable laws and theories.

Many sociological theorists use the philosophical notion of emergence to argue that collective phenomena are collaboratively created by individuals yet are not reducible to individual action (Sawyer 2001). In the social sciences, emergence refers to processes and mechanisms of the micro-to-macro transition (the "transformational mechanisms" of Hedström and Swedberg 1998, 22-23). Many of these accounts argue that although only individuals exist, collectives possess emergent properties that are irreducibly complex and thus cannot be reduced to individual properties (Archer 1995; Bhaskar 1979/1998; Sawyer 2001).

However, as I noted above, emergence has also been invoked by methodological individualists in sociology and economics. Thus, contemporary sociological uses of emergence are contradictory and unstable; two opposed sociological paradigms both invoke the concept of emergence and draw opposed conclusions. The same problem holds for contemporary versions of SM, and for the same reasons. I have attempted to address this problem by developing a sufficiently robust account of emergence so that the competing claims of individualists and collectivists can be reconciled. Using arguments from the philosophy of mind, I have argued for a version of emergence that I called *nonreductive individualism* (NRI; Sawyer 2001; 2002b). NRI holds to a form of *property dualism* in which social properties may be irreducible to individual properties, even though social entities consist of nothing more than mechanisms composed of individuals. After all, social realists generally are at least implicitly committed to the view that social kinds operate similarly to natural kinds (see Sawyer 2002b). NRI is an argument for realism concerning emergent social properties and is an argument against MI. The emergentist nature of NRI is compatible with a more limited form of mechanism, but one that is elaborated in a socially realist direction.

The NRI account of emergence is midway between individualism and holism. Counter individualism, I argue that an emergentist perspective is consistent with emergent social properties that have causal power, and SM generally rejects such properties (e.g., Little 1998). Counter holism, a focus on emergence accepts the need for a focus on individuals, their networks, and their interactions, and holists reject any need for consideration of this micro level. An exploration of the mechanisms of emergence has the potential to unify micro and macro sociological perspectives and to resolve longstanding debates between individualists and holists in the social sciences (cf. Sawyer 2001, 2002b, 2003c).

The argument for NRI uses notions of supervenience, multiple realizability, and wild disjunction to argue that the relation between emergent social properties and their realizing mechanisms is one of token identity but not of type identity. In other words, on any token occasion, a social property must be realized by a mechanism involving its components individuals (or more broadly, in Bunge's (2004) terms, realized in a system defined as a CESM quadruple). However, on different occasions, the same social property might be realized by different mechanisms (multiple realizability), and those different mechanisms might not be similar in any sociologically meaningful way (wild disjunction). This possibility would limit the scope of SM explanation.

Supervenience

Supervenience states that if a collection of individuals with distinct properties organized into a system operating according to distinct mechanisms causes a certain social property to obtain on one occasion, then that same collection of individuals in that same set of systemic relations operating according to the same mechanisms on another occasion will cause the same social property to obtain (Kim 1997, 188). Supervenience is consistent with mechanism qua MI: the idea is that every social property must be "realized" or "implemented" in some system of individuals and every social property emerges from a mechanism describable in terms of that system. Yet it does not entail MI due to multiple realizability (Heil 1998, 1999).

Multiple Realizability

If supervenience is to be used to ground a socially realist stance, one must develop a version of supervenience that argues that the reductionist approach of MI is not possible for some social properties. The argument that convinced most philosophers of mind that mental properties could not be reduced to physical properties was Putnam and Fodor's *multiple realizability* argument (Fodor 1974; Putnam 1967); nonreduction due to multiple realizability is now the consensus position in the philosophy of mind (Block 1997; Heil 1999).

Social properties are multiply realizable at the individual level (cf. Kincaid 1997, 17-20). For example, properties of groups like "being a church" or "having an argument" can be realized by a wide range of organizational structures, cultural practices, interactional patterns, and individual beliefs and dispositions. SM attempts to explain group properties by identifying their realizing mechanisms. If a group property is multiply realized, then there may be many mechanisms that realize it, a different one for each token instance. A successful translation of a token instance of a social property into its realizing mechanism is not sufficient to ground MI; MI claims not only that a given token event's social description can be reduced to its mechanism but that social properties qua types can be reduced. For example, it claims that one can identify a mechanism for the social property "having an argument" such that all possible realizations of this property can be described in systemic terms. If each token instance of the social property has a different mechanistic realization, then the mechanism explaining any one token instance would not be a complete explanation of the property.

But if all of the different realizing mechanisms of a social property are meaningfully related, then a broader mechanistic explanation could still be considered to have reductively explained the social property. MI would only be false if the realizing mechanisms were not meaningfully related. How could a social property have different realizations that would require radically different mechanistic explanations? Sawyer (2002b) showed that a social property might be multiply realized in different systems that are not meaningfully related, using Fodor's (1974) term *wild disjunction* for such higher-level properties. "Having an argument" or "being a church" could be realized on one token instance by one mechanism and realized on another occasion by a radically different type of mechanism. If the different realizing mechanisms are not meaningfully related, then the mechanistic, methodologically individualist approach would have limited explanatory power for that social property; an identification of the mechanism on one occasion would indeed provide an explanation of how that social property emerged on that occasion, but the explanation would not extend to the other realizing mechanisms contained within that same social property.

Like Fodor (1974) in the philosophy of mind, and Kincaid (1997) in the philosophy of social science, in such cases I argue that causal explanation need not cite underlying mechanisms. For example, the social property "being a church" can participate in causal laws, even though its realizations are wildly disjunctive. In fact, this sort of realist stance is quite standard in science more generally; each science standardly holds that its types and properties are real—"natural kinds" in realist parlance—and that its properties participate in causal laws. Pressure laws hold regardless of what particles make up the gas and regardless of the details of their relations and interactions. Airfoil laws hold with cloth sails and steel airplane wings, in varying temperature and humidity.

Together, multiple realization and wild disjunction provide an argument for social realism and against MI. MI could explain a social property only if the wildly disjunctive realizations at the individual level had lawful correspondences to the emergent social property. But if the disjunctive lower-level realizations are not lawfully related, then macrosocial laws are irreducible because of the way the world is put together: not all social properties correspond to a single mechanistic description in terms of individuals and their relations (cf. Fodor 1974, 131).

WHICH SOCIAL PROPERTIES ARE EMERGENT? LESSONS FROM COMPLEX SYSTEMS THEORY

The above account of multiple realizability grounds a relatively strong version of emergence: some social properties may participate in causal laws even though they cannot be reductively explained in terms of their lower-level realizing mechanisms. Such irreducible macro properties engage in macro-to-micro causal relations that have been called *downward causation* (Sawyer 2003d) or *situational mechanisms* (Hedström and Swedberg 1998). Thus, NRI is compatible with the existence of irreducible higher level causal laws; I refer to this as *supervenient causation*. NRI suggests that mechanist versions of social emergence are too reductionist in assuming that any emergent social property can ultimately be completely explained in terms of its realizing mechanism.

Whether or not a social property is emergent—in the sense that it is multiple realized in wildly disjunctive sets of lower-level mechanisms—can only be determined empirically (Fodor 1974; Sawyer

2002b). If we grant to individualism that higher-level social properties are not ontologically autonomous, but are supervenient on disjunctions of realizing mechanisms, then the individualist could respond that this seems compatible with MI. As Kim (1992) argued, the higher-level property is identical with the disjunction of all its realizing mechanisms. Thus, the individualist can accept wild disjunction and nonetheless respond, like Watkins (1957), that the only way to attain "rock bottom" explanations is to pursue the mechanistic description of the phenomenon. Thus, the wild disjunction argument does not conclusively demonstrate that SM is limited.

Thus, a key issue remains unresolved: which social properties are likely to submit to the reductionist form of explanation advocated by methodological individualists? Individualists generally take a pragmatic approach: assume that a social property can be so reduced until it becomes clear that no reductive explanation can be identified. In practice, SM accepts many limitations to individualist explanation: because many social groups have emergent properties that are the end result of long causal histories, it is essentially impossible to fully explain those social properties in the methodologically individualist fashion. Rather, sociologists must take certain macro properties as given and incorporate them into their explanation (Hedström and Swedberg 1998, 12-13). Although wild disjunction implies that the project will be difficult and complex, it does not entail its impossibility. In fact, recent developments in multi agent system simulations are rapidly increasing the scope of SM explanation.

In the rest of this article, I provide a framework that helps us to determine which properties are likely to submit to methodologically individualist reduction—such that "explanation" consists of describing the realizing system and its associated mechanisms—and which properties are not. For the latter properties—real emergent social properties—explanation may of necessity involve irreducible social properties and laws. I will draw on analyses of complexity, self-organization, and emergence to explore these questions. I conclude that for certain social mechanisms, we may be required to pursue macrosociological explanation in terms of emergent social properties, not only because of epistemological limits to current science but because of the structure of reality.

Here I argue that wild disjunction is a derivative property of complexity and that reductionism will fail for systems that cross some minimum threshold of complexity. If not all social properties can be explained by reduction, then which ones can? Complex dynamical systems theory provides sociologists with several hints about which social properties are likely to submit to reductionist explanation in terms of realizing mechanisms and which are unlikely to.

Number of Units

Most theories of emergence and complexity have proposed that as the number of component units increases, the likelihood of emergent higher-level properties increases. These arguments have been made by both philosophers of mind and artificial life researchers (Baas 1994; Cariani 1991; Darley 1994). The human brain contains around 100 billion neurons, and this large number has contributed to claims that the mind is a complex system that generates irreducibly emergent macro properties. The largest modern human societies contain several hundred million units (individuals), and they are highly interconnected.

Nonaggregativity

Wimsatt (1986) equated emergence with nonaggregativity. Aggregative properties meet four criteria, and most social properties do not satisfy them. First, the system property is not a product of the way the system is organized; the parts are *intersubstitutable* without affecting the system property. In social systems, individualists and collectivists alike agree that individuals and subsystems are not intersubstitutable because the network of relationships between individuals is significant. Second, an aggregative property should remain qualitatively similar under addition or removal of a part from the system. Third, the composition function for the property remains invariant under operations of decomposition and reaggregation of parts. Individualists and collectivists agree that these conditions do not hold of many social systems; for example, many social movements manifest threshold phenomena such that the addition or removal of the *N*th individual may result in a qualitative change in the system, even though individuals *N-1* and before did not. Fourth, there are no cooperative or inhibitory interactions among the parts; thus, the relation between parts and whole is linear (Bechtel and Richardson 1993, 266). Again, individualists and collectivists agree that this condition does not hold of most social systems, because relationships between individuals are often cooperative or inhibitory.

Most social properties are not aggregative and thus are emergent. But in Wimsatt's (1986) account, a property could be nonaggregative and nonetheless be reducible; even for nonaggregative properties, there must exist some composition function that relates the emergent property to a decomposition of the system into parts with relationships. Thus, individualists can easily accept that social systems are not aggregative. The disadvantage of Wimsatt's account is that it is so general that essentially all properties of complex systems will be nonaggregative (as Wimsatt acknowledged: 1997, S382), and because it does not address the issue of reducibility, it fails to speak to the essence of the sociological debate, which centers on whether a mechanistic account in terms of individuals and their interactions will be sufficient. Nonetheless, the characteristics associated with nonaggregativity are likely to contribute to the difficulty of reducing a given social property.

Near Decomposability

Decomposable systems are modular, with each component acting primarily according to its own intrinsic principles. Each component is influenced by the others only at its inputs; its function (processing of those inputs) is not itself influenced by other components (Simon 1969). In such a system, the behavior of any part is *intrinsically determined:* it is possible to determine the component's properties in isolation from the other components, despite the fact that they interact. The organization of the entire system is critical for the function of the system as a whole, but that organization does not provide constraints on the internal functioning of components. This was the concept of system that Parsons (1951) borrowed from cybernetics; Parsonsian structural-functionalism assumed decomposability in its elaborate identification of systems and subsystems and in its focus on status-role sets as decomposable components of systems.

In contrast, in nondecomposable systems, the overall system organization is a significant influence on the function of any component; thus, component function is no longer intrinsically determined. Dependence of components on each other is often mutual and may even make it difficult to draw firm boundaries between components (Bechtel and Richardson 1993, 26-27). Parsons acknowledged this possibility with his concept of "interpenetration" (e.g., Parsons and Shils 1951, 109), but this phenomenon remained a challenge to the essentially decomposable emphasis of his systems model. Systems that are not nearly decomposable are likely to have emergent system properties that are wildly disjunctive at the level of description of the components, and such systems are thus less likely to submit to mechanist explanation.

Localization

A system is localizable if the functional decomposition of the system corresponds to its physical decomposition and each property of the system can be identified with a single component or subsystem. Functional localizability was a foundational assumption running through Parsons's systems theory. In his AGIL scheme, each of the four major systems is defined in terms of its function. Likewise, the lowest-level components—roles—are defined in terms of the function they serve for the system (Parsons 1951, 115). Collectivities are likewise conceptualized in role terms and thus defined in terms of their functions (Parsons and Shils 1951, 190-97). The allocation process—which together with the integration process allows systems to maintain equilibrium—serves the function of allocating functions to roles and to subsystems (Parsons and Shils 1951, 108, 198).

If social properties cannot be identified with components, but are instead distributed within the system, that system is not localizable (Bechtel and Richardson 1993, 24). Many social properties are not localizable. For example, "being a church" cannot be localized to any of the individuals belonging to the church or to any subnetwork of those individuals. Social properties that are not localizable are likely to have wildly disjunctive descriptions at the level of individual mechanism and thus are more likely to resist mechanist explanation (Bechtel and Richardson 1993, 228). Such systems are more likely to manifest emergence than the localizable systems proposed by Parsons.

The brain is generally agreed to be nonlocalizable in this sense, and much of the theory about localizability has been inspired by connectionist models of brain function in cognitive science. Connectionist models suggest that the density of network connections is related to localizability and decomposability of the system. Likewise, social systems with a high dynamic density are less likely to be decomposable or localizable and as such are more likely to manifest social properties that are wildly disjunctive at the individual level of description. In modern societies, dynamic density increases as communication and transportation technology advance, increasing the number and frequency of network connections among individuals (cf. Durkheim 1895/1964, 114-15; Sawyer 2002a).

Complexity of Communication

All emergentists, systemists, and mechanists agree that interaction between components is central to micro-macro process accounts. In complexity theory, the above criteria of nonaggregativity, nondecomposability, and nonlocalizability are all defined in terms of the complex systemic relations among components. Consequently, several complexity theorists have suggested that the complexity of each interaction between components may be another variable contributing to emergence (Baas 1994; Darley 1994; Sawyer 2003c). Artificial society theorists have proposed that as the communication language between agents becomes more complex, the likelihood of irreducible macro patterns emerging increases (Sawyer 2003b). The extreme complexity of human symbolic interaction thus contributes to the irreducibility of social properties.

In mechanical and biological systems, component relations are relatively well understood and well defined. Because they are inspired by such systems, complex dynamical systems models tend to assume extremely simple interactions. Yet although human communication is qualitatively more complex, emergentists have not connected their theories to the study of symbolic interaction in groups. If there is a qualitative difference in the complexity of this communication and that in natural complex systems, then social theories of emergence may need to incorporate a theory of symbolic interaction (cf. Sawyer 2003b).

Nonlinearity

Many complexity theorists emphasize the *nonlinearity* of dynamical systems. In nonlinear systems, the effect may not be proportional to the cause; a small change in initial conditions can lead to a radical change in the later state of the system—the so-called butterfly effect—or inversely, a large change in initial conditions might not lead to any significant change in the later state of the system. Chaotic equations are a specific type of nonlinear equation. Systems described by chaotic equations are often observed to possess *attractors*, relatively steady states that the system tends to gravitate towards. Chaos theorists argue that the reductive methods of traditional science are difficult to apply to such systems because many of the mathematical tools used in traditional science were developed for linear systems and are not applicable to nonlinear systems (Casti 1994).

Gell-Mann (1994, 111) argued that in chaotic systems, reduction to explanation in terms of components is practically impossible although possible in principle (also see Baas 1994, 519). Kauffman (1993) took an even stronger position that the derivation is not possible even in principle. This unpredictability is in large part due to extreme sensitivity to initial conditions, so much so that scientists could never identify and characterize the initial conditions with the requisite degree of detail to allow prediction of the dynamic path of the system. Thus, for both practical and theoretical reasons, complexity theorists argue that nonlinearity makes explanation by reduction to realizing mechanism impossible.

Open and Closed Systems

Open systems are systems in which energy, information, or matter flows between the system and its environment. Many theorists argue that earlier versions of cybernetics-derived theory are only appropriate for *closed systems*. A common citation in this context is the work of Nobel-prize-winning physicist Ilya Prigogine on dissipative structures, systems far from thermodynamic equilibrium (Nicolis and Prigogine 1977; Prigogine and Stengers 1984). The most well-known dissipative structure is the Bénard convection cell. These cells form when a layer of liquid is heated from below; at a given temperature, the heated molecules at the bottom of the layer start to rise, and the cooler ones on top start to sink. If the temperature is too high, the liquid reaches a turbulent state without any regular, stable structure; but at just the right temperature, the flow between the top and bottom of the liquid results in the emergence of regular, stable patterns known as Bénard cells, typically with a hexagonal shape, with the hotter fluid rising up the center of the cell and the cooler fluid sinking at the edges. Although a system in this state is quite chaotic, and individual particles are moving rapidly, the cells remain remarkably stable. This is an open system because it requires the application of an external heat source to occur. In fact, it is generally thought that the structures of open, nonequilibrium many-particle systems are more complex than the structures found in systems at equilibrium.

Sociologists generally accept that social systems are open (McIntyre 1996; Parsons 1970). Talcott Parsons may have been familiar with some of Prigogine's early work when he wrote in 1970 about "the strong emphasis in modern biological thinking that living systems are open systems, engaged in continual interchange with their environments" (p. 850). In the 1950s and 1960s, cybernetics and information theory became a big influence on Parsons, who began to think that "control in action systems was of the cybernetic type" and not coercive.

Computational Theory and Mathematical Undecidability

Computational emergence is the "finite analogue of formal undecidability" (Darley 1994, 416). Undecidability is a concept regarding the halting problem—one cannot know if the computation can be decided in finite time. Several mathematicians have demonstrated that various behaviors of collectives are undecidable (Meyer and Brown 1998; Wolfram 1985). Although the mechanistic description of the system at time t might in principle predict the mechanistic description of the system at time $t + 1$, the rules describing the connection might be so chaotic as to be mathematically noncomputable. A mechanistic explanation can never be found if the relation between a social property and its realizing mechanisms is demonstrably noncomputable. Although the microexplanation is in principle not possible, the macrolevel explanation may nonetheless be possible, because the macrolevel explanation is an approximation which admits of exceptions (Fodor 1974).

The Cost of Explanation

von Neumann, one of the founders of computer science, was the first to suggest that the simplest description of a complex system might be its simulation (von Neumann 1949/1966, 31-41, 47; also see Dupuy 2000, 140-43; this argument is more commonly associated with Simon 1969). For such a system, one cannot deduce all of its properties from the description of its mechanism; rather, the simulation must be run to determine its properties. Such arguments have become increasingly widespread in complexity science, particularly among those who use multi agent systems (this is also Wolfram's 2002 argument concerning cellular automata).

These insights have complex and unresolved implications for SM, first because they raise the possibility that the only potential mechanistic "explanation" would be a run of the simulation (and philosophers have not elaborated what sort of explanation this would constitute), and second because running and then analyzing the simulation might be less efficient than prediction and explanation in terms of the higher-level laws. Dupré (1993) noted that reductionist work in the human sciences can give us good lower level theories of *how* systems do what they do but not exactly *what* those systems do; likewise, mechanists including Bunge and Hedström and Swedberg agree that mechanistic explanations cannot predict, only explain. Lower-level mechanisms do not make predictions about how the system will change over time; to address these dynamic questions, we may need to use the higher level, even when we already have a good mechanistic understanding of the realizing system (Godfrey-Smith 1999, 177). Higher-level properties may be ineliminable because they provide the lowest-cost and highest-benefit descriptions of the regularities in the phenomena at that level (Wimsatt 1976). If so, there are grounds for the retention of causal explanations at the higher level.

CONCLUSION

Mechanistic metaphors for social systems have existed for centuries; recently, philosophers of mind, biology, and social science have rejuvenated this metaphor. I have extended this work by drawing on complex dynamical systems theory. This recent tradition has developed artificial society simulations of social mechanisms, and the analysis of these simulations has much to offer to the philosophy of mechanism. I drew on this tradition to explore an unaddressed issue: which emergent social properties will submit to mechanistic explanations, and what is the potential usefulness and scope of such explanations?

The philosophers best known for emphasizing the importance of explanation are those who have advocated a close focus on specific cases—the interpretivists—and by doing so, they also deny the possibility of general social laws. The extreme case of wild disjunction would be a social property for which no two token realizing instances are meaningfully related. This limit case would lead to the interpretivist position that the explanation of any given social phenomenon, while valuable, tells us nothing about any other cases. Bunge realized this, and it is why he (alone among the social mechanists) repeatedly distinguished his position from the interpretivists and why he emphasized that mechanism is compatible with covering-law approaches rather than a rejection of them. It remains a risk that a mechanistic approach could provide wonderful explanations of specific token instances but degenerate into interpretivism, thereby losing all generality and thus being of limited scientific value.

To address such problems, I summarized the framework of NRI, which draws on arguments of supervenience, multiple realizability, and wild disjunction. I argued that whether or not a complex social system is best described in terms of its mechanisms or in terms of its macro behavior is an empirical question. It cannot be known a priori whether a given social property can be given a useful mechanist explanation. Working within the mechanistic approach of complexity science, I then presented several features of complex systems that are likely to lead them to generate emergent properties. These system features lead to irreducibility because they result in wildly disjunctive realizations of social properties, such that those properties are realized by different mechanisms on different occasions. These implications of multiple realizability are rarely noted in the writings of the social mechanists.

Even if individualist mechanism is so limited, sociologists could still develop scientific explanations in terms of systems and mechanisms, but the description of the system and mechanism would use the terms and properties of macro sociology in addition to individual properties and relations. Although mechanism is commonly associated with MI—because its advocates assume that a social mechanism must be described in terms of individual's intentional states and relations (e.g., Elster 1998)—there is no reason why science cannot include systems and mechanisms at higher levels of analysis (Wight 2004 [this issue]). After all, individual properties such as intentional states are themselves realized in the lower-level substrate of neurons and their synaptic connections; on what grounds does SM hold that individual properties should be allowed in mechanistic explanation but not social properties (Sawyer 2002a)?

I believe that the history of research in the social sciences shows that social systems meet many of the criteria for emergent complex systems, and I have provided some brief citations to such features in this paper (also see Sawyer 2001; 2002b; 2003c). Ultimately, the success or failure of microexplanation in terms of individuals must be determined through empirical research. The social mechanism approach is valuable exactly because it provides a framework that allows us to carry out this exploration.

REFERENCES

Abbott, Andrew. 1995. Things of boundaries. *Social Research* 62:857-82.

———. 1996. Mechanisms and relations. Paper presented in the Conference on Social Mechanisms, Stockholm, Sweden.

Archer, Margaret S. 1995. *Realist social theory: The morphogenetic approach*. New York: Cambridge University Press.

Aronson, Jerrold L., Rom Harré, and Eileen Cornell Way. 1995. *Realism rescued: How scientific progress is possible*. Chicago: Open Court.

Axelrod, Robert. 1997. *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press.

Baas, Nils A. 1994. Emergence, hierarchies, and hyperstructures. In *Artificial life III, SFI studies in the sciences of complexity, Proc. Vol. XVII*, edited by Christopher G. Langton, 515-37. Reading, MA: Addison-Wesley.

Barth, Fredrik. 1981. *Process and form in social life: Selected essays of Fredrik Barth*. London: Routledge & Kegan Paul.

Bechtel, William. 2001. The compatibility of complex systems and reduction: A case analysis of memory research. *Minds and Machines* 11:483-502.

Bechtel, William, and Robert C. Richardson. 1993. *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

Bhaskar, Roy. 1975/1997. *A realist theory of science*. New York: Verso Classics.

———. 1979/1998. *The possibility of naturalism*. New York: Routledge.

Block, Ned. 1997. Anti-reductionism slaps back. *Philosophical Perspectives* 11:107-32.

Boudon, Raymond. 1999. Foreword. In *The sociology-philosophy connection*, edited by Mario Bunge, xi-xv. New Brunswick, NJ: Transaction Publishers.

Bunge, Mario. 1967. *Scientific research*. Berlin: Springer-Verlag.

———. 1969/1973. The metaphysics, epistemology and methodology of levels. In *Method, model, and matter*, edited by Mario Bunge, 160-68. Dordrecht, the Netherlands: D. Riedel. (Adapted from *Hierarchical structures*, edited by Lancelot Law Whyte, Albert G. Wilson, and Donna Wilson, 17-28, New York: American Elsevier, 1969.)

———. 1977. General systems and holism. *General Systems* 22:87-90.

———. 1996. *Finding philosophy in social science*. New Haven, CT: Yale University Press.

———. 1997. Mechanism and explanation. *Philosophy of the Social Sciences* 27:410-65.

———. 2004. How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences* 34:182-210.

Cariani, Peter. 1991. Emergence and artificial life. In *Artificial life II, SFI studies in the sciences of complexity, Vol. X*, edited by C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, 775-97. Reading, MA: Addison-Wesley.

Casti, John L. 1994. *Complexification: Explaining a paradoxical world through the science of surprise*. New York: HarperCollins.

Cederman, Lars-Erik. 2002. Computational models of social forms: Advancing generative macro theory. Paper presented at Agents 2002: Social Agents: Ecology, Exchange, and Evolution, October 11-12, Chicago.

Coleman, James S. 1990. *Foundations of social theory*. Cambridge, MA: Harvard University Press.

Collins, Randall. 1981. On the microfoundations of macrosociology. *American Journal of Sociology* 86:984-1014.

Craver, Carl. 2001. Role functions, mechanisms and hierarchy. *Philosophy of Science* 68:31-55.

———. 2002. Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science* 69:S83-S97.

Darley, Vince. 1994. Emergent phenomena and complexity. In *Artificial life IV, Proceedings of the fourth international workshop on the synthesis and simulation of living systems*, edited by Rodney A. Brooks and Pattie Maes, 411-16. Cambridge, MA: MIT Press.

Dupré, John. 1993. *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge, MA: Harvard University Press.

Dupuy, Jean-Pierre. 2000. *The mechanization of the mind: On the origins of cognitive science*. Princeton, NJ: Princeton University Press.

Durkheim, Émile. 1895/1964. *The rules of sociological method*. New York: Free Press. (Orig. pub. as *Les régles de la mèthode sociologique*, Paris: Alcan, 1895.)

Ellis, Donald G. 1999. Research on social interaction and the micro-macro issue. *Research on Language and Social Interaction* 32:31-40.

Elster, Jon. 1989. *Nuts and bolts for the social sciences*. New York: Cambridge University Press.

———. 1998. A plea for mechanism. In *Social mechanisms: An analytical approach to social theory*, edited by Peter Hedström and Richard Swedberg, 45-73. New York: Cambridge University Press.

Emirbayer, Mustafa. 1997. Manifesto for a relational sociology. *American Journal of Sociology* 103:281-317.

Epstein, Joshua M. 1999. Agent-based computational models and generative social science. *Complexity* 4:41-60.

Epstein, Joshua M., and Robert Axtell. 1996. *Growing artificial societies: Social science from the bottom up*. Cambridge, MA: MIT Press.

Fodor, Jerry A. 1974. Special sciences (Or: The disunity of science as a working hypothesis). *Synthese* 28:97-115.

Gell-Mann, Murray. 1994. *The quark and the jaguar: Adventures in the simple and the complex*. New York: W. H. Freeman.

Giddens, Anthony. 1984. *The constitution of society: Outline of the theory of structuration*. Berkeley: University of California Press.

Glennan, Stuart S. 1996. Mechanisms and the nature of causation. *Erkenntnis* 44:49-71.

Godfrey-Smith, Peter. 1999. Procrustes probably: Comments on Sober's "Physicalism from a probabilistic point of view." *Philosophical Studies* 95:175-81.

Harré, Rom. 1970. *The principles of scientific thinking*. London: Macmillan.

Hedström, Peter, and Richard Swedberg, eds. 1998. *Social mechanisms: An analytical approach to social theory*. New York: Cambridge University Press.

Heil, John. 1998. Supervenience deconstructed. *European Journal of Philosophy* 6:146-55.

———. 1999. Multiple realizability. *American Philosophical Quarterly* 36:189-208.

Hempel, Carl G. 1965. *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.

Homans, George C. 1961. *Social behavior: Its elementary forms*. New York: Harcourt, Brace, & World.

———. 1964. Commentary. *Sociological Inquiry* 34:221-31.

Kauffman, Stuart A. 1993. *The origins of order: Self-organization and selection in evolution*. New York: Oxford University Press.

Kim, Jaegwon. 1992. Multiple realizability and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52:1-26.

———. 1997. The mind-body problem: Taking stock after forty years. *Philosophical Perspectives* 11:185-207.

Kincaid, Harold. 1997. *Individualism and the unity of science*. New York: Rowman & Littlefield.

Lawler, Edward J., Cecilia Ridgeway, and Barry Markovsky. 1993. Structural social psychology and the micro-macro problem. *Sociological Theory* 11:268-90.

Layder, Derek. 1990. *The realist image in social science*. New York: St. Martin's.

Little, Daniel. 1998. *Microfoundations, method and causation: On the philosophy of the social sciences*. New Brunswick, NJ: Transaction Publishers.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67:1-25.

Macy, Michael W., and Robert Willer. 2002. From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology* 28:143-66.

McAdam, Doug, Sidney Tarrow, and Charles Tilly. 2001. *Dynamics of contention*. New York: Cambridge University Press.

McIntyre, Lee C. 1996. *Laws and explanation in the social sciences: Defending a science of human behavior.* Boulder, CO: Westview.

Meyer, D. A., and T. A. Brown. 1998. Statistical mechanics of voting. *Physical Review Letters* 81:1718-21.

Miller, Richard W. 1987. *Fact and method: Explanation, confirmation and reality in the natural and the social sciences*. Princeton, NJ: Princeton University Press.

Nicolis, G., and I. Prigogine. 1977. *Self-organization in nonequilibrium systems*. New York: John Wiley.

Parsons, Talcott. 1951. *The social system*. Glencoe, IL: Free Press.

———. 1970. On building social system theory: A personal history. *Daedalus* 99:826-81.

Parsons, Talcott, and Edward A. Shils. 1951. Values, motives, and systems of action. In *Toward a general theory of action*, edited by Talcott Parsons and Edward A. Shils, 47-275. Cambridge, MA: Harvard University Press.

Phillips, D. C. 1976. *Holistic thought in social science*. Stanford, CA: Stanford University Press.

Prigogine, Ilya, and Isabelle Stengers. 1984. *Order out of chaos: Man's new dialogue with nature*. New York: Bantam.

Putnam, Hilary. 1967. Psychological predicates. In *Art, mind, and religion: Proceedings of the 1965 Oberlin Colloquium in Philosophy*, edited by W. H. Capitan and D. D. Merrill, 37-48. Pittsburgh, PA: University of Pittsburgh Press.

Rawls, Anne Warfield. 1987. The interaction order sui generis: Goffman's contribution to social theory. *Sociological Theory* 5:136-49.

———. 1990. Emergent sociality: A dialectic of commitment and order. *Symbolic Interaction* 13:63-82.

Ritzer, George, and Pamela Gindoff. 1992. Methodological relationism: Lessons for and from social psychology. *Social Psychology Quarterly* 55:128-40.

Sawyer, R. Keith. 2001. Emergence in sociology: Contemporary philosophy of mind and some implications for sociological theory. *American Journal of Sociology* 107:551-85.

———. 2002a. Durkheim's dilemma: Toward a sociology of emergence. *Sociological Theory* 20:227-47.

———. 2002b. Nonreductive individualism, Part 1: Supervenience and wild disjunction. *Philosophy of the Social Sciences* 32:537-59.

———. 2002c. Unresolved tensions in sociocultural theory: Analogies with contemporary sociological debates. *Culture & Psychology* 8:283-305.

———. 2003a. Artificial societies: Multi agent systems and the micro-macro link in sociological theory. *Sociological Methods and Research* 31:37-75.
———. 2003b. Communication matters: Agent communication languages and emergence in artificial societies. Manuscript, Washington University, St. Louis, MO.
———. 2003c. *Improvised dialogues: Emergence and creativity in conversation*. Westport, CT: Greenwood.
———. 2003d. Nonreductive individualism, Part 2: Social causation. *Philosophy of the Social Sciences* 33:203-24.
Simmel, Georg. 1950. *The sociology of Georg Simmel*. New York: Free Press.
Simon, Herbert A. 1969. *The sciences of the artificial*. Cambridge, MA: MIT Press.
Smart, J. J. C. 1981. Physicalism and emergence. *Neuroscience* 6:109-13.
Stinchcombe, Arthur L. 1991. The conditions of fruitfulness of theorizing about mechanisms in social science. *Philosophy of the Social Sciences* 21:367-88.
Von Neumann, John. 1949/1966. Theory and organization of complicated automata. In *Theory of self-reproducing automata*, edited by Arthur W. Burks, 29-87. Urbana: University of Illinois Press. (Orig. presented as five lectures delivered at the University of Illinois in December 1949.)
Watkins, J. W. N. 1957. Historical explanation in the social sciences. *British Journal for the Philosophy of Science* 8:104-17.
Wight, Colin. 2004. Theorizing the mechanisms of conceptual and semiotic space. *Philosophy of the Social Sciences* 34:283-299.
Wimsatt, William C. 1976. Reductionism, levels of organization, and the mind-body problem. In *Consciousness and the brain*, edited by G. G. Globus, G. Maxwell, and I. Slavodnik, 199-267. New York: Plenum.
———. 1986. Forms of aggregativity. In *Human nature and natural knowledge*, edited by Alan Donagan, Anthony N. Perovich Jr., and Michael V. Wedin, 259-291. Dordrecht, the Netherlands: D. Riedel.
———. 1997. Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science* 64 (Proceedings): S372-S384.
Wolfram, Stephen. 1985. Undecidability and intractability in theoretical physics. *Physical Review Letters* 54:735-38.
———. 2002. *A new kind of science*. Champaign, IL: Wolfram Media.

R. Keith Sawyer is an associate professor in social thought and analysis and education at Washington University in St. Louis. He is the author of several recent articles in the philosophy of science and in sociological theory. This theoretical work is inspired by two empirical studies of symbolic interaction, Pretend Play as Improvisation *(Lawrence Erlbaum, 1997) and* Improvised Dialogues *(Greenwood, 2003), both of which propose collectivist explanations for microsocial processes.*