Minds & Machines (2008) 18:443–459
DOI 10.1007/s11023-008-9122-6

Is Weak Emergence Just in the Mind?

Mark A. Bedau

Received: 4 December 2007 / Accepted: 5 May 2008 / Published online: 4 November 2008

Abstract Weak emergence is the view that a system's macro properties can be explained by its micro properties but only in an especially complicated way. This paper explains a version of weak emergence based on the notion of explanatory incompressibility and "crawling the causal web." Then it examines three reasons why weak emergence might be thought to be just in the mind. The first reason is based on contrasting mere epistemological emergence with a form of ontological emergence that involves irreducible downward causation. The second reason is based on the idea that attributions of emergence are always a reflection of our ignorance of non-emergent explanations. The third reason is based on the charge that complex explanations are anthropocentric. Rather than being just in the mind, weak emergence is seen to involve a distinctive kind of complex, macro-pattern in the mind-independent objective micro-causal structure that exists in nature. The paper ends by addressing two further questions. One concerns whether weak emergence applies only or mainly to computer simulations and computational systems. The other concerns the respect in which weak emergence is dynamic rather than static.

Keywords Weak emergence · Epistemological emergence · Dynamic emergence · Computational emergence · Micro-causal network · Micro-causal web · Explanatory incompressibility

M. A. Bedau (✉)
Department of Philosophy, Reed College, Vollum College Center 141, 3203 SE Woodstock Blvd., Portland, OR 97202-8199, USA
e-mail: mab@reed.edu

M. A. Bedau
European School of Molecular Medicine, Milan, Italy

The notion of emergence is enjoying a renaissance in philosophy and science today,[1] but it remains haunted by the worry that all apparent emergence in nature is really in one way or another just in the mind.[2] If emergent phenomena are just in the mind, then they are not real and objective phenomena; they have no independent ontological existence; they have no independent causal power; they have no objective reality outside the mind. The charge that emergence is just in the mind has recently been directed specifically at weak emergence, a view which I have developed and defended elsewhere,[3] and which I will explain and defend here. My main goal here is to show that weak emergence is not just in the mind. Rather, it is a distinctive kind of complex, macro-pattern in the mind-independent objective micro-causal structure that exists in nature. My secondary goal here is to answer two further questions about weak emergence. One concerns whether weak emergence is limited to computer simulations and computational systems. The other concerns how weak emergence is dynamic rather than static.

Weak emergence (as I intend it) can be explained in various ways.[4] My explanation here uses the concept of *explanatory incompressibility*. Previously I have defined weak emergence using the concept of underivability except by simulation (Bedau 1997; 2003). These two definitions are similarly indirect, and they are essentially equivalent. I support them both. Here I focus on explanatory incompressibility in order to highlight that weak emergence applies not just to computer simulations. It is true that weak emergence has a special connection with computer simulations, as we will see below, but it also applies equally well to a great many natural systems, especially those that motivated the original discussions of emergence by the British emergentists (McLaughlin 1992).

My defense of weak emergence will turn on its distinctively *dynamical* nature. In contrast to most recent discussions of emergence in philosophy, weak emergence is concerned with the complex dynamic processes by which certain global phenomena are generated. The dynamical nature of weak emergence helps explain the distinctive role that computer simulations play in both explaining natural emergent phenomena and artificially generating interesting new examples of emergent phenomena. It also helps explain why emergence is not merely in the mind. For if the dynamical causal processes distinctive of weak emergence are not merely in the mind, then neither is weak emergence itself.

1 See, e.g., Feltz et al. (2006), Kistler (2006), Bedau and Humphreys (2008).

2 Some who claim that weak emergence is just in the mind view this in a positive light (e.g., Newmann 1996 and McIntyre 1998), on the grounds that if weak emergence is just in the mind, then it surely exists and is real, and it has no objectionable metaphysical baggage. In this paper I am arguing that weak emergence is *not* just in the mind; it is real and objective in nature. Specifically, it is a certain kind of complex micro-causal network that has unpredictable macro effects. In addition, this kind of weak emergence has no metaphysical baggage. Each instance of weak emergence is entirely consistent with any reasonable form of naturalism.

3 Bedau (1997, 2003).

4 The different variants of weak emergence in the literature include those by Wimsatt (1986, 1997, 2000), Rueger (2000a, b), Boogerd et al. (2005), and Bedau (1997, 2003). Different conceptions of weak emergence focus on different kinds of explanatory complexity, but all agree that weak emergence involves some distinctive kind of explanatory complexity.

Weak Emergence as Explanatory Incompressibility

Emergence always involves a certain kind of relationship between global or macro-phenomena and local or micro-phenomena. Specifically, emergent macro-phenomena somehow both depend on, and are autonomous from, micro-phenomena. Dependence and autonomy can be given different interpretations, and different interpretations lead to different conceptions of emergence, including different conceptions of weak emergence.

The characteristic feature of weak emergence, in general, is that the macro is ontologically and causally reducible to the micro *in principle*, but the reductive micro-explanation is *especially complex*.[5] Different kinds of explanatory complexity create different versions of weak emergence.[6] Elsewhere I have characterized weak emergence as underivability without simulation.[7] Here I shift terms slightly and replace derivations with explanations, and replace macro-states that are underivable except by simulation with macro-states that have only incompressible explanations. Throughout I tacitly assume that we seek only true, exact, and complete explanations of how macro properties are generated from prior micro properties over time.

My definition of weak emergence is this, specifically: If *P* is a macro-property of some system *S*, then *P* is weakly emergent if and only if *P* is generatively explainable from all of *S*'s prior micro-facts but only in an incompressible way. This definition defines weak emergent macro-phenomena by the distinctive way in which we explain how they are generated from underlying micro-states.

The basic idea of an *incompressible* generative explanation is simple. An explanation is generative just in case it exactly and correctly explains how macro-events unfold over time, how they are generated dynamically. The temporally (discrete or continuous) generative explanation assumes complete information about both the micro-causal dynamics that drive the system and the system's earlier micro-states and boundary conditions. The explanation works simply by tracing through the temporal details in the complex web of micro-level causal interactions that ultimately generate the macro-events. This kind of explanation is appropriate for any system with (global) macro-features that depend on (local) micro-features in certain complex ways. In particular, it is appropriate if we can describe the system's macro-features *at a given time* by appropriately conjoining or aggregating or summing the (local) micro-features that exist at the same time. This is a synchronic

[5] See, e.g., Simon (1996).

[6] Unfortunately, the phrase "weak emergence" is used in different ways by different philosophers. For example, what I call "weak emergence" differs from what is called "weak emergence" by Stephan (2006) or by K. Balog (personal communication), and it does not apply to typical Braitenberg's vehicles (an example of weak emergence from A. Beckerman, personal communication).

[7] See Bedau (1997, 2003). I should note that my earlier use of the word "derivation" to define weak emergence does not imply any connection with, or dependence on, a Hempelian Deductive-Nomological account of explanation, which is famous for its central role for derivations (Hempel 1965). Rather, I am content with a much more loose and informal notion of explanation. Nevertheless, explanation and deduction still use and rely on various derivations.

reduction of macro to micro. Now, by starting with a completely specified initial condition, and by tracing forwards in time through the network of local micro-level causal interactions, the system's macro-features (which are aggregation of micro-features at a given time) can be explained from immediately preceding aggregations of micro-features. Explaining the generation of a system's macro-behavior by aggregating and iterating the earlier local micro-interactions over time I shall describe, for short, as *crawling the micro-causal web*.

Incompressible explanations cannot be replaced without explanatory loss by shorter explanations that avoid crawling the causal web. If an explanation of some macro-property of some system is incompressible, then there is no short-cut generative explanation of that macro-property that is true, complete, accurate, and can avoid crawling the causal web.[8] This explains the temporal signature of incompressible explanations: Explaining later behavior requires additional explanatory effort (Crutchfield et al. 1986).

On the other hand, if the explanation is *compressible*, then explaining the macro-property arbitrarily far into the future takes some fixed and finite amount of explanatory effort, no matter how far into the future your explanation reaches. The required explanatory effort is capped; explaining later behavior takes no more explanatory effort than explaining earlier behavior. A compressible explanation can achieve these economies because it avoids the incrementally growing cost of crawling the causal web.

Examples can vividly convey what a compressible explanation is, and the simplest and clearest examples involve cellular automata. A cellular automaton (or CA) is a regular spatial lattice of "cells," each of which can be in one of a finite number of states (e.g., "alive" or "dead"). The lattice typically has 1, 2, or 3 spatial dimensions. The state of each cell in the lattice is updated simultaneously in discrete time steps. Each cell is a finite state machine that outputs the next state of the cell given as input the states of the cells within some finite, local neighborhood of the lattice. Typically all cells in the lattice are governed by the same finite state machine, which typically is deterministic.

Now, consider the exceedingly trivial cellular automaton that I call *All Life*, which is completely governed by the following very simple local causal rule: A cell is alive at a given time whether or not it or any of its neighbors were alive or dead at the previous moment. It is trivial to give a short-cut explanation of any macro-property of All Life arbitrarily far into the future, because one can see that all cells will be alive at all future times, no matter what the initial aggregate local configuration of the cellular automaton.

All Life is atypical compared with the cellular automata usually discussed, such as the Game of Life (GOL). The most interesting CAs are those like GOL that are

[8] The explanation can apply to indeterministic systems by including complete information about the system's inderministic micro-state changes in the information from which the explanation is sought. System boundary conditions are handled in a similar fashion. See Bedau (1997). My notion of incompressible explanation is closely connected with Chaitin's notion of random sequence (Chaitin 1975, 1988) and Wolfram's subsequent notion of incompressible computation (Wolfram 1985, 2002), as well as the notion of a dynamical system that must be simulated to discover its generic properties (Crutchfield et al. 1986).

complex and known to require incompressible explanation.[9] The behavior of complex cellular automata typically cannot be explained except by crawling the causal web. The same holds throughout many other kinds of computational systems, such as soft artificial life systems like Ray's Tierra, Holland's Echo, and Packard's Bugs.[10] The more simulations of natural complex adaptive systems you study, the more weak emergence you find. This is a contingent empirical claim, but it is still true.

One might ask exactly much micro-causal complexity is sufficient for weak emergence, and exactly how much is necessary. This question makes sense if there is a bright line separating weak emergent properties from merely resultant properties, but the truth is more complicated. Weak emergence comes in degrees. Assad and Packard (1992, p. 232) describe a scale for degrees of emergence, ranging over behavior that is "immediately deducible upon inspection of the specification or rules generating it", to behavior that is "deducible in hindsight from the specification after observing the behavior", and continuing to behavior that is "deducible in theory, but its elucidation is prohibitively difficult", and finally reaching behavior that is "impossible to deduce from the specification". Explanatory incompressibility can be arrayed into similar stages. So, since weak emergence depends on explanatory incompressibility, weak emergence also comes in stages or degrees.[11] The paradigm case of weak emergence involves properties with incompressible explanations. A lower degree of emergence involves properties with compressible explanations that are so complicated that in practice no one can use the explanation except with a computer simulation. A higher degree of emergence involves properties that can be simulated but not in any finite simulation. These examples illustrate some of the sorts of ways in which weak emergence comes in degrees.[12]

My indirect definition of weak emergence applies to all systems with macro-behavior with only incompressible explanations. Now, if a system's behavior requires incompressible explanations, that is presumably because of the complexity of the system's micro-causal interactions. If we could directly identify what it is about micro-causal interactions that make them incompressible, then we might be able to construct a direct definition of weak emergence. In the meantime, experience in the field with our indirect definition suggests two conclusions about the intrinsic properties of micro-causal dynamics that require incompressible explanations. First, complex systems with weak emergent behavior typically involve massively parallel

[9] The classic reference on the Game of Life is Berlekamp et al. (2004), and Wolfram (1985, 2002) are important references on cellular automata in general. For more on emergence and cellular automata, see Bedau (1997, 2003) and the references therein.

[10] E.g., Bedau (1997) reviews the supple adaptation to the edge of disorder that emerges from Packard's Bugs model.

[11] See, also, Bedau (2003, p. 163).

[12] In this paper I will not take a stand on how to measure amounts of weak emergence. The most precise and explicit formal definition of amounts of weak emergence known to me is due to Paul Hovda (2008). Hovda defines the amount of simulation effort needed to derive something. This formalism could be interpreted as the amount of effort required for something's generative explanation.

micro-level populations of independent and autonomous agents that interact with their neighbors and their local environment. Second, the interactions among the agents and their environments are typically non-linear and synergistic, so that the behavior of an agent is highly sensitive to its local context, including other agents. These two factors make the behavior of complex systems' impossible to predict, even given complete prior micro information, short of crawling the causal web. On the other hand, crawling the micro-causal web provides an explanation that is guaranteed to be perfect, at least in principle.

The possibility of completely explaining weak emergent phenomena by crawling the causal web entails that weak emergence is consistent with reductionism. Many philosophers and scientists assume that emergence and reduction are *incompatible*. One typical form of reductionism is mereological supervenience (Kim 1978), the view that wholes are completely determined, ontologically and causally, by their parts. And it is certainly true that some kinds of emergence are incompatible with reduction; for example, strong forms of emergence are often defined in terms of reductive failure (e.g., Kim 1999). However, weak emergence differs from strong emergence because it is *consistent* with many forms of reduction.[13] To see this, consider ontological, causal, and explanatory reductions, in turn (which, respectively, involve reducing ontologies, causal relations, and explanations). Each concrete physical embodiment of weak emergence is ontologically nothing more than some kind of aggregation of smaller embodied objects. For example, the ontological substance of a traffic jam is nothing more than a certain kind of aggregation of cars on a road, and the ontological substance of a vesicle is nothing more than a certain kind of closed spherical bilayer aggregation of amphiphilic molecules in water. Furthermore, the causes and effects of each concrete instance of any kind of weak emergent macro-phenomenon are reducible to the iteration of the aggregation of the causes and effects operating at the micro-level, at least in principle. So, each example of macro-level weak emergence is ontologically and causally reducible to micro-level phenomena. However, in practice, typically nobody can understand or follow such a micro-causal reduction unless they simulate the micro-causal web on a computer, because the micro-level causal web is so complex. In a wealth of interesting cases, studied in fields like soft artificial life, computer simulations make it possible to crawl the causal web.[14]

The distinction between explanations or reductions that hold only *in principle*, versus those that also hold *in practice*, deserves further discussion. A reductive generative explanation of macro from micro might exist, in principle, but be unhelpful for explaining weak emergent phenomena, in practice, for a variety of reasons. One is that some relevant micro-level details required for the explanation might be unknown and inaccessible. Furthermore, even if all those details were known, the explanation might still be too complex and tedious for anyone to work through without the aid of something like a computer simulation. Nevertheless,

[13] A point often emphasized by Wimsatt (1986, 1997, 2000).

[14] It should be noted that in many cases we still do not know how to explain some natural regularities or patterns we seem to see in nature. One good example is the arrow of complexity in the evolution of life on Earth (Bedau, forthcoming).

given enough time and patience, anyone could work through all those micro-level explanatory details, as least in principle.[15] And working through those details is exactly what a computer simulation does. So, weak emergent phenomena always have complete and accurate explanations solely from micro-phenomena, at least in principle. These explanations rely on complete prior micro-level information, and they necessarily proceed by crawling the causal web. It is easy to see why it is typically impossible for anyone to grasp or understand how the emergent phenomena unfold from the micro, in practice, without resorting to computer simulations. This leads to the special connection between weak emergence and computer simulations, discussed below.

The distinction between explanation and reduction *in principle* and *in practice* helps explain how weak emergence fits the two hallmarks of emergence we mentioned earlier: the dependence of the macro on the micro, and the autonomy of macro from the micro. In cases of weak emergence, the macro depends on the micro because, in principle, each instance of the macro ontologically and causally is nothing more than the aggregation of micro-causal elements. For example, the ontological and causal state of a cellular automaton macro structure is nothing more than the aggregation of the ontological and causal states of its micro constituents. At the same time, weak emergence exhibits a kind of macro autonomy because of the incompressibility of the micro-causal generative explanation of the macro structure. Because the explanation is incompressible, it is useless in practice (except in so far as it serves as the basis for a good simulation of the system).

The subtle way in which weak emergence balances principles and practices is summarized with the awkward but apt notion of *in principle irreducibility in practice*. Although weak emergence phenomena have a true, complete, and exact micro-level generative explanations, at least in principle, incompressibility makes the explanations of little use, in practice. In practice, we have no alternative but to simulate the systems micro-level behavior, if we want to observe what macro behavior will emerge. This is a practical limitation, a limitation on irreducibility in practice. Furthermore, this practical limitation holds *in principle* for any naturalistic epistemic agent that is trying to explain the behavior of complex systems. We can put these points together by saying that weak emergent phenomena are in principle irreducible in practice.

As an aside, we should note that in many contexts an especially important subspecies of weak emergence is *robust* weak emergence. Weak emergence is robust when it involves causally salient law-like patterns involving weak emergent macro-properties (Bedau 2003). These robust emergent patterns recur in regular statistical patterns. Being typical or generic, they have some explanatory force. Many properties of the emergent patterns are insensitive to the details of the local micro-interactions that produce the patterns, so the emergent patterns have multiple realizations. One interesting special case of robust weak emergence are the physical systems that exhibit what physicists call "universal" behavior, especially around

[15] The computer-generated proof of the four colors theorem is one specific kind of example of a proof that one could work through in principle but not in practice.

phase transitions, such as when a solid melts into a liquid.[16] Physicists in some instances have mathematically proved that the critical behavior of some large class of physical systems is insensitive to almost all details about the system, but in most cases one has merely empirical evidence that a physical system exhibits universal behavior. Nevertheless, this empirical evidence can be very strong.[17] Since the evidence is empirical, sometimes we are wrong when we think we have strong evidence that a system's behavior is weakly emergent. But this is not a weakness in the notion of weak emergence. It is the expected consequence of all empirically justified claims.

Why Weak Emergence is Not Just in the Mind

There are a number of reasons why weak emergence might seem to be just in the mind, and I will discuss some of the most important and influential ones. I will argue that weak emergence goes beyond our minds and concerns actual objective causal relations in nature. In that sense, weak emergence is not just in the mind.

The first worry is especially simple to explain: The *existence* of in principle irreducible downward causation is an ontological matter, because it involves the real existence of a certain kind of causal process. Many people, especially in the philosophy of mind, are interested in a strong form of emergence that entails the existence of this kind of in principle irreducible downward causation. For example, Silberstein and McGeeve (1999) define "ontological emergence" as "features of systems of wholes that possess causal capacities not reducible to any of the intrinsic causal capacities of the parts nor to any of the (reducible) relations between the parts" (p. 182), and continue: "Emergent properties are properties of a system taken as a whole which exert a causal influence on the parts of the system consistent with, but distinct from, the causal capacities of the parts themselves" (p. 182). This statement illustrates how emergence sometimes is equated only with strong emergence, specifically, the sort that involves in principle irreducible downward causation.

Now, weak differs from strong emergence on precisely this point, for weak emergence *bars* in principle irreducible downward causation. Weak emergence does involve a certain kind of downward causation, and that kind of downward causation is irreducible in practice, due to explanatory incompressibility. But weakly emergent phenomena can always be given an explanation by crawling the causal web.[18] The web could always be crawled in principle, given complete information about the initial conditions and boundary conditions, and given enough time and effort. This is what computer simulations do. This difference between reducibility *in principle* and *in practice* is the difference between strong and weak emergence.

By defining their terms appropriately, Silberstein and McGeeve (1999) brand weak emergence as merely "epistemological" and not genuinely "ontological".

[16] See, e.g., Laughlin and Pines (2000), Laughlin (2006). Batterman (2002) is the first philosopher to emphasize the connection between emergence and universality in physics.

[17] See, e.g., Stanley (1971).

[18] Bedau (2003) elaborates this claim.

They define epistemological emergence to apply to any property that is "reducible to or determined by the intrinsic properties of the ultimate constituents of the objects or system" but "is very difficult for us to explain, predict or derive… on the basis of the ultimate constituents" (p. 186). A reductive consequence for macro phenomena immediately follows: "*In principle*, in such cases the higher-level feature, rule or law is a logical consequence of some lower-level feature, rule or law" (186, emphasis added). Nevertheless, micro-reductive explanations fail in practice, and "at each stage, entirely new laws, concepts and generalizations will be necessary (*though not in principle*) to explain or predict the phenomena with relative ease" (186, emphasis added).

Note that all of these properties of "epistemological" emergence apply to weak emergence. I have stressed how weak emergent macro phenomena in principle can always be explained solely from complete prior micro phenomena, by crawling the causal web. But those explanations (or reductions) exist only in principle. In practice, the explanations are so incompressible that we can explain the emergent phenomena only if we resort to computer simulations or to appeal to empirically justified macro-level patterns, regularities, or laws. So, weak emergence meets the definition of "epistemological" emergence from Silberstein and McGeever.

But this definition does not provide a reason to conclude that weak emergence is merely in the mind. It is just a definition; it has no force as an argument. Silberstein and McGeever *define* "epistemological" emergence as any form of emergence that rejects in principle irreducible downward causation, and according to that definition weak emergence is "epistemological." But it does not follow that weak emergence is *merely* in the mind or *merely* epistemological. Weak emergent phenomena might also have ontological, non-epistemological aspects. In particular, the distinctively incompressible micro-causal explanations of weak emergence presumably are due to a distinctively incompressible form of micro-causal structure in reality. It is presumably not an accident that one sort of micro-causal structure is incompressible and another sort is compressible. So, though weak emergence meets the Silberstein and McGeever definition of "epistemological" emergence, weak emergence is not *merely* epistemological. It is not *just* in the mind. Instead, weak emergence results from incompressible macro-level structure in the network of micro-level causal connections. This causal web is embodied and brought to life in real ontological substances with real causal powers, and it really generates certain macro-level ontological and causal phenomena.

Let us now turn to a second reason for thinking that emergence is merely in the mind, due to Hempel (Hempel and Oppenheim 1965). Hempel construed emergence as irreducibility of the macro from the micro given the full resources of the best scientific theories of the day. He presumed that all apparently emergent phenomena are merely apparent, and have a true, reductive and non-emergent explanation. This implies that attributions of emergence are merely admissions of our ignorance of the true, reductive and non-emergent explanation. If our best scientific theories construe certain phenomena as emergent (because in principle they are irreducible), that does not show us anything about nature. Rather, it just shows that we need a better scientific theory. You can sum up Hempel's complaint against emergence this way:

If our best scientific theories imply the existence of weak emergent phenomena, that merely reflects our ignorance of their true, non-emergent explanation.

Cellular automata provide an especially clear and simple illustration why Hempel's complaint does not apply to weak emergence. Consider the Game of Life (GOL).[19] We know the complete micro-theory for the behavior of each cell in the GOL: A cell at a given moment is alive just in case it was alive at the previous moment and had two or three living neighbors, or it was dead at the previous moment and had exactly three living neighbors. There is no ignorance whatsoever in our understanding of the rules that completely determine the micro-behavior of any call in the GOL at any time, because the behavior of every cell in the GOL is a trivial application of the GOL birth-death rule. Thus, any weak emergence that exists in the GOL is not merely the result of our ignorance of true non-emergent explanations; it is not just something we have imagined. If it exists, it is real and objective. The explanatory incompressibility in the GOL arises from the context-sensitivity of the birth-death rule; living cells arise or persist only if their immediate neighborhood contains just the right level of living cells, and neither too few nor too many. The GOL is also synergistic; the effect of the state of a given cell depends on the states of neighboring cells, and this is symmetric. In addition, the rule enables the existence of emergent macro-level causal structures, such as so-called "gliders" that propagate and interact in a family of reactions. A universal Turing machine can even be constructed in the GOL (Berlekamp et al. 2004). Gliders enable CAs to be programmed to perform various desired forms of complex parallel computations, such as density classification and synchronization (Crutchfield and Mitchell 1995; Crutchfield et al. 2003). In general, these interesting macro-behaviors of CAs cannot be explained even given the complete and accurate theory governing micro-state behavior (along with contingent information about initial micro-states and boundary conditions), except, of course, by crawling the causal web. This shows that the behavior of those CAs is weakly emergent. In the GOL we know that this kind of weak emergence is not a sign of our ignorance of some true, underlying, non-emergent explanation of the macro-behaviors. Emergent macro-behaviors have no non-emergent, compressible explanation, but they can always be explained by crawling the causal web. So, contemporary echoes of Hempel's criticism provide no reason to think that weak emergence is merely in the mind.

There is a third reason why weak emergence might seem to be merely in the mind: Weak emergence is anthropocentric and concerns a limitation in human epistemological capacities. In fact, weak emergence is defined in this paper in terms of its incompressible explanations, and elsewhere in terms of its underivability without simulation. Both definitions are indirect and identify emergent phenomena in terms of our distinctive epistemological relationship with them (the incompressibility of their explanation, and their underivability except by simulation). Since weak emergence is defined by reference to these human epistemological limitations, doesn't it follow that weak emergence is just in our minds?[20]

[19] For details about the GOL, see Bedau (1997, 2003) and the references cited therein.

[20] I might mention that the indirectness itself of these definitions of weak emergence does not make the definitions dubious. Indirect definitions can still be perfectly useful and accurate.

In truth, though, weak emergence has nothing to do with specifically human epistemological limitations. Weak emergence does not involve phenomena that are too complex for humans to explain but simple enough for smarter naturalistic epistemic agents to explain.[21] Rather, when weak emergence arises, the actual underlying local micro-causal processes are so complex that, in principle, complete and accurate explanations of macro-behavior are all incompressible. The emergent phenomena that arise from complex synergistic micro-causal explanations are explanatorily incompressible for any naturalistic epistemic agent. No matter how fast and infallibly inferences are made, no matter how perfect memories remain, naturalistic epistemic agents trying to explain weak emergent phenomena can produce only incompressible explanations. The practical limitations of explanations of weak emergence apply in principle to any epistemic agent; this is in principle inexplicability in practice.

If something has an indirect epistemological definition, it does not follow that it is just in the mind. Instead, the indirect epistemological definition is produced by and reflects a distinctive underlying ontological status or structure in nature. Incompressibility of explanations is a consequence of the objective complexity of the local micro-causal interactions that are ultimately generating the emergent behavior being explained. The micro-causal web is real and objective, and the incompressible causal pathways of weak emergent phenomena have a distinctive epistemological consequence. Note that the explanatory incompressibility that defines weak emergence applies to the explanations of any naturalistic epistemic agent, in principle. Just like us, any non-human epistemic agent will have to work through the objective complexity of the local micro-causal interactions. Thus, weak emergence is not merely in the mind, but refers to objective complexity in the objective natural world that is in principle irreducible in practice.

We can now turn to two further questions about weak emergence: in what way weak emergence applies only or mainly to computer simulations, and in what way weak emergence is inherently dynamic rather than static. It turns out that answering these questions has a connection with why weak emergence is not just in the mind.

Computer Simulations and Weak Emergence

Weak emergence has an especially close connection with computer simulations and computational systems, but the link is sometimes misunderstood. Some of the best examples of weak emergence come from computer simulations, and my previous definitions of weak emergence rely centrally on the notion of "underivability except by simulation" (Bedau 1997, 2003).[22] Someone might incautiously infer that weak

[21] By "naturalistic epistemic agent" I mean one with no magical abilities, such as an infinite amount of storage space.

[22] See also Humphreys's (2007a) discussion of computational emergence.

emergence applies merely or primarily to computer simulations or other kinds of computational systems. But that would be a mistake.

In truth, explanatory incompressibility typifies the behavior of a great many natural systems. For example, macro-level traffic jams are composed of a loose and changing group of individual micro-level automobiles, and traffic jams exhibit interesting macro-behavior. For example, jams suddenly spontaneously form when the traffic density crosses a critical value, and jams move slowly backwards in the traffic flow. It is easy to explain these macro-behaviors by iterating and aggregating all the simple local interactions among individual vehicles (Nagel and Rasmussen 1996; Sugiyama et al. 2008), but as far as anyone knows it is impossible to give any short-cut explanation of this behavior from complete information about micro-states and boundary conditions.

The same holds for the behavior of many natural chemical processes, such as the self-assembly, growth, and subsequent division of vesicles formed from amphiphiles in appropriate aqueous solutions (Hanczyc et al. 2003). Likewise, explanatory incompressibility seems to characterize a vast number of global properties of complex systems in molecular and cellular biology, including regulatory gene networks, metabolic networks, and the process by which proteins fold into three-dimensional structures. The same can be said for many systems studied by psychology and the social sciences. So, according to our best current explanations of complex systems, weak emergence applies throughout nature. It is not limited only or mainly to computational systems and computer simulations.

At the same time, one must acknowledge that computer simulations and computational systems have two important roles in helping us to understand weak emergence. First, certain computational systems produce some of the most striking *examples* of weak emergent phenomena. Computer simulations of complex, non-linear, dynamical, hierarchical systems in nature comprise one class of computational embodiment of weak emergence. Another class consists of complex computational systems that are not simulations of something else but are studied in their own right. Both kinds of computational systems have a massively parallel architecture with non-linear local interactions. Cellular automata and other software systems studied in artificial life provide plenty of good examples. Those familiar with these computational systems know that the global patterns they produce comprise many interesting and vivid examples of weak emergent phenomena.

Note that these patterns and regularities produced by computational systems are not mere simulations of emergent phenomena. Rather, they are computational embodiments of real emergent phenomena. That is, the computer produces something that is weakly emergent in its own right. If the computer happens to be simulating some natural system, that natural system might also exhibit its own emergent phenomena. Further, if the simulation is accurate in the relevant respects, it might explain why and how the natural system's phenomena is weakly emergent. But the computer simulation itself, considered as an object it its own right, is also exhibiting emergent behavior.

This points to a distinctive role that computer simulations play in our *evidence for* weak emergence in complex natural systems.[23] We typically study the behavior of complex systems by computer simulations, because we typically have no practical alternative. It is no accident that computer simulations fill the study of complex natural systems in virtually all disciplines. Computer simulations provide our only useful evidence about how complex systems will behave, about what global patterns emerge from their myriad micro-interactions.[24] That is why it is possible to define weak emergence as that which is "underivable except by simulation" (Bedau 1997, 2003).

The evidence for weak emergence provided by computer simulations, like other empirical evidence, can be misleading. So our beliefs about which systems are weakly emergent can be mistaken, when empirical evidence leads us astray. If we discover that there is a compressible explanation for some complex behavior that we thought was weakly emergent, this is not a flaw with the notion of weak emergence. It shows merely that we were wrong about an example of weak emergence. The possibility of this kind of error is an expected consequence of the empirical nature of the evidence for emergence in simulations.

The indirect epistemological role of computer simulations in explaining weak emergence might fuel a revival of the belief that weak emergence is in some sense merely epistemological. But this would be a mistake. The weak emergence exhibited in jamming traffic and dividing vesicles is not merely epistemological. Traffic jams and vesicles require incompressible explanations because of their objective, intrinsic micro-causal complexity. Traffic jams and vesicles are not just in the mind.

[23] Different kinds of computational systems have been called "simulations" so I should clarify what I mean. The simulations I have in mind are those that crawl the causal web (recall above) and generate global properties out of myriad local interactions. In addition to cellular automata, so-called "agent-based" models are good examples of simulations that crawl the causal web. They explicitly describe how local causal processes unfold over time, and global properties are merely certain kinds of aggregations of local properties.

[24] A tangential issue arises here: How can we tell if a computer model corresponds to reality, especially if the model is much simpler than the natural system being studied? This complex issue is beyond the scope of this footnote, but I would like to mention one point—that some computer simulations aim to explain only certain very general and robust global patterns and regularities in the behavior of certain complex systems. They do not attempt to explain the system's detailed behavior. Further, sometimes a complex system's robust global patterns and regularities are due to relatively simple and abstract features of the system; many of the details about the system do not materially affect its robust global behavior. In these cases, a very simple and abstract model can adequately explain the system's robust behavior.

One example might be Shelling's famous simple models of social phenomena such as segregated neighborhoods (Schelling 1968). These models abstract away from almost all the details of actual social neighborhoods. But they preserve certain key property—such as each agent's awareness of the social class of its local neighbours and itself, and its preference for local neighbors of the same social class—and the models explain how global segregation can result merely from those simple facts. Furthermore, you can empirically test whether people actually do know the social classes of their immediate neighborhoods, so even simple models can be empirically grounded.

Weak Emergence and the Dynamics of Causal Processes

Weak emergence has a distinctive *dynamical* nature.[25] Most recent philosophical discussions of emergence concern the static, synchronic relationships between different kinds of instantaneous phenomena or states (e.g., McLaughlin 1997, Kim 1999). The canonical example concerns someone's mental state (or some aspect of it) at some moment is thought to emerge from the person's brain state *at the same moment*. That kind of emergence is static, because it concerns states and conditions that are synchronic, i.e., that all happen at the same moment. By contrast, weak emergence concerns the processes by which certain global phenomena are generated from an aggregation of local phenomena. These generative processes are essentially dynamic. They happen over time, and are caused by local interactions. So, something is weakly emergent at a given moment not merely because of the aggregation of its local states at that time, but rather because of how its global states arose from previous aggregations of previous local states. That is, weak emergence is a historical property; it requires having a certain kind of causal history.

We noted above that the micro-level causal processes that underlie weak emergence form a large, context sensitive, massively parallel network of local causal interactions. Many of those interactions are nonlinear. In general, the only way to explain the system's eventual global (macro) behavior from the sum of its local (micro) behavior is to crawl the causal web. Starting with a completely specified initial micro-condition, one propagates the local micro influences forward in time, and then aggregates the local micro results into the global state of the system. You can think of this as "deriving" the system's later global states from its earlier micro states. These complex causal networks require a distinctive incompressible kind of explanation. The local causal networks really involve a large number of components that interact locally and nonlinearly, sometimes in a heterogeneous variety of ways. The net effect is that the system's global behavior is unpredictable, except by crawling the causal web.

Complex micro-causal webs exhibit different kinds of dynamical behavior. Sometimes the causal dynamics produces a global equilibrium state that is essentially constant and fixed over time. Sometimes the causal dynamics produce a chaotically changing sequence of global states, which remain far from equilibrium and behave unpredictably. Sometimes weak emergent global dynamics are robust, indicating a generic statistical regularity in the dynamical emergent behavior of a class of complex systems.

The momentary, static global state of a system exhibiting weak emergence is trivially derivable from the system's local state at that same moment. For the global static state is nothing more than simply the sum or aggregation of all the local states at that time. Return to our earlier examples of weak emergence: the formation and behavior of traffic jams, and the spontaneous self-assembly, growth and division of vesicles. Note that both are dynamic causal processes that unfold in time. The causal processes that generate traffic jams and cause vesicles to grow and divide are not

[25] Rueger (2000b) and Humphreys (2007a, 2007b) stand out in contemporary philosophical discussions for their focus on dynamic forms of emergence.

just in the mind. They really exist and really involve a complex network of causal relations among micro-level entities. Nature is full of this sort of complex causal systems, especially those parts of nature that are alive, or have a mind, or involve social relations and technology. In this way, the dynamic nature of weak emergence underscores why weak emergence is not merely in the mind.

Conclusions

This paper has defended a form of weak emergence that is based on the notion of explanatory incompressibility. This weak emergence is clearly metaphysically innocent and consistent with any reasonable form of naturalism, for it rests entirely on merely the dynamical micro-causal processes that underlie and generate complex phenomena. These dynamical causal processes occur not just in computer simulations or computer systems; they also occur in a vast number of the complex systems in the natural world.

So, weak emergence is not just in our minds. It concerns not just how we explain things. It is produced not merely by human mental, explanatory, or epistemic limitations. It does not mask ignorance of true non-emergent explanations. Rather, weak emergence is an objective phenomenon that exists in nature. Any naturalistic epistemic agent who tries to explain it will have to use incompressible explanations. Weak emergence is the macro-level mark of incompressible complexity in a network of micro-causal interactions. When the objective micro-causal web is sufficiently complex, all explanations of its macro-behavior are incompressible. The resulting weak emergence is not just in the mind

Acknowledgement I am grateful for helpful comments from Kellyn Bardeen, Paul Humphreys, and Philippe Huneman.

References

Assad, A., & Packard, N. H. (1992). Emergence. In M. A. Bedau & P. Humphreys (Eds.). (2008). *Emergence: Contemporary readings in philosophy and science* (pp. 231–234). Cambridge: MIT Press; page citations refer to this printing. Originally appeared as Sect. 2 of “Emergent colonization in an artificial ecology”. In F. Varela & P. Bourgine (Eds.), *Towards a practice of autonomous systems: Proceedings of the First European Conference on Artificial Life* (pp. 143–152). Cambridge, MA: The MIT Press.

Batterman, R. (2002). *The devil in the details*. New York: Oxford University Press.

Bedau, M. A. (1997). Weak emergence. *Philosophical Perspectives* 11, 375–399.

Bedau, M. A. (2003). Downward causation and autonomy in weak emergence. *Principia Revista Internacional de Epistemologica, 6*, 5–50. (Reprinted in M. A. Bedau & P. Humphreys (Eds.). (2008). *Emergence: Contemporary readings in philosophy and science*. Cambridge: MIT Press, page citations refer to this printing).

Bedau, M. A. (Forthcoming). The evolution of complexity. In T. Pradeu, et al. (Eds.), *Mapping the future of biology: Evolving concepts and theories*. Heidelberg: Springer.

Bedau, M. A., & Humphreys, P. (Eds.). (2008). *Emergence: Contemporary readings in philosophy and science*. Cambridge: MIT Press.

Berlekamp, E. R., Conway, J. H., & Guy, R. K. (2004). What is life? Chapter 25 in *Winning ways for your mathematical plays* (2nd ed., Vol. 4, pp. 927–961). Wellseley, MA: AK Peters.

Boogerd, F. C., Bruggeman, F. J., Richardson, R. C., Stephan, A., & Westerhoff, H. V. (2005). Emergence and its place in nature: A case study of biochemical networks. *Synthese, 145*, 131–164.
Chaitin, G. J. (1975, May). Randomness and mathematical proof. *Scientific American, 232*, 47–53.
Chaitin, G. J. (1988, July). Randomness in arithmetic. *Scientific American, 259*, 80–85.
Crutchfield, J. P., Farmer, J. D., Packard, N. H., & Shaw, R. S. (1986). Chaos. *Scientific American, 255*, 46–57.
Crutchfield, J. P., & Mitchell, M. (1995). The evolution of emergent computation. *Proceedings of the National Academy of Sciences USA, 92*, 10742–10746.
Crutchfield, J. P., Mitchell, M., & Das, R. (2003). Evolutionary design of collective computation in cellular automata. In J. P. Crutchfield, & P. K. Schuster (Eds.), *Evolutionary dynamics—exploring the interplay of selection, neutrality, accident, and function* (pp. 361–411). New York: Oxford University Press.
Feltz, B., Crommelinck, M., & Goujon, P. (Eds.). (2006). *Self-organization and emergence in life sciences*. Heidelberg: Springer.
Hanczyc, M. M., Fujikawa, S. M., & Szostak, J. W. (2003). Experimental models of primitive cellular components: Encapsulation, growth, and division. *Science, 302*, 618–622.
Hempel, C. (1965). Aspects of scientific explanation. In C. Hempel (Ed.), *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331–496). New York: Free Press.
Hempel, C., & Oppenheim, P. (1965). On the idea of emergence. In C. Hempel (Ed.), *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 258–264). New York: Free Press.
Hovda, P. (2008). Quantifying weak emergence. *Minds and Machines* (this issue). doi:10.1007/s11023-008-9123-5
Humphreys, P. (2007a). Computational and conceptual emergence (preprint).
Humphreys, P. (2007b). Pattern emergence (preprint).
Kim, J. (1978). Supervenience and nomological incommensurables. *American Philosophical Quarterly, 15*, 149–156.
Kim, J. (1999). Making sense of emergence. *Philosophical Studies, 95*, 3–36.
Kistler, M. (Ed.). (2006). New perspectives on reduction and emergence in physics, biology, and psychology. *Synthese (Special Issue), 151*(3).
Laughlin, R. (2006). *A different universe: Reinventing physics from the bottom down*. New York: Basic Books.
Laughlin, R., & Pines, D. (2000). The theory of everything. *Proceedings of the National Academy of Science USA, 97*, 28–31.
McIntyre, L. (1998). Complexity: A philosopher's reflections. *Complexity, 3*, 26–32.
McLaughlin, B. P. (1992). The rise and fall of British Emergentism. In A. Beckerman, H. Flohr, & J. Kim (Eds.), *Emergence or reduction? Essays on the prospects of nonreductive physicalism* (pp. 49–93). Berlin: Walter de Gruyter.
McLaughlin, B. P. (1997). Emergence and supervenience. *Intellectica, 25*, 25–43.
Nagel, K., & Rasmussen, S. (1996). Particle hopping models and traffic flow theory. *Physical Review E, 53*, 4655–4672.
Newmann, D. V. (1996). Emergence and strange attractors. *Philosophy of Science, 63*, 245–261.
Rueger, A. (2000a). Robust supervenience and emergence. *Philosophy of Science, 67*, 466–489.
Rueger, A. (2000b). Physical emergence, diachronic and synchronic. *Synthese, 124*, 297–322.
Schelling, T. (1968). *Micromotives and macrobehavior*. New York: Norton.
Silberstein, M., & McGeeve, J. (1999). The search for ontological emergence. *Philosophical Quarterly, 49*, 182–200.
Simon, H. (1996). Alternative views of complexity. In his *The sciences of the artificial* (3rd ed.). Cambridge: MIT Press.
Stanley, H. E. (1971). *Introduction to phase transitions and critical phenomena*. New York: Oxford University Press.
Stephan, A. (2006). The dual role of 'emergence' in the philosophy of mind and in cognitive science. *Synthese, 151*, 485–498.
Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Tadaki, S., & Nakayama, A. (2008). Traffic jams without bottlenecks—experimental evidence for the physical mechanisms of the formation of a jam. *New Journal of Physics, 10*, 033001. doi:10.1088/1367-2630/10/3/033001
Wimsatt, W. C. (1986). Forms of aggregativity. In A. Donagan, A. N. Perovich Jr., & M. V. Wedin (Eds.), *Human nature and natural knowledge* (pp. 259–291). Dordrecht: Reidel.

Wimsatt, W. C. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science, 64*, S372–S384.

Wimsatt, W. C. (2000). Emergence as nonaggregativity and the biases of reductionisms. *Foundations of Science, 5*, 269–297.

Wolfram, S. (1985). Undecidability and intractability in theoretical physics. *Physical Review Letters, 54*, 735–738.

Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.