DOWNWARD CAUSATION AND THE AUTONOMY OF WEAK EMERGENCE

MARK BEDAU
Reed College

Abstract

Weak emergence has been offered as an explication of the ubiquitous notion of emergence used in complexity science. After outlining the problem of emergence and comparing weak emergence with the two other main objectivist approaches to emergence, this paper explains a version of weak emergence and illustrates it with cellular automata. Then it explains the sort of downward causation and explanatory autonomy involved in weak emergence.

1. The problem of emergence

Emergence is a perennial philosophical problem. Apparent emergent phenomena are quite common, especially in the subjects treated by biology and psychology, but emergent phenomena also seem metaphysically objectionable. Some of these objections can be traced to the autonomy and downward causation that are distinctive of emergent phenomena. Emergence is receiving renewed attention today, in part because the notion repeatedly arises in certain contemporary approaches to understanding complex biological and psychological systems. I have in mind such approaches as neural networks, dynamical systems theory, and agent-based models—what for simplicity I'll call *complexity science*. For anyone interested in understanding emergence, two things about complexity science are striking. First, it aims to explain exactly those natural phenomena that seem to involve emergence, the range of phenomena covered by complexity science are about as broad as the examples of apparent emergence in nature.

© *Principia*, 6(1) (2002), pp. 5–50. Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

Second, the models in complexity science are typically described as emergent, so much so that one could fairly call the whole enterprise the science of emergence (e g , Holland 1998, Kauffman 1995) A good strategy, then, for understanding emergence is to turn to complexity science for guidance A few years ago I introduced the notion of weak emergence to capture the sort of emergence involved in this scientific work (Bedau 1997) This paper expands on that project

There are a variety of notions of emergence, and they are contested We can provide some order to this controversy by distinguishing two hallmarks of how macro-level emergent phenomena are related to their micro-level bases

(1) Emergent phenomena are *dependent* on underlying processes
(2) Emergent phenomena are *autonomous* from underlying processes

These two hallmarks are vague There are many ways in which phenomena might be dependent on underlying processes, and there are also many ways in which phenomena might be autonomous from underlying processes Any way of simultaneously meeting both hallmarks is a candidate notion of emergence The hallmarks structure and unify these various notions and provide a framework for comparing them

Taken together, the two hallmarks explain the controversy over emergence, for viewing macro phenomena as both dependent on and autonomous from their micro bases seems metaphysically problematic inconsistent or illegitimate or unacceptably mysterious It is like viewing something as both transparent and opaque The problem of emergence is to explain or explain away this apparent metaphysical unacceptability

We should not assume that there is just one solution to the problem of emergence Some philosophers search for the one true account of emergence and for the one correct solution to the problem of emergence, but that is not my goal For one thing, while the two hallmarks set boundary conditions on notions of emergence, different notions may fit this bill in different ways So different concepts of emergence might provide different useful perspectives on the problem of emergence Capturing a distinctive feature of the phenom-

ena explained by complexity science is the utility of my preferred notion of emergence Furthermore, I doubt that there is a single, specific, useful, pre-theoretical concept of emergence, so traditional conceptual analysis is of questionable value in this context Defining a metaphysically acceptable and scientifically useful notion of emergence might involve inventing new concepts that revise our view of the world My project is open to what Peter Strawson termed "revisionary" rather than "descriptive" metaphysics (Strawson 1963)

The problem has two main kinds of solutions One concludes that emergence has no legitimate place in our understanding of the real world This strategy construes apparent emergent phenomena as misleading appearances to be explained away The other strategy treats apparent emergent phenomena as genuine Success with the second strategy requires explicating a precise notion of emergence, showing that it applies to apparent emergent phenomena, and then explaining away the appearance of problematic metaphysics I defend a version of this second strategy

The proper application of the term "emergence" is controversial Does it apply properly to properties, objects, behavior, phenomena, laws, whole systems, something else? My answer is pluralistic, I think we can apply the term in all these ways and more Being alive, for example, might be an emergent property, an organism might be an emergent entity, and the mental life of an organism might be an emergent phenomenon These different subjects of emergence can be related in a straightforward way—for example, an entity with an emergent property is an emergent entity and an emergent phenomenon involves an emergent entity possessing an emergent property—and they all can be traced back to the notion of an emergent property So I will first explain the notion of an emergent property, and then extend the notion of emergence to other contexts This will allow me to talk of emergent properties, entities, phenomena, etc , as the context suggests

Before explaining my preferred notion of emergence, I will sketch a broader canvas containing different kinds of emergence Then I will explain my notion of weak emergence and illustrate it with cellular automata—a typical kind of system studied in complexity science [1] Finally, I will examine downward causation and autonomy in the con-

text of weak emergence—two connected problems that tend to pull in opposite directions In the end we will see that weak emergence avoids the problems of downward causation, and that a certain kind of robust weak emergence has an interesting metaphysical autonomy I conclude that this robust weak emergence is philosophically acceptable and scientifically illuminating, it is all the emergence to which we are now entitled

2. Three kinds of emergence

It is useful to distinguish three kinds of emergence nominal, weak, and strong [2] These are not narrow definitions but broad conceptions each of which contains many different instances My classification is not exhaustive It ignores some views about emergence, such as the view that attributes emergence on the subjective basis of observer surprise (Ronald et al 1999) The classification's utility is that it captures three main objectivist approaches to emergence Emphasizing the underlying similarities within each view and the differences between the contrasting views highlights the strengths and weaknesses of the view that I will defend

The classification of kinds of emergence assumes a distinction between a micro level and a macro level, and the issue is to specify what it is for the macro to emerge from the micro We might be interested in how an individual cell in an organism emerges out of various biomolecules and their chemical interactions, or we might be interested in how an organism emerges out of various cells and their biological interactions As this example shows, a macro level in one context might be a micro level in another, the macro/micro distinction is context dependent and shifts with our interests In addition, a nested hierarchy of successively greater macro levels gives rise to multiple levels of emergence Any final theory of emergence must clarify what such levels are and how they are related

Macro entities and micro entities each have various kinds of properties Some of the kinds of properties that characterize a macro entity can also apply to its micro constituents, others cannot For example, consider micelles These are clusters of amphiphilic poly-

mers arranged in such a way that the polymers' hydro-phillic ends are on the outside and their hydro-phobic tails are on the inside Those polymers are themselves composed out of hydro-phyllic and –phobic monomers In this context, the micelles are macro objects, while the individual monomeric molecules are micro objects [3] The micelles and the monomers both have certain kinds of physical properties in common (having a location, mass, etc) By contrast, some of the properties of micelles (such as their permeability) are the kind of properties that monomers simply cannot possess Here is another example The constituent molecules in a cup of water, considered individually, cannot have properties like fluidity or transparency, though these properties do apply to the whole cup of water.

This contrast illustrates a core component of all three kinds of emergence the notion of a kind of property that can be possessed by macro objects but cannot be possessed by micro objects The simplest and barest notion of an emergent property, which I term mere *nominal emergence*, is simply this notion of a macro property that is the kind of property that cannot be a micro property Nominal emergence has been emphasized by Harré (1985) and Baas (1994), among others It should be noted that the notion of nominal emergence does not *explain* which properties apply to wholes and not to their parts Rather, it *assumes* we can already identify those properties, and it simply terms them nominally emergent Full understanding of nominal emergence would require a general theory of when macro entities have a new kind of property that their constituents cannot have

Nominal emergence easily explains the two hallmarks of emergence Macro-level emergent phenomena are dependent on micro-level phenomena in the straightforward sense that wholes are dependent on their constituents, and emergent phenomena are autonomous from underlying phenomena in the straightforward sense that emergent properties do not apply to the underlying entities When dependence and autonomy are understood in these ways, there is no problem in seeing how emergent phenomena could simultaneously be both dependent on and autonomous from their underlying bases

The notion of nominal emergence is very broad It applies to a large number of intuitive examples of emergent phenomena and cor-

responds to the compelling picture of reality consisting of a hierarchy of levels Its breadth is its greatest weakness, though, for it applies to all macro-level properties that are not possessed by micro-level entities Macro-properties are traditionally classified into two kinds genuine emergent properties and mere "resultant" properties, where resultant properties are those that can be predicted and explained from the properties of the components For example, a circle consists of a collection of points, and the individual points have no shape So being a circle is a property of a "whole" but not its constituent "parts"—that is, it is a nominal emergent property However, if you know that all the points in a geometrical figure are equidistant from a given point, then you can derive that the figure is a circle So being a circle is a resultant property To distinguish emergent from resultant properties one must turn to more restricted kinds of emergence The two more restricted kinds of emergence simply add further conditions to nominal emergence [4]

The most stringent conception of emergence, which I call *strong emergence*, adds the requirement that emergent properties are supervenient properties with irreducible causal powers [5] These macro-causal powers have effects at both macro and micro levels, and macro-to-micro effects are termed "downward" causation We saw above that micro determination of the macro is one of the hallmarks of emergence, and supervenience is a popular contemporary interpretation of this determination Supervenience explains the sense in which emergent properties depend on their underlying bases, and irreducible macro-causal power explains the sense in which they are autonomous from their underlying bases These irreducible causal powers give emergent properties the dramatic form of ontological novelty that many people associate with the most puzzling kinds of emergent phenomena, such as qualia and consciousness In fact, most of the contemporary interest in strong emergence (e g , O'Conner 1994, Kim 1992, 1997, 1999, Chalmers 1996) arises out of concerns to account for those aspects of mental life like the qualitative aspects of consciousness that most resist reductionistic analysis

The supervenient causal powers that characterize strong emergence are the source of its most pressing problems One problem is the so-called "exclusion" argument emphasized by Kim (1992, 1997,

1999) This is the worry that emergent macro-causal powers would compete with micro-causal powers for causal influence over micro events, and that the more fundamental micro-causal powers would always win this competition I will examine downward emergent causation at length later in this paper The exclusion argument aside, the very notion of strong emergent causal powers is problematic to some people By definition, such causal powers cannot be explained in terms of the aggregation of the micro-level potentialities, they are primitive or "brute" natural powers that arise inexplicably with the existence of certain macro-level entities This contravenes *causal fundamentalism*—the idea that macro causal powers supervene on and are determined by micro causal powers, that is, the doctrine that "the macro is the way it is in virtue of how things are at the micro" (Jackson and Pettit 1992, p 5) Many naturalistically inclined philosophers (e g , Jackson and Pettit) find causal fundamentalism compelling, so they would accordingly be skeptical about any form of emergence that contravenes causal fundamentalism Still, causal fundamentalism is not a necessary truth, and strong emergence should be embraced if it has compelling enough supporting evidence But this is where the final problem with strong emergence arises All the evidence today suggests that strong emergence is scientifically irrelevant Virtually all attempts to provide scientific evidence for strong emergence focus on one isolated moribund example Sperry's explanation of consciousness from over thirty years ago (e g , Sperry 1969) There is no evidence that strong emergence plays any role in contemporary science The scientific irrelevance of strong emergence is easy to understand, given that strong emergent causal powers must be brute natural phenomena Even if there were such causal powers, they could at best play a primitive role in science Strong emergence starts where scientific explanation ends

Poised between nominal and strong emergence is an intermediate notion, which I call *weak emergence*[6] It involves more than mere nominal emergence but less than strong emergence Something could fail to exhibit weak emergence in two different ways either by being merely resultant or by being strongly emergent Weak emergence refers to the aggregate global behavior of certain systems The system's global behavior derives just from the operation of micro-

level processes, but the micro-level interactions are interwoven in such a complicated network that the global behavior has no simple explanation The central idea behind weak emergence is that emergent causal powers can be derived from micro-level information but only in a certain complex way As Herbert Simon puts it, "given the properties of the parts and the laws of their interaction, it is not a trivial matter to infer the properties of the whole" (1996, p 184) In contrast with strong emergence, weak emergent causal powers can be explained from the causal powers of micro-level components, so weak and strong emergence are mutually exclusive In contrast with mere nominal emergence, those explanations must be of a certain complicated sort, if the explanation is too simple, the properties will be merely resultant rather than weakly emergent Weak emergence is a proper subset of nominal emergence, and there are different specifications of the special conditions involved (e g , Wimsatt 1986, 1997, Newman 1996, Bedau 1997, Rueger 2000)

The strengths and weaknesses of weak emergence are both due to the fact that weak emergent phenomena can be derived from full knowledge of the micro facts Weak emergence attributes the apparent underivability of emergent phenomena to the complex consequences of myriad non-linear and context-dependent micro-level interactions These are exactly the kind of micro-level interactions at work in natural systems that exhibit apparent emergent phenomena, so weak emergence has a natural explanation for these apparent emergent phenomena Weak emergence also has a simple explanation for the two hallmarks of emergence Weakly emergent macro phenomena clearly depend on their underlying micro phenomena So weak emergent phenomena are *ontologically* dependent on and reducible to micro phenomena, their existence consists in nothing more than the coordinated existence of certain micro phenomena Furthermore, weakly emergent causal powers can be explained by means of the composition of context-dependent micro causal powers So weakly emergent phenomena are also *causally* dependent on and reducible to their underlying phenomena, weak emergence presumes causal fundamentalism (More on this below) At the same time, weakly emergent macro phenomena are autonomous in the sense that they can be derived only in a certain non-trivial way

In other words, they have *explanatory* autonomy and irreducibility, due to the complex way in which the iteration and aggregation of context-dependent micro interactions generate the macro phenomena (Section 6 develops the ramifications of distinguishing two forms of this explanatory autonomy) There is nothing metaphysically illegitimate about combining this explanatory autonomy (irreducibility) with ontological and causal dependence (reducibility), so weak emergence dissolves the problem of emergence

Some apparent emergent macro phenomena like consciousness still resist micro explanation, even in principle This might reflect just our ignorance, but another possibility is that these phenomena are strongly emergent The scope of weak emergence is limited to what has a micro-level derivation (of a certain complex sort) So those who hope that emergence will account for irreducible phenomena will find weak emergence unsatisfying

My project in this paper is to develop and defend a version of weak emergence that is ubiquitous in complexity science My main aim is to explain how it avoids the problems of downward causation and how it can involve metaphysical autonomy My arguments may generalize (with some modifications) to other versions of weak emergence, but I will not explore those generalizations here because I think my preferred notion of weak emergence has the greatest general utility in understanding emergence in nature

3. Weak emergence as underivability except by simulation

For ease of exposition, I will first explain weak emergence in a certain simple context and then extend it more broadly Assume that some system has micro and macro entities Assume also that all the macro entities consist of nothing more than appropriate kinds of micro entities appropriately configured and arranged (The micro entities might be constituted by entities at a yet lower level, but we can ignore that here) All of the ultimate constituents of any macro entity are simply micro entities, macro entities are ontologically dependent on and reducible to micro entities The system's

micro and macro entities have various kinds of properties Some of the macro properties might be nominally emergent, i e , not the kind of property found at the micro level Nevertheless, we assume that all the macro properties are structural properties, in the sense that they are constituted by micro entities possessing appropriate micro-level properties That is, a macro entity has a macro property only in so far as its constituent micro entities have an appropriate structure (are appropriately related to each other) and have the appropriate micro properties The state of a micro entity consists of its location and its possession of intrinsic properties, and its state changes if these change A macro entity also has a state, and this consists simply in the aggregation of the states of all its component micro entities and their spatial relations The fundamental micro-level causal dynamics of the system—its "physics"—is captured in a set of explicit rules for how the state of a micro entity changes as a function of its current state and the current states of its local neighboring entities Macro entities and their states are wholly constituted by the states and locations of their constituent micro entities, so the causal dynamics involving macro objects is wholly determined by the underlying micro dynamics Thus, causal fundamentalism reigns in such a system, macro causal powers are wholly constituted and determined by micro causal powers The micro dynamics is context sensitive since a micro entity's state depends on the states of its micro-level neighbors The context sensitivity of the system's underlying causal dynamics entails that understanding how a micro entity behaves in isolation or in certain simple contexts does not enable one to understand how that entity will behave in all contexts, especially those that are more complicated *Locally reducible* systems are those that meet all the conditions spelled out in this paragraph

The notion of weak emergence concerns the way in which a system's micro facts determine its macro facts A system's micro facts at a given time consist of its micro dynamic and the states and locations of all its micro elements at that time If the system is open, then its micro facts include the flux of micro entities that enter or leave the system at that time Its micro facts also include the micro-level accidents at that time, if the system's micro dynamics is nondeterministic Since causal fundamentalism applies to locally reducible systems, the

micro facts in such systems determine the system's subsequent evolution at all levels Given all the system's micro facts, an explicit simulation could step through the changes of state and location of each micro element in the system, mirroring the system's micro-level causal dynamics Since macro entities and states are constituted by the locations and states of their constituent micro entities, this explicit simulation would reflect the evolution over time of the system's macro facts Such an explicit simulation amounts to a special kind of *derivation* of the system's macro properties from its micro facts It is an especially "long-winded" derivation because it mirrors each individual step in the system's micro-level causal dynamics A locally reducible system's macro properties are always derivable from the micro facts by a simulation However, in some situations it is possible to construct a quite different "short-cut" derivation of a system's macro properties, perhaps using a simple mathematical formula for the evolution of certain macro properties arbitrarily far into the future Such short-cut derivations are the bread and butter of conventional scientific explanations They reveal the future behavior of a system without explicitly simulating it

It is now easy to define weak emergence 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 Weak emergence also applies to systems that are not locally reducible, when they contain locally reducible subsystems that exhibit weak emergence [7] Notice that the notion of weak emergence is relative to a choice of macro and micro levels A macro property could be weakly emergent with respect to one micro level but not with respect to another (although in my experience this is just an abstract possibility) It is usually obvious which levels are appropriate to choose in each context, so I will usually leave this implicit

My goal here is not a complete account of weak emergence but just an analysis of some paradigmatic cases It is natural to extend in various ways the core notion of an emergent property exhibited by a system given complete micro facts Note that the core definition allows a given property to be weakly emergent in one context with one set of micro facts, but not weakly emergent in another con-

text with different micro facts Abstracting away from any particular context, one could define the notion of an emergent property in a system as a kind of property that is emergent in that system in some context It is natural to think of certain macro objects or entities as emergent, and the natural way to define these is as objects with some weak emergent property [8] A weak emergent phenomenon can be defined as a phenomenon that involves emergent properties or objects, and a weak emergent system can be defined as one that exhibits some weak emergent phenomenon, object, or property A weak emergent law could be defined as a law about weak emergent systems, phenomena, objects, or properties The notion of weak emergence can be extended into further contexts along similar lines

I have been speaking of underivability except by simulation as if there were a sharp dividing line separating weak emergent properties from merely resultant properties, but this is an oversimplification (Assad and Packard 1992) One can define various sharp distinctions involving underivability except by simulation, but focusing on one to the exclusion of the others is somewhat arbitrary The underlying truth is that properties come in various degrees of derivability without simulation, so there is a spectrum of more or less weak emergence A core concept of weak emergence concerns properties that in principle are underivable except by finite feasible simulation A slightly weaker notion of emergence concerns properties that in principle are derivable without simulation, but in practice must be simulated A slightly stronger notion of emergence concerns properties that are underivable except by simulation, but the requisite simulation is unfeasible or infinite A variety of even weaker and stronger notions also exist Nevertheless, the paradigm concept along this scale is weak emergence as defined above

It is important to recognize that my notion of weak emergence concerns how something *can* be derived, not whether it *has* been derived It concerns which derivations exist (in the Platonic sense), not which have been discovered Perhaps nobody has ever worked through a short-cut derivation of some macro property Nevertheless, if there is such a derivation, then the macro property is not weakly emergent If a genius like Newton discovers a new short-cut derivation for macro properties in a certain class of system, this

changes what properties we *think* are weakly emergent but not which properties *are* weakly emergent Notice also that weak emergence does not concern some human psychological or logical frailty It is not that human minds lack the power to work through simulations without the aid of a computer Nor is it that available computing power is too limited (e g , detailed simulations of the world's weather are beyond the capacity of current hardware) Rather, it involves the formal limitations of any possible derivation performed by any possible device or entity To dramatize this point, consider a Laplacian supercalculator that could flawlessly perform calculations many orders of magnitude faster than any human Such a supercalculator would be free from any anthropocentric or hardware-centered limitation in reasoning speed or accuracy Nevertheless, it could not derive weakly emergent properties except by simulation The Laplacian supercalculator's derivations of weak emergence might look instantaneous to us, but their logical form would be just like the logical forms of our derivations Each derivation iterates step by step through the aggregation of local interactions among the micro elements

The phrase "derivation by simulation" might seem to suggest that weak emergence applies only to what we normally think of as simulations, but this is a mistake Weak emergence also applies directly to natural systems, whether or not anyone constructs a model or simulation of them A derivation by simulation involves the temporal iteration of the spatial aggregation of local causal interactions among micro elements That is, it involves the local causal processes by which micro interactions give rise to macro phenomena The notion clearly applies to natural systems as well as computer models So-called "agent-based" or "individual-based" or "bottom-up" simulations in complexity science have exactly this form [9] They explicitly represent micro interactions, with the aim of seeing what implicit macro phenomena are produced when the micro interactions are aggregated over space and iterated over time My phrase "derivation by simulation" is a technical expression that refers to temporal iteration of the spatial aggregation of such local micro interactions We could perhaps use the phrase "derivation by iteration and aggregation," but that would be cumbersome Since "simulation" is coming to mean exactly this kind of process (Rasmussen and Barrett 1995), I adopt

the more economical phrase "derivation by simulation." Derivation by simulation is the process by which causal influence typically propagates in nature. Macro processes in nature are caused by the iteration and aggregation of micro causal interactions. The iteration and aggregation of local causal interactions that generate natural phenomena can be viewed as a computation (Wolfram 1994), just like the causal processes inside a computer. These intrinsic natural computations are a special case of derivation by simulation. Natural systems compute their future behavior by aggregating the relevant local causal interactions and iterating these effects in real time. They "simulate" themselves, in a trivial sense. Thus, derivation by simulation and weak emergence apply to natural systems just as they apply to computer models.

The behavior of weakly emergent systems cannot be determined by any computation that is essentially simpler than the intrinsic natural computational process by which the system's behavior is generated. Wolfram (1994) terms these systems "computationally irreducible." The point can also be expressed using Chaitin's (1966, 1975) notion of algorithmic complexity and randomness: roughly, the macro is random with respect to the micro, in the sense that there is no derivation of the macro from the micro that is shorter than an explicit simulation. Computational irreducibility—that is, weak emergence—is characteristic of complex systems and it explains why computer simulations are a necessary tool in their study.

4. Weak emergence and reduction in cellular automata

Some examples can make the ideas of weak emergence and derivation by simulation more concrete. The examples also illustrate the sort of systems studied in complexity science.[10] One advantage of such systems is that we have exact and total knowledge of the fundamental laws govern the behavior of the micro elements. The examples are all cellular automata, consisting of a two-dimensional lattice of cells, like an infinitely large checker board. Each cell can be in either of two states, which we'll refer to as being alive and being dead. (You can think of them equivalently as being in state 0 and 1, or black and white.)

Time moves forward in discrete steps The state of each cell at a given time is a simple function of its own state and the states of its eight neighboring cells at the previous moment in time, this rule is called the system's "update function " Assume that one of these systems is started with some initial configuration of living and dead cells (These initial states could be chosen by somebody or determined randomly) The next state of each cell is completely determined by its previous state and the previous state of its neighbors, according to the update function Notice that causal fundamentalism holds in cellular automata The only primitive causal interactions in the system are the interactions between neighboring cells, as specified by the system's update function If there are any higher-level causal interactions in the system, they all can be explained ultimately by the interactions among the system's elementary particles the individual cells

The only difference between the cellular automata that we will consider is their update functions The first updates the state of each cell as follows

> **All Life** 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

My name for this update function should be obvious, and so should its behavior No matter what configuration of living and dead cells the system has initially, at the next moment and for every subsequent moment every cell in the system is alive Given this update function, it is a trivial matter to derive the behavior of any individual cell or clump of cells in the system at any point in the future All regions at all times in the future consist simply of living cells

Part of what makes the All Life rule so trivial is that a cell's state does not make a difference to its subsequent state Living and dead cells alike all become alive The second update rule is slightly more complicated, as follows

> **Spreading Life** A dead cell becomes alive if and only if at least of its neighbors were alive at the previous moment, once a cell becomes alive it remains alive

The behavior of this system is also quite trivial to derive, and its name reflects this behavior Life spreads at the speed of light (one cell per moment of time) in all directions from any living cell Once a dead cell is touched by a living cell, it becomes alive and then remains alive forever after Life spreads from a single living cell in a steadily growing square If the initial configuration contains a random sprinkling of living cells, a square of life spreads from each at the speed of light Eventually these spreading squares overlap to form a connected shape growing at the speed of light

The third system we will consider is the most famous of all cellular automaton the so-called "Game of Life" devised in the 1960s by John Conway (Berlekamp et al 1982, see also Gardner 1983 and Poundstone 1985) It has the following update rule

> **Game of Life** A living cell remains alive if and only if either two or three of its neighbors were alive at the previous moment, a dead cell becomes alive if and only if exactly three of its neighbors were alive at the previous moment

The Game of Life's update rule is more complicated than the rules for All Life and Spreading Life, but it is still quite simple It is easy to calculate the subsequent behavior of many initial configurations For example, an initial configuration consisting of a single living cell will turn to all dead cells after one tick of the clock, and it will remain that way forever Or consider a 2x2 block of living cells Each of the cells in this initial configuration has three living neighbors, so it remains alive Each of the dead cells that border the block has at most two living neighbors, so it remains dead Thus the 2×2 block alone remains unchanging forever—an example of what is called a "still life" in the Game of Life Another interesting configuration is a vertical strip of living cells three cells long and one cell wide The top and bottom cells in this strip die at the first clock tick, since each has only one living neighbor. The middle cell remains alive, since it has two living neighbors But this is not all The two dead cells adjacent to the middle cell have three living neighbors—the three cells in the strip—so they each become alive Thus, after one clock tick, there is a horizontal strip of living cells, three cells long and one cell wide By parity of reasoning, one more clock tick turns this

configuration back into the original vertical strip Thus, with each clock tick this configuration changes back and forth between vertical and horizontal 3×1 strips—an example of what is called a "blinker" in the Game of Life

Still lives and blinkers do not begin to exhaust the possibilities One particular configuration of five living cells changes back into the same pattern in four clock ticks, except that the pattern is shifted one cell along the diagonal Thus, over time, this pattern glides across the lattice of cells at one quarter the speed of light, moving forever in a straight line along the diagonal—an example of a "glider." Other gliding patterns leave various configurations of living cells in their wake—these are called "puffers " Other configurations periodically spawn a new glider—these are called "glider guns " Still other configurations will annihilate any glider that hits them—these are called "eaters " Gliders moving at ninety degrees to each other sometimes collide, with various kinds of outcome, including mutual annihilation or production of a new glider

Streams of gliders can be interpreted as signals bearing digital information, and clusters of glider guns, eaters, and other configurations can function in concert just like AND, OR, NOT, and other logic switching gates These gates can be connected into circuits that process information and perform calculations In fact, Conway proved that these gates can even be cunningly arranged so that they constitute a universal Turing machine (Berlekamp et al 1982) Hence, the Game of Life can be configured in such a way that it can be interpreted as computing literally any possible algorithm operating on any possible input As Poundstone vividly puts it, the Game of Life can "model every precisely definable aspect of the real world" (Poundstone 1985, p 25)

For our present purposes, the most important respect in which the Game of Life differs from All Life and Spreading Life is that many properties in the Game of Life are weakly emergent For example, consider the macro property of indefinite growth (i e , increase number of living cells) Some initial configurations exhibit indefinite growth, and others do not Any configuration consisting only of still lifes and blinkers will not exhibit indefinite growth By contrast, a configuration consisting of a glider gun will exhibit indefinite growth,

since it will periodically increase the number of living cells by five as it spawns new gliders Other configurations are more difficult to assess The so-called R pentomino—a certain five-cell pattern that resembles the shape of the letter R—exhibits wildly unstable behavior Poundstone (1985, p 33) describes its behavior this way "One configration leads to another and another and another, each different from all of its predecessors On a high-speed computer display, the R pentomino roils furiously It expands, scattering debris over the Life place and ejecting gliders " Now, does the R pentomino exhibit indefinite growth? If the R pentomino continually ejects gliders that remain undisturbed as they travel into the infinite distance, for example, then it would grow forever But does it? The only way to answer this question is let the Game of Life "play" itself out with the R pentomino as initial condition That is, one has no option but to observe the R pentomino's behavior As it happens, after 1103 time steps the R pentomino settles down to a stable state consisting of still lifes and blinkers that just fits into a 51-by-109 cell region Thus, the halt to the growth of the R pentomino is a weakly emergent macro state in the Game of Life

By contrast, the behavior of any initial configuration in both All Life and Spreading Life are trivial to derive There is no need to observe the behavior of All Life and Spreading Life to determine whether the R pentomino in those cellular automata exhibits indefinite growth, for example The same holds for any other macroproperty in All Life and Spreading Life They exhibit no weakly emergent behavior

Figure 1 (a) Time evolution of the Game of Life starting from a 50 x 50 random initial condition in which 30% of the cells are alive

Figure 1 (b) The same pattern after 10 time steps

Figure 1 (c) The same pattern after 100 time steps

It is noteworthy how much of the interesting behavior of the Game of Life depends on the precise details of its cellular birth-death rule To get a feel for this, consider the time evolution of the Game of Life given a randomly generated initial condition, shown in Figure 1 (a)–(e) The Game starts at (a) with a 50×50 random initial condition in which 30% of the cells are alive, and in 10 time steps it has evolved into (b) By time 100 it is at (c), now a number of still lifes and blinkers are evident and a glider leaving from the upper left, but the pattern also contains many randomly structured piles of "muck" (d) shows time 300, now the pattern has grown slightly,

having spawned another glider (right side), preserved some still lifes and blinkers while creating and destroying others, and continuing to roil in two large unstructured piles of "muck" This pattern continues to grow slowly, and after 700 time steps at (e) it is mostly stable, consisting only of still lifes, blinkers, some gliders (out of the picture) moving off into the distance, and one random pile of muck near the top Eventually all the piles of muck dissipate and after many hundreds of time steps the pattern stabilizes with over sixty still lifes and blinkers spread out over a region about three times the size of the initial random pattern, with a few gliders wiggling off to infinity

Figure 1 (d) The same pattern after 300 time steps

Figure 1 (e) The same pattern after 700 time steps This pattern eventually reaches a stable configuration

But if we make a minor change in the birth-death rule, the resulting system's behavior changes completely For example, consider what happens to exactly the same random initial condition if survival is a little harder. (I will adopt the convention of naming an update function with the number of neighbors required to give birth to a new living cell followed by the number of neighbors required to keep a living cell alive)

> **3-3 Life** A dead cell becomes alive if and only if exactly three of its neighbors were alive at the previous moment, a living cell remains alive if any only if exactly three of its neighbors were alive at the previous moment

Figure 2 (a) The state of 3-3 Life, a near cousin of the Game of Life, after 10 time steps, after it has been started from exactly the same initial condition shown in Figure 1 (a)

Figure 2 (b) The same pattern in 3-3 Life after 14 time steps, at which point it has reached a completely stable configuration

Figure 2 (a) and (b) shows the behavior of 3-3 Life given the same random initial condition displayed in Figure 1 In stark contrast to the behavior of the Game of Life in Figure 1, 3-3 Life quickly reduces this pattern to a small stable configuration of still lifes and blinkers By time step 10 in (a), the configuration has collapsed to a small number of living cells, and by time step 14 in (b) its behavior has become stable, consisting of six blinkers and one still life The interesting thing about 3-3 Life is that all other initial conditions exhibit the same kind of behavior, they all quickly reduce to a small stable pattern consisting of at most some still lifes and blinkers This collapse to a few isolated periodic subpatterns is a universal generalization about 3-3 Life's global behavior This is a general macro-level law about 3-3

Life, somewhat analogous to the second law of thermodynamics for our world

Now, consider a different minimal change of the Game of Life's update function, one that makes birth a little easier but survival a little harder

> **2-2 Life** A dead cell becomes alive if and only if exactly two of its neighbors were alive at the previous moment, a living cell remains alive if and only if exactly two of its neighbors were alive at the previous moment.

This cellular automaton exhibits a completely different kind of behavior from both the Game of Life and 3-3 Life A typical example of its behavior is shown in Figure 3 (a) and (b), after it has been started with the same initial condition used in Figures 1 and 2 In this case, though, a random "slime" of living and dead cells steadily grows and eventually spreads over the entire world After 10 time steps, it has evolved into (a), and a random "slime" pattern of cells can already be seen to be growing By time 100 in (b), the random slime has increased in size by more than a factor of four (note reduced size scale) This random slime pattern will continue to grow indefinitely Similar random slime patterns grow from virtually all other initial conditions

Figure 3 (a) The time evolution of 2-2 Life, another near cousin of the Game of Life, after started with exactly the same random initial condition in Figures 1 and 2

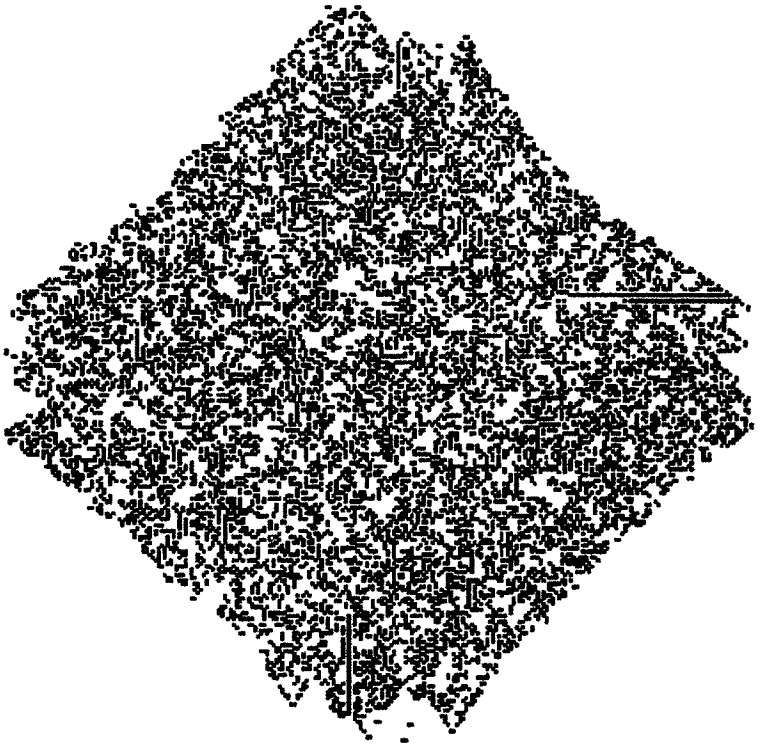

Figure 3 (b) The same pattern after 100 time steps The pattern will continue to grow indefinitely

in 2-2 Life [11] Figures 4 (a) and (b) show similar but nonidentical slime patterns growing from two initial configurations that differ only in the position of one living cell This spreading chaos is a general macro-level law about 2-2 Life

A little experimentation is all it takes to confirm the typical behavior of 3-3 Life and 2-2 Life collapse to isolated periodicity and spreading chaos From a statistical point of view, their global behavior is very easy to predict Changing the state of a cell here or there in an initial condition makes no difference to the quality of their global behavior, indeed, neither does *drastically* changing the initial configuration By contrast, the Game of Life has no typical global behavior Some configurations quickly collapse into stable periodic patterns

Other very similar configurations continue to change indefinitely Changing the state of one cell can completely change the system's global behavior Neither 3-3 Life nor 2-2 Life has the exquisite sensitivity and balance of order and disorder that allows the Game of Life to exhibit complex macro-level patterns such as switching gates, logic circuits, or universal computers

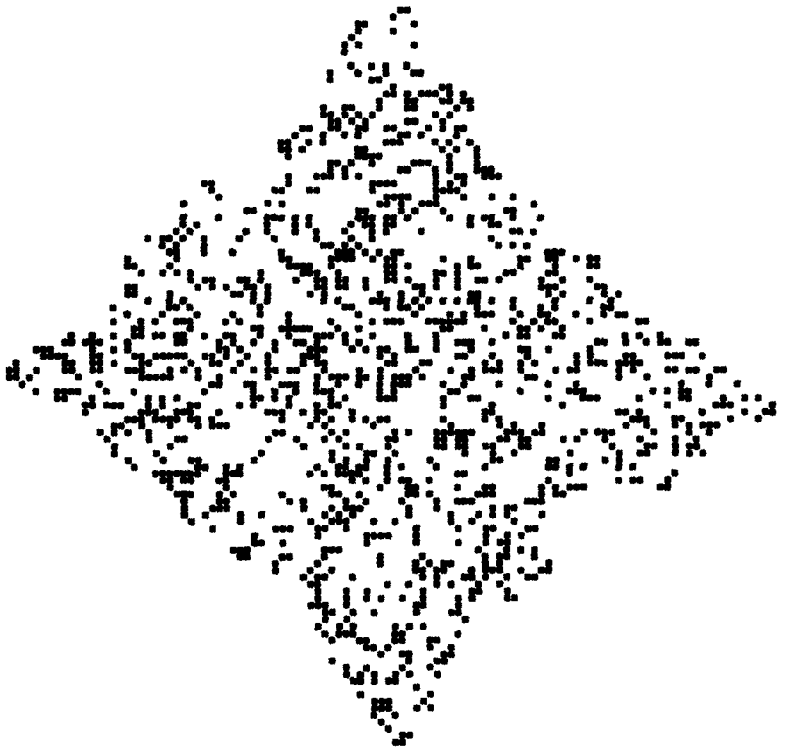

Figure 4 (a) The random "slime" pattern in 2-2 Life after 50 time steps from an initial configurations consisting of 30 living cells confined with a 10 x 10 region

Figure 4 (b) The random "slime" pattern in 2-2 Life after 50 time steps from an initial configuration that differs from (a) only in the position of one living cell Note that this pattern is qualitatively similar to that in (a), and also to that in Figure 3 (b)

Nevertheless, all three cellular automata exhibit weak emergence This is easily recognizable from the fact that their *exact* global behavior (whether statistically predictable or not) can be derived only by simulation—iterating through time the aggregate local effect of the update function across all cells Sufficient experience with 3-3 Life and 2-2 Life provides empirical evidence for macro-level laws about the *kind* of behavior they each exhibit But the only way to tell exactly *which* instance of that behavior will be produced from a given initial configuration is to watch how the system unfolds in time—i e , to "simulate" it Given a random initial configuration, you can

be sure that 3-3 Life will quickly reduce to a collection of isolated still lifes and blinkers, and that 2-2 Life will produce a steadily growing chaotically changing mixture of living cells But the only way to determine exactly which collection of still lifes and blinkers, or exactly which chaotically changing sequence of living cells, is to step through the behavior of the whole system This is the signature of systems with weak emergent properties

Earlier we saw that it is often difficult to tell whether a given initial condition in the Game of Life leads to indefinite growth 3-3 Life and 2-2 Life are different in this respect Given 3-3 Life's law of collapse to isolated periodicity, we know that 3-3 Life never shows indefinite growth Likewise, given 2-2 Life's law of spreading chaos, we know that 2-2 Life (virtually) always shows indefinite growth However, the presence or absence of indefinite growth is still a weak emergent property in 3-3 Life and 2-2 Life Our knowledge that 3-3 Life never exhibits indefinite growth depends on having learned its law of collapse to periodicity, and analogously for 2-2 Life's law of spreading chaos But these macro-laws are *emergent laws*—that is, laws about the system's emergent properties—and they are discovered empirically Our knowledge of these laws comes from our prior empirical observations of how the systems behave under different initial conditions This is analogous to how we know that a rock can break a window Weak emergence concerns the derivation of macro-properties, and these derivations involve exact and absolutely certain inferences from the system's micro facts Empirically grounded generalizations about the system's behavior play no part in such derivations Thus, in the sense that is relevant to weak emergence, it is not possible to derive the presence or absence of indefinite growth in 3-3 Life or 2-2 Life

It is important to note that all five of our cellular automata are ontologically and causally on a par, though not all exhibit weak emergence Each cellular automaton is nothing but a lattice of cells, and the behavior of its cells is wholly determined by a local update function Any large-scale macro patterns exhibited by the cellular automata are derived from iterating the behavior of each cell over time according to the system's update function and aggregating the cells over the lattice In other words, the macro behavior of each system

is constituted by iterating and aggregating local causal interactions

Emergence is sometimes contrasted with reduction, but this oversimplifies matters, especially for weak emergence The three kinds of reduction we distinguished earlier (reduction of ontology, causation, and explanation) need not go hand in hand Ontological reductionism and causal reductionism hold for all cellular automata—indeed, for all weak emergence Local causal influence propagates in space and time in the same way in all cellular automata The distinctive feature of those cellular automata that exhibit weak emergence is not the lack of ontological or causal reductionism, nor the lack of context-sensitive derivation of macro properties It is simply having micro-level context-sensitive interactions that are complex enough that their aggregate effect has no short-cut derivation Macro properties in All Life and Spreading Life always have a short-cut derivation But this is not so for 2-2 Life, 3-3 Life, or the Game of Life

Embracing ontological and causal reduction permits weak emergence to avoid one of the traditional complaints against emergence J J C Smart (1963), for example, objected that emergence debarred viewing the natural world as a very complicated mechanism However, weak emergence postulates just complicated mechanisms with context-sensitive micro-level interactions Rather than rejecting reduction, it requires (ontological and causal) reduction, for these are what make derivation by simulation possible

5. Downward causation of weak emergence

Ordinary macro causation consists of causal relations among ordinary macro objects Examples are when a rock thrown at a window cracks it, or an ocean wave hits a sand castle and demolishes it [12] But macro-level causes can also have micro-level effects This is termed "downward causation " Downward causation is a straightforward consequence of ordinary macro causation To see this, choose some micro piece of the macro effect and note that the macro cause is also responsible for the consequent changes in the micro piece of the macro effect For example, consider the first molecular bond that broke when the window cracked The rock caused that molecular

bond to break Or consider the violent dislocation of a particular grain of sand at the top of the castle The wave caused its dislocation

Figure 5 A glider gun which has so far shot six gliders moving away along the southeast diagonal

Emergence is interesting in part because of emergent causal powers Emergent phenomena without causal powers would be mere epiphenomena Weak emergent properties, objects, phenomena, etc often have causal powers For example, the property of being a glider gun is a weak emergent property of a certain macro-level collection of cells in the Game of Life, and it has the causal power of generating a regular stream of gliders—a macro-level pattern of cells propagating in space For example, the glider gun shown in Figure 5 shoots another glider every forty-six time steps This weak emergent macro-level causation brings downward causation in its train To pick just one example, as successive gliders are shot from the gun they cause a certain pattern of behavior in the individual cells in their path To make things concrete, consider one specific cell (call it cell 17) the left-most glider in the second glider in the stream When a glider first touches cell 17, the cell becomes alive While the glider passes,

cell 17 remains alive for three more generations Then it becomes dead and remains so for forty-two more time steps, until the next glider touches it Clearly, this repeating pattern in cell 17's behavior is caused by the macro-level glider gun It is a micro effect of a macro cause, i e , it is an example of downward causation of the glider gun

Campbell (1974) called attention to emergent downward causation, because he wanted to combat excessive reductionism and bolster the perceived reality of higher-level emergent biological organization Downward causation is also emphasized recently by advocates of strong emergence (e g , Kim 1992 and 1999, O'Conner 1994), because the characteristic feature of strong emergence is irreducible downward causal power

Downward causation is now one of the main sources of controversy about emergence There are at least three apparent problems The first is that the very idea of emergent downward causation seems incoherent in some way Kim (1999, p 25) introduces the worry in this way

> The idea of downward causation has struck some thinkers as incoherent, and it is difficult to deny that there is an air of paradox about it After all, higher-level properties arise out of lower-level conditions, and without the presence of the latter in suitable configurations, the former could not even be there So how could these higher-level properties causally influence and alter the conditions from which they arise? Is it coherent to suppose that the presence of X is entirely responsible for the occurrence of Y (so Y's very existence is totally dependent on X) and yet Y somehow manages to exercise causal influence on X?

The upshot is that there seems to be something viciously circular about downward causation

The second worry is that, even if emergent downward causation is coherent, it makes a difference only if it violates micro causal laws (Kim 1997) This worry arises because of a background presumption that micro events are caused by prior micro events according to fundamental micro laws If emergent downward causation brought about some micro event E, there would be two unattractive possibilities One is that E is also brought about by some micro cause, in

which case the emergent macro cause of E is irrelevant The other possibility is that the macro and micro causes conflict because micro causation would have brought about an incompatible micro effect, E', so the downward causation would violate the fundamental micro laws [13]

Even if emergent downward causation is coherent and consistent with fundamental micro laws, a third worry still arises This worry also grows out of the fact that micro-level events have sufficient micro-level causes Any macro-level cause that has a micro-level effect (i e , any downward causation) will compete for explanatory relevance with the micro-level explanation But the micro-level explanation is more fundamental So the micro-level explanation of the micro-level effects will preempt the macro-level explanation This "exclusion" argument has been emphasized by Kim (1992,1999), and it has provoked extensive contemporary discussion (e g , Chalmers 1996)

I want to show that these worries present no problems for weak downward causation There is a simple two-step argument that shows this The first step is to note that ordinary downward causation is unproblematic An ocean wave demolishes a sand castle, causing the violent dislocation of a grain of sand A vortex in the draining bathtub causes a suspended dust speck to spin in a tight spiral A traffic jam causes my car's motion to slow and become erratic I take it as uncontroversial that such ordinary cases of downward causation are philosophically unproblematic They violate no physical laws, they are not preempted by micro causes, and they are not viciously circular or incoherent The second step is to note that weak downward causation is simply a species of ordinary downward causation Many ordinary macro objects with downward causal effects are weakly emergent Waves, vortices, and traffic jams are all plausible candidates for weak emergence Their macro causal powers are constituted by the causal powers of their micro constitutents, and these are typically so complicated that the only way to derive their effects is by iterating their aggregate context-dependent effects—i e , by simulation In any event, weak emergent properties and objects have the kind of relation to their micro-level bases that ordinary macro-scale physical properties and objects have to their bases Weak emergent

causal powers are constituted by the causal powers of the micro constituents The weak emergent macro cause is nothing but the iteration of the aggregate micro causes Ontological and causal reduction holds Since weak downward causation is just a subset of ordinary macro causation, the one is no more problematic than the other

This defense of weak downward causation is confirmed when we examine each of the three worries First, since a weak macro cause is identical with the aggregation and iteration of micro causes, weak macro causation cannot violate micro causal laws In fact, since weak macro causation is constituted by the appropriate context-sensitive micro causation, weak macro causation *depends* on the micro causal laws They are the mechanism through which weak macro causation is realized Second, since a weak macro cause is nothing more than the aggregation of micro causes, macro and micro causes are not two things that can compete with each other for causal influence One constitutes the other So, the micro causes cannot exclude weak macro causes Third, once we see that weak downward causation does not violate fundamental micro explanations and is not preempted by them, the apparent incoherence or vicious circularity of emergent downward causation reduces to the worry that downward causal effects must precede their causes But weak downward causation is diachronic Higher-level properties can causally influence the conditions by which they are sustained, but this process unfolds over time The higher-level properties arise out of lower-level conditions, and without those lower-level conditions the higher-level properties would not be present But a weak macro cause cannot alter the conditions from which it arose At most it can alter the conditions for its subsequent survival, and this is neither viciously circular nor incoherent

These abstract considerations are concretely exemplified by our earlier discussion of weak downward causation in the Game of Life the glider gun that causes a repeating pattern in cell 17 (Figure 5) First, the downward causation is diachronic, the micro effects are subsequent to their macro causes So there is no vicious circularity Second, this downward causation is brought about simply by aggregating the state changes in each cell, given the appropriate initial condition, and then iterating these aggregated local changes over

time The glider gun (a macro object consisting of a special aggregation of micro elements) creates the context for qualitatively distinctive (macro- and) micro-level effects, but this violates no micro laws Indeed, it exploits those micro laws Third, macro glider gun explanation does not compete with micro update-rule explanation The macro explanation is constituted by iterating the aggregated micro explanation So explanatory exclusion is no threat

6. The autonomy of weak emergence

The preceding discussion of downward causation emphasized that weak emergent phenomena are nothing more than the aggregation of the micro phenomena that constitute them This prompts a final worry about whether the explanations of weak emergent phenomena are sufficiently autonomous Consider some weak emergent macro property *P* This property is brought about by the aggregation of a collection of micro causal histories—the causal histories of all the micro properties that constitute *P* So, isn't the underlying explanation of *P* just the aggregation of the micro explanations of all the relevant micro elements? If the underlying explanation of the macro phenomena is merely the aggregation of micro explanations, then all the real explanatory power resides at the micro level and the macro phenomena are merely an effect of what happens at the micro level [14] In this case, weak emergent phenomena have no real macro-level explanatory autonomy

Some of the plausibility for this line of argument comes from the ontological and causal reducibility of weak emergent phenomena Since their existence and causal powers are nothing more than the existence and causal powers of the micro elements that instantiate them, wouldn't their real underlying explanation also be at the micro level? Weak emergent macro phenomena have various macro explanations, and these explanations may be convenient and useful for us In particular, the overwhelming complexity of their aggregate micro explanation typically overwhelms us, preventing us from grasping how they are generated [15] Hence we resort to computer simulations, observing the resulting macro properties and experimentally manip-

ulating micro causes to see their macro effects The computer can aggregate micro causal histories fast enough for us to see their weak emergent macro effect But isn't the explanation of the macro effect exhausted by the micro causal processes?

The nub of this worry is that, if weak emergence has any macro explanatory autonomy, the autonomy is just our inability to follow through the details of the complicated micro causal pathways It amounts to nothing more than an epistemic obstacle to following the ontological and causal reduction We study the weak emergent effects of these micro causal processes by observing the macro effects directly (in nature or in computer simulations) But the macro phenomena are mere effects of micro causal processes This explanatory autonomy is merely epistemological rather than ontological It reflects just our need for macro explanations of certain phenomena, it does not reflect any distinctive objective structure in reality In particular, it does not reflect any autonomous and irreducible macro-level ontology [16] Or, at least, that is the worry

The correct response to this worry takes different branches for different kinds of weak emergence In some cases the worry is sound All weak emergence has a certain epistemic autonomy, for the context-sensitive micro causal interactions can be explained only by iterating the aggregated effect of all the micro interactions [17] Thus, as a practical matter, we must study them through simulation Some weak emergence is nothing more than this Such weak emergent phenomena are mere effects of micro contingencies and their explanatory autonomy is merely epistemological

One example of such merely epistemological weak emergence is a configuration in the Game of Life that accidentally (so to speak) emits an evenly spaced stream of six gliders moving along the same trajectory What is crucial is that this configuration contains no glider gun It's an irregular collection of still lifes, blinkers, and miscellaneous piles of "muck" that happens to emit six gliders It might be somewhat like the configuration in Figure 1 at time 100, which has just emitted a glider from the northwest corner, except that it happens to emit five more evenly spaced gliders in the same direction The configuration is always changing in an irregular fashion, and there is no overarching explanation for why the six gliders stream

out The explanation for the gliders is just the aggregation of the causal histories of the individual cells that participate in the process The macro-level glider stream is a mere effect of those micro contingencies

Contrast the accidental glider stream with the configuration of cells shown in Figure 5 This configuration of cells also emits an evenly spaced stream of six gliders heading in the same direction Furthermore, the aggregation of the causal histories of the individual cells that participate in the process explains the glider stream However, there is more to the explanation of this second stream of gliders, because the configuration of cells is a *glider gun* and glider guns always emit evenly spaced gliders in a given direction The glider gun provides an overarching, macro-level explanation for the second glider stream Furthermore, this same macro explanation holds for any number of other guns that shoot other gliders There are many kinds of gliders and many kinds of glider guns (Figure 6 shows two more glider guns) The aggregate micro explanation of the second glider stream omits this information Furthermore, this information supports counterfactuals about the stream The same glider stream would have been produced if the first six gliders had been destroyed somehow (e g , by colliding with six other gliders) Indeed, the same glider stream would have been produced if the configuration had been changed into any number of ways, as long as the result was a gun that shot the same kind of gliders Any such macro gun would have produced the same macro effect Thus, the full explanation of the six gliders in Figure 5 consists of more than the aggregation of the causal histories of the relevant micro cells There is a macro explanation that is not reducible to that aggregation of micro histories If those micro histories had been different, the macro explanation could still have been true The macro explanation is autonomous from the aggregate micro explanation

Consider another example the chaotically changing "slime" that spreads at the speed of light from an initial configuration in 2-2 Life (Figures 3 and 4) These examples illustrate a general macro law that I mentioned earlier 2-2 Life always generates such random slime, provided the initial configuration is dense enough for any life to grow Each instance of spreading slime can be explained by aggregating the

causal histories of the micro cells that participate in the pattern But this aggregate micro explanation leaves out an important fact the random slime macro law Alter the initial condition (and thus the micro histories) in virtually any way you want, and the same kind of macro behavior would still be generated The fact that the same kind of behavior would have been produced if the micro details had been different is clearly relevant to the explanation of spreading slime observed in any particular instance The macro law explanation is autonomous from the aggregation of micro histories in each particular instance

Figure 6 Two more guns shooting gliders on the southeast diagonal Note that the configuration of cells constituting these two guns and the gun in Figure 5 all differ

Notice that weak emergent phenomena in the real world have the same kind of macro explanatory autonomy Consider a transit strike that causes a massive traffic jam that makes everyone in the office late to work Each person's car is blocked by cars with particular and idiosyncratic causal histories The traffic jam's ability to make people late is constituted by the ability of individual cars to block other cars Aggregating the individual causal histories of each blocking car explains why everyone was late However, the aggregate micro explanation obscures the fact that everyone would still have been late if the micro causal histories had been different The transit strike raised the traffic density above a critical level So, even if different individual cars had been on the highway, the traffic still would have been jammed and everyone still would have been late The critical traffic density provides a macro explanation that is autonomous from any particular aggregate micro explanation

My strategy for showing that macro explanations of some weak emergent phenomena are autonomous is analogous to well-known strategies for showing that explanations in special sciences can be autonomous from the explanations provided in underlying sciences [18] One complementary strategy for defending special sciences emphasizes that macro explanations can contain causally relevant information that is missing from micro explanations (e g , Jackson and Pettit 1992, Sterelny 1996) My arguments above have this form The original defense of special sciences focussed on multiple realization and the resulting irreducibility of macro explanations (e g , Fodor 1974 and 1997) My arguments can be recast in this form Note that glider guns are multiply realizable in the Game of Life, as are random slimes in 2-2 Life, as are traffic jams, and none is reducible to any particular collection of aggregate micro phenomena

Either way the argument is put, the conclusion is the same Macro explanations of some weak emergent phenomena have a strong form of autonomy Note that this is not the mere epistemological autonomy that comes with all weak emergence The accidental glider stream discussed above is just an effect of micro contingencies By contrast, the glider gun, the random slime, and the traffic jam are instances of larger macro regularities that support counterfactuals about what would happen in an indefinite variety of different micro

situations An indefinite variety of micro configurations constitute glider guns in the Game of Life, and they all shoot regular streams of gliders An indefinite variety of micro configurations constitute random slime in 2-2 Life, and they all spread in the same way An indefinite variety of micro configurations constitute traffic jams, and they all block traffic Each macro-level glider gun, random slime, and traffic jam is nothing more than the micro-level elements that constitute it But they participate in macro regularities that unify an otherwise heterogeneous collection of micro instances Fodor argues that macro regularities in the Game of Life and similar systems have micro reductions because the macro regularities are "logical or mathematical constructions" out of micro regularities (1997, n 5) But Fodor fails to appreciate that micro realizations of the macro regularities in cellular automata are as wildly disjunctive as any in the special sciences

So, the explanatory autonomy of weak emergence can take two forms When the emergent phenomena are mere effects of micro contingencies, then their explanatory autonomy is merely epistemological The explanatory autonomy does not signal any distinctive macro structure in reality But weak emergent phenomena that would be realized in an indefinite variety of different micro contingencies can instantiate robust macro regularities that can be described and explained only at the macro level The point is not just that macro explanation and description is irreducible, but that this irreducibility signals the existence of an objective macro structure This kind of robust weak emergence reveals something about reality, not just about how we describe or explain it So the autonomy of this robust weak emergence is ontological, not merely epistemological [19]

Not all weak emergence is metaphysically or scientifically significant In some quarters emergence *per se* is treated as a metaphysically significant category that signals a qualitative difference in the world This is not the perspective provided by weak emergence Much weak emergence is due just to complicated micro-level context-sensitivity, the same context-sensitivity that is ubiquitous in nature In some cases, though, these context-sensitive micro interactions fall into regularities that indicate an objective macro structure in reality These macro regularities are important scientifically, for they explain the

generic behavior of complex systems in nature The Game of Life instantiates fantastically complicated macro structures like universal Turing machines only by exploiting the ability of glider guns to send signals arbitrary distances in time and space The law of spreading random slime in 2-2 Life is a hallmark of one of the four fundamental classes of cellular automata rules identified by Wolfram (1994) Explaining robust traffic patterns necessitates identifying the critical role of traffic density A significant activity in complexity science is sifting through the emergent behavior of complex systems, searching for weak emergent macro properties that figure in robust regularities with deep explanatory import

7. Conclusions

The problem of emergence arises out of attempting to make sense of the apparent macro/micro layers in the natural world I have argued that what I call weak emergence substantially solves this problem The weak emergence perspective is ontologically and causally reductionistic, and this enables it to avoid many of the traditional worries about emergence, such as those involving downward causation But weak emergence is still rich enough for an ontology of objective macro-level structures Indeed, the search for robust weak emergent macro-structures is one of the main activities in complexity science— exactly the science that attempts to explain the apparent emergent phenomena in nature Could there be a better guide for understanding emergence in nature than complexity science?

Weak emergence is prevalent in nature, but it is unclear whether it is all the emergence we need In particular, some aspects of the mind still strenuously resist ontological and causal reduction, examples include fine-grained intentionality, the qualitative aspects of consciousness, freedom, and certain normative states Weak emergence can get no purchase on these phenomena until we have a (context-sensitive) reductionistic account of them As long as this is in doubt, so is the final reach of weak emergence However this turns out, weak emergence should still illuminate a variety of debates and confusions about the relations between macro and micro

These range from long-standing controversies over the autonomy of the special sciences to newer debates about whether macro evolutionary patterns are mere effects of micro processes or reflect genuine species selection (Vrba 1984, Sterelny 1996)

Emergence is often viewed synchronically An organism at a given time is thought to be more than the sum of its parts that exist at that time Your mental states at a given time are thought to emerge from your neuro-physical states at that time By contrast, the primary focus of weak emergence is diachronic It concerns how the macro arises over time from the micro, i e , the causal process (derivation) by which the micro constructs the macro This is a bottom-up generative process, rooted in context-sensitive micro-level causal interactions

The advent of modern philosophy is conventionally presented as the Cartesian triumph over Aristotelian scholasticism An Aristotelian thesis that attributed natures on the basis of a rich dependence on generating context was supplanted by a Cartesian antithesis that attributed reductionistic essences independent of context Computer simulations allow weak emergence to extend reductionism into new territory, but they do so by embodying the idea that something's nature can depend on its genesis Thus, the macro can depend on the context-sensitive process from which it arises and by which it is maintained In this way, weak emergence can be viewed as a new synthesis [20]

References

Assad, A M , and N H Packard (1992) Emergent Colonization in an Artificial Ecology In Varela and Bourgine (eds) *Towards a practice of autonomous systems* Cambridge MIT Press, pp 143–52

Baas, N A (1994) Emergence, hierarchies, and hyperstructures In C G Langton (ed) *Artificial life III* Redwood City Addison-Wesley, pp 515–37

Beckner, Morton (1974) Reduction, hierarchies and organicism In F J Ayala and T Dobzhansky (eds) *Studies in the philosophy of biology Reduction and related problems* Berkeley University of California Press, pp 163–76

Bedau, M A (1997) Weak emergence *Philosophical Perspectives* 11 375–99

Bedau, M A , J McCaskill, N Packard, S Rasmussen (eds) (2000) *Artificial life VII* Cambridge MIT Press

Berlekamp, E R , J H Conway, and R K Guy (1982) *Winning ways for your mathematical plays* Vol 2 New York Academic Press

Campbell, Donald T (1974) 'Downward causation' in hierarchically organised biological systems In F J Ayala and T Dobzhansky, eds , *Studies in the philosophy of biology Reduction and related problems* Berkeley University of California Press, pp 179–86

Chaitin, G J (1966) On the length of programs for computing finite binary sequences *Journal of the Association of Computing Machinery* 13 547–69

—— (1975) A theory of program size formally identical to information theory *Journal of the Association of Computing Machinery* 22 329–40

Chalmers, D J (1996) *The conscious mind In search of a fundamental theory* New York Oxford University Press

Dennett, Daniel (1991) Real patterns *Journal of Philosophy* 87 27–51

Farmer, J D , Lapedes, A , Packard, N , and Wendroff, B (eds) (1986) *Evolution, games, and learning Models for adaptation for machines and nature* Amsterdam North Holland

Fodor, Jerry (1974) Special sciences *Synthese* 28 97–115

—— (1997) Special sciences still autonomous after all these years *Philosophical Perspectives* 11 149–63

Forrest, S (ed) (1989) *Emergent computation Self organizing, collective, and cooperative phenomena in natural and artificial computing networks* Amsterdam North-Holland

Gardner, M (1983) *Wheels, life, and other mathematical amusements* New York Freeman

Gaussier, P, and Nicoud, J -D (eds) (1994) *From perception to action* Los Alamitos, Calif IEEE Computer Society Press

Gillett, Carl (Unpublished) Strong emergence as a defense of non-reductive physicalism A physicalist metaphysics for 'downward' determination

Harré, Rom (1985) *The philosophies of science* Oxford Oxford University Press

Holland, John (1998) *Emergence From chaos to order* Reading, MA Helix Books

Jackson, F, and P Pettit (1992) In defense of explanatory ecumenism *Economics and Philosophy* 8 1–21

Kauffman, Stuart (1995) *At home in the universe The search for the laws of self-organization and complexity* New York Oxford University Press

Kim, Jaegwon (1992) "Downward-causation" in emergentism and nonreductive physicalism In A Beckerman, H Flohr, and J Kim (eds) *Emergence or reduction? Essays on the prospects of nonreductive physicalism* Berlin Walter de Gruyter, pp 119–38

—— (1997) The mind-body problem taking stock after forty years *Philosophical Perspectives* **11** 185–207

—— (1999) Making sense of emergence *Philosophical Studies* **95** 3–36

Langton, C , C E Taylor, J D Farmer, S Rasmussen (eds) (1992) *Artificial life II* SFI Studies in the Sciences of Complexity, Vol X Reading, Calif Addison-Wesley

Newman, David V (1996) Emergence and strange attractors *Philosophy of Science* **63** 245–61

O'Conner, T (1994) Emergent properties *American Philosophical Quarterly* **31** 91–104

Poundstone, W (1985) *The recursive universe* Chicago Contemporary Books

Rasmussen, S , N A Baas, B Mayer, M Nilson, and M W Olesen (2001) Ansatz for dynamical hierarchies *Artificial Life* **7** 329–353

Rasmussen, S , and C L Barrett (1995) Elements of a theory of simulation In F Morán, A Moreno, J J Merelo, and P Chacón (eds) *Advances in artificial life* Berlin Springer, pp 515–29

Rueger, Alexander (2000) Physical emergence, diachronic and synchronic *Synthese* **124** 297–322

Ronald, E M A , M Sipper, and M S Capcarrere (1999) Design, observation, surprise! A test of emergence *Artificial Life* **5** 225–239

Silberstein, M and J McGeever (1999) The search for ontological emergence *The Philosophical Quarterly* **49** 182–200

Simon, Herbert A (1996) *The sciences of the artificial* Cambridge MIT Press

Smart, J J C (1963) *Philosophy and scientific realism* London Routledge and Keagan Paul

Sperry, R W (1969) A modified concept of consciousness *Psychological Review* **76** 532–6

Sterelny, Kim (1996) Explanatory pluralism in evolutionary biology *Biology and Philosophy* **11** 193–214

Strawson, P F (1963) *Individuals* Garden City, N Y Doubleday

Varela, F , and P Bourgine (1992) *Towards a practice of autonomous systems* Cambridge, Mass MIT Press

Vrba, E S (1984) What is species selection? *Systematic Zoology* 33 318–29
Wimsatt, William (1986) Forms of aggregativity In A Donagan, A N Perovich, Jr, and M V Wedin (eds) *Human nature and natural knowledge* Dordrecht Reidel, pp 259–91
—— (1997) Aggregativity reductive heuristics for finding emergence *Philosophy of Science* 64 (Proceedings), S372–S384
Wolfram, S 1994 *Cellular automata and complexity* Reading, Mass Addison-Wesley

Keywords

emergence, downward causation, autonomy

Department of Philosophy
Reed College
Portland OR 97202
USA
mab@reed edu
http //www reed edu/~mab

Notes

[1] These cellular automata include the Game of Life so the present paper illustrates the philosophical versatility of cellular automata, which Dennett (1991) recently emphasized

[2] There is no standard accepted terminology for referring to different kinds of emergence, so my terminology of "nominal," "weak" and "strong" might clash with the terminology used by some other authors In particular, Gillett (unpublished) means something else by "strong" emergence

[3] Although my point in the text is unaffected by this, note that this example really involves multiple levels of emergence, for we could split these levels more finely into macro (micelles), meso (polymers), and micro (monomers) See Rasmussen et al (2001) for an analysis and a model of this situation

[4] Since the two more restricted notions of emergence are proper subsets of nominal emergence, they of course exhibit the two hallmarks of emergence that characterize nominal emergence However, they each also capture their own distinctive and specific forms of dependence and autonomy, as the subsequent discussion shows

[5] Supervenient properties, in this context, are macro properties that can differ only if their micro property bases differ, there can be no difference in supervenient properties without a difference in their micro bases

[6] The qualifier "weak" is intended to highlight the contrast with the "strong" irreducible macro causal powers characteristic of strong emergence I need some qualifier, since weak emergence is just one among many kinds of emergence, but "weak" has the drawback of vagueness I would prefer a more descriptive term, but I have not found an appropriate one For example, "reductive" would emphasize weak emergence's ontological and causal reducibility, but it would obscure its explanatory irreducibility One sometimes sees weak emergence described as "innocent" emergence (e g , Chalmers 1996) This calls attention to our metaphysical evaluation of weak emergence, but it does not identify the source of this evaluation This is unfortunate since different kinds of emergence are metaphysically innocent for different reasons (compare nominal and weak emergence) Unfortunate for a related reason is "statistical" emergence "Statistical" does bring to mind a picture of macro phenomena arising out of the aggregation of micro phenomena, but it does not help distinguish the special kind of aggregation involved in weak emergence Terms like "explainable" or "non-brute" emergence have the same problem Thus, I will continue to use "weak" until I find a better alternative

[7] Thus, weak emergence can be exhibited by systems that also involve strong emergence The fates of weak and strong emergence are independent

[8] A "backward looking" emergent object is one the existence of which is weakly emergent, and a "forward looking" emergent object is one with weak emergent behavior, causal powers, etc

[9] There are different kinds of simulations My account of weak emergence fits best the agent-based simulations that explicitly represent micro causal interactions, but it can be extended to other simulation methods like those based on differential equations

[10] A variety of other kinds of systems studied in complexity science can be found by surveying conference proceedings, such as Farmer et al 1986, Forrest 1989, Langton et al 1992, Varela and Bourgine 1992, Gaussier and Nicoud 1994, and Bedau et al 2000

[11] The rare exceptions arise when the initial configuration is too sparse to support any life

[12] I will speak of macro objects as causes, where referring to their macro properties or events involving them as causes might be more appropriate I trust that no confusion will result

[13] This worry made Beckner (1974) conclude that emergent downward causation would require micro-level indeterminism, so that macro causes can have micro effects without violating micro physical laws Micro-level indeterminism is clearly an unsatisfactory way to save emergent downward causation, though There is no guarantee that the indeterminism would be available exactly where and when it is needed, and brute downward causal determination of micro-indeterministic events would be mysterious

[14] This would be the analog of Vrba's effect hypothesis about macro evolutionary properties (Vrba 1994)

[15] The opacity of the aggregate micro-level causal mechanisms in the agent-based models is a current source of unease about complexity science

[16] Note also that if weak emergence has mere epistemological autonomy, then weak emergent macro causation is spurious rather than genuine causation For the apparent macro causation is really nothing more than an effect of micro causal processes It would follow that weak downward causation is also spurious So, the fate of genuine weak downward causation hinges on weak emergence having more than epistemological autonomy

[17] Context-sensitive micro interactions are necessary for weak emergence but they are not sufficient All Life and Spreading Life have context-sensitive micro interactions but they are so trivial that the resulting macro properties are not weakly emergent

[18] Nonreductive physicalism in contemporary philosophy of mind is probably most plausible to cast as an instance of weak emergence However, my defense of weak emergence here is not tied to the fate of nonreductive physicalism

[19] Some, such as Silberstein and McGeever (1999) and perhaps Gillett (unpublished) will still classify this robust weak emergence as mere epistemological emergence, on the grounds that it embraces ontological and causal reduction (mereological supervenience) However, I think this is an excessively liberal view of epistemological emergence Consider an analogy Is the difference between the (presumably hypothetical) world in which all special sciences are reducible to fundamental physics and the (presumably actual) world in which they are autonomous merely epistemological? Is there nothing in the ontological structure of the second world that *makes* the special sciences autonomous? Presumably not

[20] Thanks for helpful comments to audiences at SCTPLS'99 in Berkeley CA (July 1999), at ISHPSSB'01 in Quinnipiac CT (July 2001), and at the philosophy department at the University of Oklahoma (October 2001), where some of the ideas in this paper were presented Thanks also for helpful discussion to Carl Gillett, Paul Hovda, Brian Keeley, Dan McShea,

Norman Packard, Steen Rasumssen, David Reeve, Edmund Ronald, Andre Skusa, Kelly Smith, and Pietro Speroni di Fenizio