Self-Assembling Systems

Paul Humphreys†

Starting with the view that methodological constraints depend upon the nature of the system investigated, a tripartite division between theoretical, semitheoretical, and empirical discoveries is made. Many nanosystems can only be investigated semitheoretically or empirically, and this aspect leads to some nanophenomena being weakly emergent. Self-assembling systems are used as an example, their existence suggesting that the class of systems that is not Kim-reducible may be quite large.

1. Introduction. Suppose we ask the question, What is the most effective route to knowledge in a given domain? There are various answers to that question, drawing on an array of methods that includes theories, models, observation, experiment, measurement, statistical inference, and computational science. These are all effective routes to scientific knowledge, but which methods to choose and in which combination depends upon system-specific constraints. A partial list of such constraints would include the number of degrees of freedom in a system, the system's complexity, the speed and available memory of computational devices, and the size of the system. Some of these constraints, such as the available memory, come into play only because other constraints that are a result of the nature of the system itself, such as system complexity, are in operation. Others, such as the number of degrees of freedom in the system, are properties of the systems themselves. Thus, one answer to our question is based upon what we can call the realist's credo: *It is the nature of the phenomena in the domain at hand, either directly or indirectly, that determines the most effective routes to knowledge for that domain.*

The credo is not a particularly exciting statement, but it does have some bite. For example, it leaves as a factual matter what works in a given domain and it thus requires us to reject certain philosophical positions that impose universal or noncontingent constraints on our methods. Some

†To contact the author, please write to: Paul Humphreys, Department of Philosophy, 120 Cocke Hall, P.O. Box 400780, University of Virginia, Charlottesville, VA 22904; e-mail: pwh2a@cms.mail.virginia.edu.

Philosophy of Science, 73 (December 2006) pp. 595–604. 0031-8248/2006/7305-0012$10.00

examples of these positions are unification accounts of explanation, methodological physicalism, and noncontingent versions of humean supervenience. The credo does not preclude general methods—it just leaves open as a matter of contingent fact how generally applicable a given method is. Looked at from that perspective, empiricism, whether constructive or traditional, should be construed as a subject-matter-dependent method rather than a universal epistemological constraint because some features of some systems lend themselves to empiricist methods, whereas others do not. But the constraints empiricism imposes come from the wrong direction. This is because many antirealist positions, including empiricism and instrumentalism, have placed constraints on scientific methods using human limitations as the primary criterion, rather than emphasizing the constraints imposed on scientific methods by the world itself. Indeed, one particular thing that the credo calls to our attention is that empiricism and realism need not be in conflict, even though they have traditionally been seen to be. Some constraints result from the fact that the system properties do not lend themselves to direct observation, but that constraint is only one among many and it can be overcome in many cases.[1] For a considerable portion of the physical and biological sciences, at least, these anthropocentric constraints can be relaxed by using nonhuman, often automated procedures that have proved to be remarkably effective at providing scientific knowledge. Astronomical telemetry, genome sequencing, the scientific investigation of nanomaterials, each of these and many other techniques are carried out by automated processes, and it is the nature of the interaction between those instruments and those systems (which themselves can be at the nanolevel in some cases) that constrains what we can know. For empiricists, the size of the system often serves as a proxy for its degree of observability, but one of the principal morals of this paper is that in the nanoscale realm it is frequently neither the size of the objects involved nor their claimed unobservability that creates epistemological problems, but other features of the system.

The nature of nanophenomena places special, although not unique, constraints on effective methods. Within the area of nanoscience, computer simulations, automated instruments, and data-reading devices such as the atomic force microscope have played key roles. There is one method used within nanoscience that has a somewhat different flavor to it and that usefully illustrates both the realist's credo and how automated procedures can lead to knowledge. This is the method of self-assembly of molecules and other systems, sometimes also called self-organization. I

1. Arguments for this claim are given in detail in Humphreys (2004).

shall start with some more general issues before addressing the specifics of self-assembling systems.

2. Theoretical, Semitheoretical, and Irreducibly Empirical Discoveries. What exactly constitutes nanophenomena is not precisely defined, but nanoscale phenomena are generally considered to lie in a size range that is intermediate between individual molecules and bulk materials. If we take the nano-area to be roughly the size region between 10^{-10} and 10^{-7} meters, then the choice of feasible methods and the constraints on them can result either directly from the size of the phenomena, or indirectly from features contingently associated with the size, such as the energy scale, the complexity of the systems considered, or the surface-to-volume ratio. It is the need to deal with this supramolecular realm, with its multiple interacting entities having complex internal structure, rather than the scale itself, that produces most of the constraints in the area.

Scale invariance is not a standard feature of laws of nature, although there are structural features that are known to be scale invariant. Perhaps the best-known examples are systems with features that are described by power laws. For example, some features of self-avoiding random walks are scale invariant, and the entire field of self-organized criticality is based upon this self-similarity at different scales. At least in theoretical models, large and small avalanches in sandpiles are subject to the same laws. But most systems do not have this feature, and the failure of scale invariance is illustrated by the common gap between laboratory research and industrial application, resulting from the fact that not all discoveries in the lab can be numerically scaled up to be mass produced.

In fact, and this is one of the interesting things about nanotechnology, matter sometimes behaves quite differently at different length scales. For example, gold, which is largely inert at macroscopic scales, is both a chemical catalyst and bioactive at the nanoscale. One explanation for this change is that the volume fraction of the surface atoms and the interface atoms increases as the size of the gold particles becomes smaller, and as a result the electronic structure of the gold particles changes. This is one reason why the simplistic identification of gold as being identical with whatever has the atomic number 79 is wrong. Collections of gold atoms (which are not aggregates in the colloquial sense of that term) have different properties depending upon the size of the collection and how it is spatially arranged.

Now suppose that we need to discover how a given system will behave at some length scale. In what follows, I shall restrict my attention to dynamical systems. Using some familiar terminology, consider the following definitions:

> **Definition 1**. A state of a system S is discovered *purely theoretically* just in case S comes to be known without any use of empirical data.

Some things to note about this definition: It allows that a given state can be discovered both purely theoretically and in other ways. It is couched in actualist terms and appeals to what can be done in practice, rather than at some in-principle limit to which we do not have epistemic access. Predictions are special cases of discovery, but a temporal component is excluded from the definition in order to allow for retrospective discoveries and theoretical procedures that are too slow to arrive at a correct description of S before S occurs. Finally, I have used the term "theoretical" rather than "a priori" because the latter has anthropocentric associations that I want to avoid for reasons mentioned earlier.

Purely theoretical discoveries are familiar features of science. Although the terminology is by no means uniform, call a method of calculating energy levels in an atom or a molecule an ab initio method if it carries out an exact or approximate calculation of those levels based on Schrödinger's equation using only the positions of the nuclei and the electrons in the Hamiltonian. Then an example of a purely theoretically discovered state is a successful ab initio calculation from Schrödinger's equation of an energy state of an atom.[2] A second example of a purely theoretical discovery is the successful prediction of the mean square end-to-end distance of a two-dimensional random walk. The system involved can be formal, as it will be when predicting computational states[3] in a computer simulation; or material, as it will be when predicting the states of some physical system.

> **Definition 2**. A state S of a system is discovered *semitheoretically* just in case S comes to be known when at least one, but not every, computational or deductive step in a theoretical discovery of S is replaced by a step that uses empirically gathered data.

For example, the calculation of gross features of large molecules can be carried out by considering only the outer (valence) electrons, and using the net overall charge of the nucleus plus the inner electrons for the rest. Such predictions are semitheoretical in our sense because in the Huckel approach, to take one method, the integrals use, in part, energy values derived from spectroscopic observations of real atoms. Examples of semitheoretical methods include most of what are usually called semiempirical

2. Note that many so-called ab initio methods add modeling approximations to the standard apparatus of quantum theory and do not rely only upon Schrödinger's equation and a choice of basis vectors.

3. As distinct from the physical states that implement the computational state.

methods in molecular chemistry and elsewhere. I say "most," rather than "all," because although some semiempirical methods do use experimentally determined values as a substitute for infeasible computations, others employ values that are computed using models of similar systems and the values obtained are then transferred to the system under consideration.

> **Definition 3**. A state S of a system is discovered *empirically* just in case S is predicted by letting the system itself dynamically evolve. If there are no available processes by means of which S can be discovered theoretically or semitheoretically, call the discovery of S *irreducibly empirical.*

Irreducibly empirical discoveries should be familiar. Recall David Marr's distinction between type-1 and type-2 theories. A type-2 theory of a process is one that predicts the evolution of the process in such a way that the process is its own simplest description. There is an obvious conceptual connection between processes described by type-2 theories and data sequences that are classified as random under the Kolmogorov/Chaitin computational complexity approach to randomness within which a sequence of syntactical objects is random just in case the shortest program that has the sequence as output is equal to the length of the sequence itself. A type-1 theory, in contrast, allows for information compression; there exits a syntactically simple theory from which the development of the process can be predicted in the traditional sense favored by philosophers. One thing to note is that Marr was quite explicit in claiming that the type-1/type-2 distinction does not constitute a dichotomy. There are many degrees of complexity between the two extremes, and in that region lie the subjects of many semitheoretical discoveries.

A second kind of phenomenon that straightforwardly falls into the domain of irreducibly empirical discoveries is weak emergence. Weak emergence is defined by Bedau in the following way: "Assume that P is a nominally emergent property possessed by some locally reducible system S. Then P is weakly emergent if and only if P is derivable from all of S's micro facts but only by simulation" (Bedau 2003, 12). Some elaboration: A nominally emergent property is a property that can be possessed at the macro-level but is in principle incapable of being possessed at the micro-level. For example, individual cells in a cellular automaton can only be square and so the property 'is rectangular' is nominally emergent within such a system. A locally reducible system is, roughly, a system in which all of the macro-properties are structural properties, that is, the state of a micro-entity consists of its location and intrinsic properties, whereas the state of a macro-entity is simply the aggregate of the states of its micro-constituents together with their spatial relations. A simulation, in the special sense used here, is a step-by-step process that replicates the time

development of the system at the micro-level. It is then obvious that a state identified by the possession of a weakly emergent property by some object at a time will, if discovered, be discovered in an irreducibly empirical way.

Now for the philosophical point: Despite its position firmly within what has been considered the domain of the unobservable, the constraints on knowledge about systems in the nanoscale region are frequently not the result of what empiricists consider to be unobservable features of the system. Instead, many of the constraints are computational in form, and this indicates that the empiricist's qualms about size relative to the human scale are not the primary epistemological issue here. Moreover, contrary to a widespread view that theory is clean and accurate, whereas empirical measurements are messy and less accurate, one reason why theoretical methods of discovery are replaced with semitheoretical methods is that the accuracy of theoretical predictions is often far outrun by the accuracy of experimental data, as this quote indicates: "Needed values for energy levels, transition wavelengths, ionization potentials, polarizabilities, fine and hyperfine structure splittings, transition probabilities, level lifetimes etc. can be determined experimentally for complex atoms more precisely than they can be specified using the best available theoretical methods" (Curtis 2003, 5).

Nor is it easy to improve the accuracy of the theoretical predictions. Even with the approximations and simplifications introduced by both theoretical and semitheoretical methods, calculations of the energy levels in atoms and molecules remain computationally intensive. As a result, a major constraint on the choice of which method to use is memory demands. For example, among the methods commonly used, the leapfrog method of numerical integration has the advantage of low memory requirements but has the drawback of requiring a small time step, whereas the predictor-corrector method allows larger time steps but requires a larger memory. Monte Carlo methods are computationally efficient because they do not require computing the derivatives of energies and so are used for finite temperature equilibrium properties. Methodological issues arising from the choice of representation often go deep and the trade-offs are somewhat different. Most atomistic models, within which individual atoms are the basis, are infeasible because quantum mechanical treatments of atomic bonding are too computationally intensive. So they are often replaced by models in which the material is treated as a continuum and fields are used as the basic quantities. This means that the quantities of interest are defined at every point in the body. They are easier to interpret than atomistic models and usually computationally more efficient but tend to be less accurate than atomistic models. Despite their differences, these methods of investigation, although widely employed in

nanoscience, are by no means unique to it. In contrast, self-assembling systems tend to be more characteristic of the domain, and so I shall now turn to them.

3. Self-assembly. One of the most common methods of fabricating materials at the nanolevel is by using the self-assembling properties of various materials. The basic idea behind self-assembly is that large-scale structure spontaneously emerges in a dynamic system of interacting constituents (i.e., emerges without the need for external interventions) solely in virtue of local interactions between those constituents. The emergence of this structure is not accidental in the sense that, were the system to be restored to its initial state and the micro-dynamics rerun, similar large-scale structure would, with a high degree of probability, reappear. There are many examples of self-assembling systems, and their self-assembly often involves the minimization of the total energy of the system, resulting in a stable state. Monolayers, micelles, and membranes are well-known examples. Lipid nanotubes can be self-assembled, as can protenoid microspheres,[4] colloidal rods and spheres constructed from silica gel, C_{60} buckminsterfullerenes, and fractal structures appearing during the growth of Mn oxides on surfaces. A specific molecule that self-assembles is amphiphilic [2]catenane $5.4PF_6$. Biological systems, of course, have long engaged in self-assembly. Hydra, when chopped into small fragments, have the ability for each piece to regenerate into a complete new hydra.

The literature sometimes distinguishes between the *self-assembly* of systems and their *self-organization*, the distinction being roughly that a unit is self-assembling if it can construct itself without the aid of external inputs, whereas more complex entities of that science are self-organizing if they can be constructed from the basic units using only the properties of those units.

Representative characterizations of the two concepts are: "Self-assembly deals with the construction of discrete molecular and supramolecular assemblies—via the use of weak but specific noncovalent bonding interactions—that have programmed into them a way by which the architecture and function can be controlled" (Preece and Stoddart 1995, 3), and "Self-organization deals with bringing together identical (supra) molecular assemblies to produce large highly ordered polymolecular arrays—as a direct result of noncovalent bonding interactions—that generally act in a cooperative manner. Self-organization can be thought of as 'polymerization' via noncovalent bonding interactions, resulting in well-defined nanoscale architectures. Such systems include liquid crystals, monolayers at inter-

4. These are mixtures of amino acids that when heated and wetted organize into spherical balls.

faces, Langmuir-Blodgett films, vesicles, and micells" (Preece and Stoddart 1995, 3). However, the difference between the two is based more on degrees of complexity than anything of great ontological importance, and we can use just one definition here:

> **Definition**. An entity E *self-assembles* if its components $C_1,\ldots,C_n$ contain within themselves the information, properties, and relations that are sufficient to enable E to be constructed without the aid of input other than the C_i.

Why are self-assembling systems philosophically interesting? In part it is because they are good examples of systems that are not discovered theoretically, and some may even involve irreducibly empirical discovery processes. Let's begin by asking what sort of method is self-assembly. Self-assembly is a form of experiment, but it is unlike traditional Galilean experiment in being a synthetic rather than an analytic process and in being essentially dynamic. It is to be contrasted with many other synthetic procedures such as the use of atom force microscopes to assemble materials because those require deliberate, designed intervention from outside the systems. Is it the computational intractability of theoretical methods that forces one to rely on self-assembling processes? Not quite. In part because the mechanisms of self-assembly are not fully understood, representations of self-assembling systems usually rely on a collection of approximations, simulations, and semiempirical methods to model the process of self-assembly. That places the process within the intermediate zone of semitheoretical discoveries.

We can gain a perspective on the nature of self-assembling processes by placing them within the context of Jaegwon Kim's account of reduction. Emergence is often viewed as a mysterious and rare phenomenon, if it exists at all, but one of the things that nanoscience does is to show us that emergence is in fact quite a common phenomenon. In his 1999 article "Making Sense of Emergence" (Kim 1999), Kim proposed an account of property reduction that is significantly different from the Nagelian reduction relation between theories and has the consequence that there seem to be very few candidates for irreducible properties. There are three steps in the process of Kim reduction. First, the property to be reduced has to be functionalized. Second, a realizer of the functionalized property must be identified, usually by science. Third, a theory must be found at the level of the realizers that explains how the realizers perform the causal task that is constitutive of the original functionalized property. The first of these steps is conceptual; the second and third usually require significant scientific work. Kim's account of reduction is ingenious, powerful, and a valuable addition to our repertoire of philosophical tools. Valuable as it is, I think it inherently overstates the case for reduction.

Let me explain why. Part of the claim that Kim reduction counts as a form of reduction rests on the idea that once we have the ability to give a *theoretical* explanation or prediction of a given phenomenon, that phenomenon no longer counts as emergent under emergentist positions that are essentially epistemic in orientation, that is, those that claim a phenomenon is emergent if there is no way to predict or explain that phenomenon, even in an ideal theory. Kim contrasts this kind of theoretical prediction with an inductive prediction of the occurrence of a property E on the basis of the occurrence of properties P. Inductive predictions are based upon observing a regular association between the occurrence of P and the occurrence of E.

The connection between what I said earlier about semitheoretical and irreducibly empirical discoveries on the one hand and Kim emergence on the other is this: Many self-assembled molecules have a functional description, so the first part of Kim's account is satisfied. Now, Kim requires that an explanation of how the realizer carries out the causal role be given. As he says: "Why does this system exhibit E (the emergent property) at t? Because having E is, by definition, having a property with causal role C, and the system at t has property Q, which fills causal role C (and hence realizes E). Moreover, we have a theory that explains exactly how Q manages to fill C" (Kim 1999, 13). But many self-assembling processes are not purely theoretically predictable, and some are not even semitheoretically predictable. This means that we have a self-assembled entity that has properties that are weakly emergent in Bedau's sense, and we have a theory only in a very general sense about the assembly process. This is the situation in which we often find ourselves with self-assembling molecules, and it suggests that many self-assembling systems for which no formal discovery methods exist must be considered as emergent systems.

4. Conclusion. It would be remiss of me to leave this topic without mentioning the negative side of nanoscience. One does not need to invent science fiction scenarios of nanobots running amok to have serious concerns about the potential hazards of nanoscience. Two reasons for this follow from what I have said earlier. First, I mentioned that some substances have different properties at different size scales, and there is some evidence that gold becomes biochemically active at the nanoscale. There is no reason to think that this phenomenon is uncommon. Second, the inability to discover properties of nanoscale matter in a theoretical way, and the fact that irreducibly empirical methods might have to be used suggests that our knowledge of interactions at this scale tends to be piecemeal and incomplete. This is still the state of knowledge with relation to cigarette smoke and carcinogenesis in humans. Our ability to draw upon

decades of mortality and morbidity tables in that case is a sobering thought.

REFERENCES

Bedau, Mark (2003), "Downward Causation and the Autonomy of Weak Emergence," *Principia Revista Internacional de Epistemologica* 6: 5–50.

Curtis, Lorenzo J. (2003), *Atomic Structure and Lifetimes: A Conceptual Approach.* New York: Cambridge University Press.

Humphreys, Paul (2004), *Extending Ourselves: Computational Science, and Scientific Method*. New York: Oxford University Press.

Kim, Jaegwon (1999), "Making Sense of Emergence," *Philosophical Studies* 95: 3–36.

Preece, J. E., and J. F. Stoddart (1995), "Towards Molecular and Supramolecular Devices," in M. E. Welland and J. K. Gimzewski (eds.), *Ultimate Limits of Fabrication and Measurement*. Dordrecht: Kluwer, 1–8.