Emergence and Simulation

Alan Baker

Department of Philosophy
Swarthmore College
Swarthmore PA 19081, USA
abaker1@swarthmore.edu

Abstract. One approach to characterizing the elusive notion of emergence is to define that a property is emergent if and only if its presence can be derived but only by simulation. In this paper I investigate the pros and cons of this approach, focusing in particular on whether an appropriately distinct boundary can be drawn between simulation-based and non-simulation-based methods. I also examine the implications of this definition for the epistemological role of emergent properties in prediction and in explanation.

Keywords: emergence, simulation, explanation.

1 Philosophical Approaches to Emergence

It is generally agreed that some properties of systems are emergent and others are not. Thus at the very least, an adequate definition of emergent needs to be *non-vacuous*, i.e. it makes room for some actual properties to be emergent, and it needs to be *non-trivial*, i.e. not all properties of whole systems are automatically emergent. Mark Bedau [1] puts the issue of defining emergence in terms of the question of whether there is room for a well-defined notion of a property which is both *autonomous* but also *constituted by* the underlying processes. If we give up on the constitution condition then emergent properties threaten to 'float free' of the underlying processes and to become mysterious or magical. If we give up on the autonomy condition then it is not clear what makes emergent properties any different from (so-called) resultant properties, such as '_ weighs 9 kilograms,' which can be properties of whole systems but which are presumably not emergent in any significant sense.

More recently, some philosophers have argued for a third criterion for any adequate definition of emergence, in addition to non-vacuousness and non-triviality, namely that the defined notion be *relevant* to scientific practice. This has led to a focus on the cluster of approaches including neural networks, agent-based models, and dynamical systems theory which has come to be referred to collectively as *complexity science*. Bedau [2003, p. 1] writes that "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."

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1984 – 1996, 2009.

2 Weak Emergence

In [1] and [2], Mark Bedau puts forward a concept he terms *weak emergence*, and argues that it meets all three of the criteria mentioned in the previous section. Bedau's definition operates against the background of what he refers to as "locally reducible systems." These are systems with some specified set of components, whose macroproperties depend only on the structural properties of these components. Given such a system S, and some property P of the system then

Definition. P is *weakly emergent* if and only if P is derivable from all of S's micro facts but only by simulation. ([2], p. 8)[1]

My goal in this paper is to examine the pros and cons of this approach to defining emergence. My particular interest will be in the epistemological ramifications of the resulting notion. Questions include: How (if at all) can weak emergent properties be predicted? How (if at all) can they be explained? How can they be described? Addressing these questions will also involve looking in more detail at the notion of simulation.

My concern in this paper is with epistemological issues concerning emergent properties, if we follow an approach along the lines of Bedau's. However there is also an issue about whether Bedau's definition is itself epistemological. For example, Symons describes Bedau's approach as an "epistemological characterization of emergent properties" [3]. I shall argue that, although it is framed in terms of the resources needed for certain kinds of derivation, Bedau's definition is not 'epistemological' in any interesting sense, for it does not make the notion of emergence relative to the cognitive or observational capacities of a particular agent. Bedau himself is clear that he intends his definition to be objective, in contrast for example to characterizing emergence based on observer surprise, as in [4]. By way of analogy, consider the standard mathematical definition of irrational numbers as numbers which cannot be expressed as the ratio of two integers. Although phrased in terms of what can (or cannot) be done by some agent, it is clearly a precise and objective definition which does not depend on the mathematical capacities of a particular agent. Thus the mere fact that Bedau's definition has a certain formulation in terms of what is required for some cognitive task does not thereby make it epistemological. It may still turn out to be epistemological, but only if the notion of simulation is definable only relative to the capacities of a particular theorizing agent.

Before discussing the extent to which Bedau's definition of weak emergence meets the three criteria listed earlier, it may be helpful to look at how the definition plays out in a particular example. Most of Bedau's own detailed examples operate against the background of the well-known cellular automata system, invented by mathematician John Conway, known as the "Game of Life." [5] The background 'universe' consists of an infinite two-dimensional array of cells, where each cell can either be 'alive' (black) or 'dead' (white). At each time-step, the state of the system is updated according to the following single rule:

---

[1] Note that in [1], Bedau gives a slightly different version of this definition: "Macrostate P of S with microdynamic D is *weakly emergent* if and only if P can be derived from D and S's external conditions but only by simulation."

> *Game of Life*: A living cell remains alive if and only if either two or three of its neighbors were alive at the previous time step; a dead cell becomes alive if and only if exactly three of its neighbors were alive at the previous time step.

Depending on the initial configuration of black and white cells, the above simple rule can lead to remarkably complex and interesting behavior as the system evolves over time. One macroproperty which Bedau considers is the property of indefinite growth. Some initial configurations in the Game of Life have no upper bound on the number of live cells that occur at future time states, while others do have an upper bound. In some cases it is easy to see which situation obtains. For example, an empty initial configuration will stay empty (since no cell will ever have any live neighbors, so the condition for cells coming alive because of live neighbors will never be fulfilled). In other cases, it is very difficult. Bedau mentions the example of the (so-called) R pentomino, a five-cell pattern that resembles a letter R, whose behavior is described by Poundstone [6] as follows: "On a high-speed computer display, the R pentomino roils furiously. It expands, scattering debris over the Life place and ejecting gliders" (p. 33).[2] If we start with a single R pentomino, do we get indefinite growth? Here is what Bedau says:

> "The only way to answer this question is to 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 … that just fits into a 51-by-109 cell region." [2]

Thus, Bedau concludes, the bounded growth of the R pentomino is a weakly emergent property in the Game of Life.

Sometimes, instead of referring to simulation, Bedau talks about "iterating the microdynamic" of the given system, and this seems to capture well what is going on in the R-pentomino case. Humphries [7] uses similar language when he describes simulation, in the sense used here, as "a step-by-step process that replicates the time development of the system at the micro level." Bedau's claim is that the only route to deriving the property of bounded growth in this case is to iterate the Game of Life update rule time step by time step until stability is reached at step 1103.

3 Circumscribing Simulation

With the above example in hand, let us now return to the question of the adequacy of Bedau's definition. With respect to the third criterion, that the defined notion be scientifically relevant, it seems clear that the definition is successful. Simulation is a key part of the contemporary study of complex systems, and emergence and complexity are closely tied up with one another. By linking emergence to simulation in his very definition, Bedau effectively builds in explicitly the scientific relevance of the resulting notion.

---

[2] A glider is another configuration of five cells which changes back into the same pattern after every 4th time step, except that the pattern is shifted one cell diagonally.

The criteria of non-vacuousness and non-triviality are harder to assess. Recall that the worry was that the notion of emergence thus defined be neither too strong nor too weak. The first step is to identify which aspects of Bedau's definition connect up to these upper and lower bounds on the strength of the defined notion. Notice the logical shape of Bedau's definition. The main clause is an 'if and only if', so it gives necessary and sufficient conditions for a property to be weakly emergent. Then, within the second part of the 'if and only if' clause, is a "can … but only." The condition that weakly emergent properties *can* be derived by simulation is intended to ensure that the defined notion is non-vacuous, since it does not magically 'float free' of the underlying processes involved. The condition that the *only* means of derivation is simulation is intended to ensure that the defined notion is non-trivial, since it rules out analytically derivable properties that are merely resultant. Thus Bedau's definition has the right form to steer a middle road between vacuousness and triviality. However – and this is crucial –, the definition depends for its cogency on there being a reasonably well-defined background notion of *simulation*. Otherwise conditions framed in terms of what can or cannot be done using simulation will be unacceptably vague. In other words, the above approach to defining weak emergence is only as good as our operating notion of simulation.[3]

The task of characterizing simulation may proceed either by trying to give a purely formal definition of simulation, or by directly considering the boundaries between simulation and non-simulation. Concerning the first approach, Rasmussen and Barrett [8] argue that "unlike computation …, simulation does not have a well established conceptual and mathematical foundation." We have already seen an implicit equation, by Bedau and by Humphries, of simulation with iteration of a system's microdynamic. Rasmussen and Barrett end up settling on a similar characterization, defining simulation as

> "an *iterated mapping* of a (usually large and complicated) system. … The simulation is an iterative system in which the simulated system is represented and its dynamics calculated." [8]

This seems to capture well various canonical examples of simulation, for example computer simulations of weather systems or of the flocking behavior of birds. But does it help in placing non-arbitrary boundaries around the core notion of simulation? I shall argue that it does not.

There are two boundaries to consider. Bedau and other philosophers sympathetic to his notion of weak emergence tend to stress the *specificity* of simulation. Every step of every microelement is calculated until the given property appears.' To stipulate that simulation is the only route to deriving the presence of a particular weakly emergent property amounts to saying that there is no 'shortcut' route to determining this fact, (cf. [2]) no way around this completely specific procedure. Sometimes this point is put in terms of the computational incompressibility of emergent properties, for example by Stephan in [9] who describes it as "incompressible unfolding." So why not just insist that proper simulation involve full specificity of this sort? There are two problems with this proposal, as I hope to show below. Firstly, even fully specific

---

[3] Note that this is not the same as demanding that the notion of simulation have completely sharp boundaries. For we may end up – for independent reasons – wanting a notion of emergence which is itself a matter of degree.

simulations are only specific relative to a particular choice of microelements, rather than being specific in any absolute sense. Secondly, many of the favored examples used to illustrate (supposedly) weakly emergent properties are not fully specific even relative to the given microelements.

4 Full Simulation

The first point, about specificity being relative rather than absolute, is obscured in the case of cellular automata such as the Game of Life because, in the first instance, these are not simulations *of* any particular real-world system. Hence the individual cells in the Game of Life can be considered absolutely fundamental and 'atomic.' Once we move to canonical examples of simulations in the sciences, however, it becomes clear that microelements do not have any privileged *ontological* status. Indeed, Bedau makes it clear that he sees the macro / micro distinction as one that can shift with different contexts:

> "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." [2]

For example, in modeling a financial market we might develop a simulation in which the microelements consist of individual investors, while in the context of simulating the rational deliberations of an individual we may take that person to be the macrosystem and, say, the individual neurons of her brain to be the microelements.

If we want to stick with the requirement that any simulation that derives the presence of a weakly emergent property be "fully specific" then it seems as if there are only two principled options. On the one hand, we could just accept that full specificity here will always just be relative to the choice of microelements. The problem here is that if nothing more is said about choice of microelements other that it "shifts with our interests" then this threatens to trivialize the whole notion of simulation. Implicit in much of the philosophical discussion of weak emergence is that microelements stand in something like a part / whole relation to their respective macrosystems. But with no constraints on which pieces of a whole system count as parts, we are free to pick all kinds of gerrymandered ways of carving up the whole.

For present purposes, the most worrying possibilities are where the whole is divided into very few 'microelements.' Consider a situation in which there is an analytic equation that governs some simple dynamical system, but there are two distinct equations for the two halves of the system. For example, the system might involve fluid flow on the surface of a sphere. In the 'northern hemisphere,' the fluid moves at a constant velocity, v, in a clockwise direction around lines of longitude, while in the 'southern hemisphere,' it moves at some different constant velocity, v'. It seems clear that this is a paradigm case where there are no weakly emergent properties. Indeed by stipulation the system is governed at each point by a single analytic equation. However, we need to consider the following line of objection: there are dynamical properties of this system which are weakly emergent because the only way to predict the future position of all the points on the surface of the sphere is by dividing the sphere into two halves and calculating the position separately for each half. And since these two halves could be taken to be 'microelements', this procedure counts as a simulation.

The proper response to this objection is to point out that it ignores a second facet of the specificity of simulations. Simulations are not merely *synchronically* specific – in that they examine the behavior of each microelement – but they are also *diachronically* specific – in that they calculate the full state of the system at each time step. In the sphere example presented above, even if there is a certain sense in which the two hemispheres can be regarded as potential 'microelements,' the equations governing each half are still analytic. Thus the positions of every point on the sphere can be calculated for some arbitrary future time without going through every (or indeed any) intermediate time steps. Of course the arbitrariness of cashing full specificity out in relation to time steps is even more obvious than for the case of microelements. Indeed in most cases the system being simulated is dynamically continuous, hence whatever discrete time steps are chosen for the simulation involves some loss of specificity relative to the original system.

This is an important point, and one that it is worth pausing for to pursue a little further. As Rasmussen and Barrett highlight, typically "the simulation and the simulated system are both dynamical systems." ([8], p. 4) Given the problems discussed above with simply making 'full specificity' relative to (possibly arbitrary) choice of microelements and time step units, a second option is to insist that – at least for the purposes of defining weak emergence – simulations be 'full' in some absolute sense. (Compare [3], which talks of weakly emergent features as "those which can be derived from the microdynamics of the system only by an exhaustive simulation.") Let us focus for the moment on synchronic specificity, and think about what "full," or "complete," or "exhaustive" simulation might amount to.

One possible response is to say that a truly complete simulation of a real-world system amounts to *duplicating* the system. In other words, a complete simulation is not a simulation at all! Though not targeted at issues of complexity and emergence, physicist Max Tegmark has recently proposed what he terms "the Mathematical Universe Hypothesis" (or MUH), which is the claim that "our external physical reality is a mathematical structure." [10] His defense of this claim is strikingly direct:

> "Whereas the customary terminology in physics textbooks is that the external reality is *described by* mathematics, the MUH states that it *is* mathematics. … We write *is* rather than *corresponds to* here, because if two structures are isomorphic, then there is no meaningful sense in which they are not one and the same." [10]

This argument seems too quick. Tegmark's background assumption is that the universe is fully describable in terms of its structural properties, since any non-structural properties would make no observable difference. But even if one grants this (far from innocuous) assumption, it does not follow that the universe is no more than a complex mathematical structure. Stewart Shapiro [11] makes a useful distinction between structures which are *freestanding* and those which are not. Freestanding structures are basically those for which full simulation and instantiation coincide. Take the game of chess. Any arrangement of objects which conforms to the structural and dynamical constraints of the rules of chess is a game of chess. It does not matter what the pieces are made of, or how quickly the game proceeds. Chess games can be played with pieces of plastic, or pieces of ivory, or using people as 'pieces.' It can be played at a lightning pace, or by mail over a period of years. Chess can also be played on computers and by computers. But it would be odd, in fact not just odd but false, to say that a computer was merely 'simulating' the playing of chess. When the computer Deep Blue defeated then world chess champion Gary Kasparov in the famous 1997 match, Deep Blue was literally playing chess. Thus chess is a freestanding structure.

Shapiro thinks that many structures are non freestanding. One example he gives is of a certain defensive arrangement in baseball. It is not enough, in order to instantiate this defense, merely to have objects of some kind or other at the appropriate locations on a baseball field. Thus a structure consisting of piles of rocks at the different locations is not a baseball defense. Against Shapiro's claim here, one might argue that he is ignoring precisely the distinction that we have been targeting, between simulation and *full* simulation. Thus Deep Blue counts as literally playing chess because Deep Blue is fully simulating the structure of the game of chess. By contrast, the pile of rocks on the baseball field, while it might count as a simulation of some sort, is not fully simulating a baseball defense, because there are various structurally relevant properties that are omitted (for example, the ability to move the ball from one position on the field to another). So there are not two kinds of structures here, freestanding and non-freestanding, but rather two kinds of simulation, full simulation and partial simulation.[4]

Lest we get too far afield from our original topic, I will not pursue the above issues any further. However, it is worth pointing out that – whether or not full simulation is the same as duplication – full simulation in this absolute sense is impossible as a practical matter for the vast majority of real-world systems that are of interest to complexity scientists. Hence it would likely violate the third criterion mentioned earlier, that the notion of weak emergence being defined be relevant to scientific practice, for it to be linked to any such 'absolute' sense of simulation.[5]

5 Between Simulation and Analytic Derivation

One could claim that I have been unduly pessimistic about the first of the two options discussed in Section 4, since it is not obvious that the only alternative to absoluteness is arbitrariness. In other words, it would seem that there is room to argue for a notion of "full simulation" which is relative to choice of microelements and time units, but where these are picked out based on objective features of the system being simulated. (An analogy here might be with approaches to anchoring the objectivity of inductive reasoning in science by reference to some background ontology of natural kinds.) I am not sure whether this approach could be made to work, and if so how, but for sake

---

[4] There may still be a derivative sense of "freestanding structure" as a structure for which full simulation is possible. And it is worth noting that the points made about the game of chess seem to apply with equal force to the Game of Life. Specifically, there seems to be no distinction between simulating the Game of Life and instantiating it. Also note the slide between simulation / observation when Bedau [2] talks about the R-pentomino's bounded growth.

[5] Bedau [2] does distinguish between different 'strengths' of simulation, e.g. between finite and infinite simulations. But it is unclear how this intersects with the above issues.

of argument let us assume that it can. Plugging the result back into Bedau's definition yields a notion of weak emergence according to which weakly emergent properties can only be derived by a simulation that is completely specific relative to the given (non-arbitrary) microelements and time units.

Even granting all of the above assumptions, however, there are still potential problems with this line of analysis. For illustrative purposes, let us return to the earlier example from Bedau of the R pentomino in the Game of Life and whether it exhibits unbounded growth. Recall that, as Bedau describes the situation, the only way to see that growth in this case is bounded is to iterate the microdynamic for 1103 time steps until a stable pattern appears. Hence, the bounded growth of the R pentomino is a weakly emergent property of it.

At first glance, this seems to fit with the definition of "full simulation" that was presented above, since the process of establishing stability is completely specific: the color of each cell on the grid is calculated for each time step until step 1103. But at second glance, the situation is not so clear-cut. In particular, there is room to argue that the 'simulation' is neither synchronically nor diachronically exhaustive. From a synchronic perspective, not every cell is updated individually at each time step. Indeed this is impossible from a practical perspective since the Game of Life takes place on an infinite grid! More prosaically, given the update rule we know that any 'dead' cell that is surrounded by dead cells itself remains dead at the next time step. So the infinite array of empty cells surrounding the R-pentomino initial condition can simply be ignored for the purposes of simulation.[6] From a diachronic perspective, there is indeed complete specificity for the first 1103 time steps of the simulation, but at this point the simulation ends and a different form of reasoning takes over. For it is not – and cannot be – purely on the basis of simulation that we come to know that the state at time step 1103 is stable. It turns out that this state is made up of a collection of two sorts of sub-patterns: "still lifes" are a 2x2 blocks of living cells, which remain unchanging forever given the update rule for the Game of Life; "blinkers" which are vertical strips of three living cells, which alternate between a vertical 3-strip and a horizontal 3-strip at successive time steps.

I want to argue that our knowledge of the stability of still lifes and of blinkers is not based on simulation. If it was, then according to Bedau's analysis, the property of unchangingness of the *still life* pattern would count as weakly emergent, and this seems to go against the whole spirit of his approach. If I am right, then our knowledge of the bounded growth of the R pentomino is based partly on simulation and partly on more traditional mathematical analysis. As Bedau himself puts it, albeit not when talking about this example,

> "[I]n 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 a certain macro property arbitrarily far into the future." [2]

This poses an obvious problem for a notion of weak emergence based on full simulation, since we are forced to conclude that the property in the R-pentomino case

---

[6] More precisely, any empty cell that is at least two cells removed from a live cell can be ignored at a given time step.

is not weakly emergent because it can be (indeed it must be) derived without *fully* simulating the system.

There are at least a couple of responses that could be made here on behalf of the defender of the full-simulation definition. One response is to argue that recourse to analytic methods, though possible in the R-pentomino case, could not have been predicted in advance except by going through the simulation up to time step 1103. Hence, from an epistemological point of view, the simulation is still an indispensable part of the derivation process. A second response is to maintain that the above argument equivocates between two conditions, and thereby conflates 'being derivable only by full simulation' with 'being fully derivable by simulation.' It is the former of these two conditions which appears in the revised version of Bedau's definition of weak emergence, and this condition (unlike the second condition) is fully compatible with the full simulation being augmented by other modes of reasoning.

The effectiveness of both of these defenses may hinge in part on the nature of the (putatively) weakly emergent property being discussed. In particular, some properties involve implicit existential claims about the future state of the system, while other properties involve implicit universal claims. To take a simple case, the property of "stability" in the Game of Life involves an implicit claim that is existential, namely that there is some future state of the system which is identical to an earlier state (which thereby guarantees that the system has entered into a closed cycle). The stronger property of "unchangingness" involves an implicit universal claim, that all future states of the system are identical to the current state. The property of "bounded growth," which is the property that Bedau focused on initially in his discussion of the R-pentomino case, is more complicated. Both stability and unchangingness are sufficient for bounded growth, but neither is necessary. To see why not, recall from our earlier discussion of the Game of Life that there is another five-cell arrangement called a "glider" which changes back into the same pattern after every fourth time step, except that the pattern is shifted one cell diagonally. Consider an initial configuration consisting only of two gliders 'pointed' in opposite directions. As the system evolves over time, the two gliders move gradually off to infinity. No state is ever identical to any previous state, yet there is clearly an upper bound to the number of live cells that ever appear in any future state.

As a parenthetical point, it should be noted that there are in fact two distinct properties that fit the term "bounded growth." In my discussion thus far, I have been assuming that growth is bounded if and only if there is some upper bound on the number of live cells in any (individual) future state of the system. However, the fact that Bedau talks of the stable state reached at step 1103 as "just fitting into a 51-by-109 cell region" indicates that he himself may have a different property in mind, such that growth is bounded if and only if there is some finite area outside of which no live cells of any future state appear. The first of these two readings yields a simpler property than the second, since spatial boundedness implies numerical boundedness, but not vice-versa (as the 'double-glider' example above illustrates).

My general claim – which I shall not further argue for here – is that properties associated with implicit universal claims cannot be derived by full simulation. The only way to justify a categorical claim about *all* future states of an open-ended system is through analytic methods. Yet there seems no reason why properties of this sort cannot be weakly emergent, indeed "bounded growth" – in the second sense specified above – seems to be a good candidate for just such a property. The key point is that for universal-linked properties, there is no sense of running a simulation and 'waiting for the first appearance of the relevant state.' Of course if we abandon the insistence on full simulation in our definition of weak emergence, then there is still clarificatory work to be done. The main challenge, once we allow in some abstraction to our simulations of weakly emergent properties, is to show a principled line can then be drawn between simulation on the one hand and fully analytic methods on the other.

6 Epistemological Ramifications of Derivation

I argued earlier, in Section 2, that Bedau's definition of weak emergence is intended to carve out a notion which is objective and independent of the cognitive capacities of any particular observer. The main issue with which we have been grappling in the intervening sections is whether an appropriate notion of "full simulation" can be circumscribed which gives teeth to Bedau's definition. In the remainder of the paper, I want to look more briefly at the epistemological features of weak emergent properties thus defined. In other words, what does it mean for such core epistemological tasks as prediction and explanation for a property to be derivable only by simulation?

I shall start with prediction, since there is a fairly clear link, at least *prima facie*, between deriving the presence of a property and predicting that the property will occur. Indeed, weak emergence is sometimes defined directly in terms of prediction. For example, Humphries interprets Bedau as equating weak emergence with computational incompressibility, and writes that "prediction of future states of computationally incompressible systems must run through each of the intermediate time steps between the initial state and the predicted state." ([10], p. 4) However, there are at least a couple of reasons for resisting any straightforward equivalence between a property being derivable and a property being predictable. Firstly, derivability is a logical notion and the bare existence of a derivation from given premises does not guarantee anything about the length or complexity of the required derivation. Hence weakly emergent properties may not be predictable *by a particular epistemic agent* if the complexity of the required simulation outruns that agent's cognitive capacities. Secondly, and even more importantly, even if an agent is in possession of a valid derivation of the presence of a given property, this will not be convertible into a genuine prediction unless the agent believes that the premises of the derivation are true.

Reflection on this point puts pressure on a different facet of Bedau's definition. The claim is that weakly emergent properties are derivable, but only by simulation. Thus far we have been assuming that the space of alternative methods to simulation, especially analytic methods, are themselves well-defined and legitimate. But derivations, even deductively valid, 'rigorous' derivations, come in many guises. Consider, for example, the following derivation, D*, of the bounded growth of the R pentomino.

D*

(D1) The initial configuration is a (single) R pentomino
(D2) If the initial configuration is a R pentomino then growth is bounded
(D3) Hence, growth is bounded

D* is clearly a valid, albeit trivial derivation of the presence of the property of bounded growth. And, equally clearly, it does not involve any simulation. So what prevents this from being a counterexample to the claim that the only way to derive the property of bounded growth for the R pentomino is by simulation?

There are a couple of ways of trying to impose conditions on acceptable derivations so as to exclude 'trick' cases such as D*. One way is to impose some sort of epistemic constraint on the premises of the derivation, for example that they be justifiably believed by whatever agent is making the prediction. Adopting this approach, D*-style cases would fall into two sorts. Either the agent has no justification for believing the truth of premise D2, in which case the derivation fails to count as legitimate. Or the agent is justified in believing D2, in which case – presumably – this justification has come either from the agent having already run a simulation, or from having been informed of the truth of D2 by someone else who has run an appropriate simulation. In either case, the derivation is legitimate but it rests, ultimately, on simulation. While this approach might therefore work as a way of excluding spurious alternative means of derivation, I suspect that it may run into trouble by also excluding many kinds of simulation. For simulations standardly begin with various simplifying assumptions – discrete time steps, simultaneous updating, division of the space into equal cells, and so on – which are typically literally *false*. As such, the 'premises' of most simulations will not be believed to be true by the agent who implements them, and if they do happen to be believed then this belief will not be justified.

A second, and somewhat different, approach to ruling out D* from contention is to argue that fails to be an appropriately *general* form of argument. In other words, there is nothing in D* which indicates what the correct answer to the bounded growth question would be for other initial configurations, nor how the R-pentomino behaves with respect to other properties. Clearly the contrast here is supposed to be with analytic derivations which typically involve equations where different figures can be substituted in to yield predictions of different phenomena. Simulations also tend to be applicable across a range of initial conditions and for a wide variety of different kinds of properties. To make this approach work, more needs to be done to clarify just what it is that weak emergence applies to. I will not undertake this work here except to note that on Bedau's account, despite it mostly being summarized as being about properties, weak emergence seems to come out as a *three*-place relation. In the R-pentomino case, for example, what Bedau is concerned to show is that, in the Game of Life, the property of bounded growth, for the initial configuration of an R pentomino, is weakly emergent. So we have here a three-place relation between system, property, and initial configuration. The insistence on generality of derivation amounts to requiring (at minimum) that a given derivation be applicable to more than just one of these triples.

While generalizability is clearly a valuable property of derivations, it is much less clear that it makes sense as a necessary condition for adequate prediction. Mathematicians, for example, tend to prefer generalizable proofs, but they seem to value such proofs for their explanatory potential as opposed to any extra justificatory power. Indeed, ultra-specific proofs are still taken to provide a perfectly adequate justification of their conclusions. This provides a nice segue into the second facet of derivability that I want to examine, namely the relation between the presence of a property being derivable only by simulation and its explicability (or lack thereof).

It is worth recalling that the main form of "autonomy" that Bedau credits to weakly emergent properties is *explanatory autonomy*. These properties are ontological and causally reducible to their underlying microproperties, but they are nonetheless explanatorily autonomous. Symons agrees that one should "distinguish computational models or explanations from reductions." ([3], p. 2) It certainly seems correct that not all derivations are also explanations. As above, one way to see this point is by considering different kinds of proof in mathematics. There are some well-known mathematical proofs, for example the proof of the 4-Color Theorem, which were carried out with the aid of computers because they are highly disjunctive and require going individually through many thousands of subcases. Aside from worries about the very involvement of computers in this process, most mathematicians find the proof of this result unsatisfying because they do not see it as explaining *why* the theorem holds.

Not only does Bedau argue for the explanatory autonomy of weakly emergent properties, in his 2003 paper he also distinguishes between two ways in which this explanatory autonomy can manifest itself. On the one hand, there are "accidental" weakly emergent properties of a system, where our explanations in terms of macrofeatures are required by the complexity of the underlying processes, but where this explanatory autonomy is "merely epistemological." On the other hand, there are "robust" weakly emergent properties whose macro-explanations are autonomous in a deeper sense. To illustrate the distinction, Bedau again turns to his favorite milieu of the Game of Life. His example of a robust property is the configuration of cells known as a "glider gun." As the name suggests, this configuration produces a steady stream of gliders that move at an even spacing in one particular direction. He contrasts this with "an irregular collection of still lifes, blinkers, and miscellaneous piles of 'muck' that happens to emit six gliders." [2] In this latter case, Bedau claims that the property of producing a stream of gliders is "accidental." He cashes out difference in counterfactual terms. In the case of the glider gun, but not in the case of the irregular configuration,

> "the aggregate micro explanation … omits information [that] supports counterfactuals about the stream. … If those micro histories had been different, the macro explanation could still have been true." [2]

Whether proper sense can be made in this context of the notion of an "accidental" property is, I think, an important and interesting question. (One intuitive reaction, for example, is that in a fully deterministic system such as the Game of Life, *nothing* is accidental …!) I suspect that at best the distinction Bedau is pointing is a matter of degree. For, aside from its greater complexity, I see nothing qualitatively different between the behavior of the 'miscellaneous piles of muck' and of the glider gun. Bedau writes, in the course of cashing out his counterfactual analysis of the latter, that "the same glider stream would have been produced if the configuration had been changed in any number of ways, as long as the result was a gun that shot the same kind of gliders." [2] But this counterfactual condition seems almost completely circular, and a precisely parallel point could be made about the miscellaneous piles of muck – gliders would still have been produced as long as the same miscellaneous piles lay at the core of the initial configuration of cells.

The second place where I part company with Bedau concerns the role of the microlevel behavior in the glider-gun example. According to Bedau this detailed causal history does explain the stream of gliders, it is just that the macro-level gives a better explanation and captures features linked to generality and robustness that are left out of the micro-level explanation. My own view is that in this sort of case, the micro-level story does not *explain* the macro behavior at all. It can be used to predict this macro behavior, at least in principle, but it plays no part in the proper explanation of it.

7 Conclusions

My concern in this paper has been to examine an approach to emergence that defines emergent properties in terms of what can – and can only – be derived by simulation. If such a definition is to be objective and to have real content then this puts pressure on the notion of simulation. In particular, without reasonably precise boundaries between simulation-based and non-simulation-based techniques, the definition threatens to collapse into uselessness. I have argued that, although the core notion of simulation is well-understood, and although canonical cases of simulation are uncontroversial, placing principles bounds on it is surprisingly difficult to do. If we are successful, then the prospects for the resulting notion of weak emergence charting a path between being 'mysteriously' strong and being trivially weak look promising. While there is every hope that such a notion will be coherent and scientifically relevant, there remains work to do in delineating the relation between possessing a simulation-based derivation of the presence of a property and being able to predict it or explain it. Exploring further the epistemological ramifications of this style of definition will provide fruitful terrain for future research.

References

1. Bedau, M.A.: Weak Emergence. Philosophical Perspectives, Mind, Causation, and World 11, 375–399 (1997)
2. Bedau, M.A.: Downward Causation and the Autonomy of Weak Emergence. Principia Revista Internacional de Epistemologica 6, 5–50 (2003)
3. Symons, J.: Computational Models of Emergent Properties. Minds and Machines (forthcoming)
4. Ronald, E.M.A., Sipper, M., Capcarrère, M.S.: Design, Observation, Surprise! A Test of Emergence. Artificial Life 5, 225–239 (1999)
5. Berlekamp, E.R., Conway, J.H., Guy, R.K.: Winning Ways for Your Mathematical Plays, vol. 2. Academic Press, New York (1982)
6. Poundstone, W.: The Recursive Universe. Contemporary Books, Chicago (1985)
7. Humphries, P.: Synchronic and Diachronic Emergence. Minds and Machines (forthcoming)
8. Rasmussen, S., Barrett, C.L.: Elements of a Theory of Simulation. In: Morán, F., Merelo, J.J., Moreno, A., Chacon, P. (eds.) ECAL 1995. LNCS, vol. 929. Springer, Heidelberg (1995)
9. Stephan, A.: The Dual Role of 'Emergence' in the Philosophy of Mind and in Cognitive Science. Synthese 151, 485–498 (2006)
10. Tegmark, M.: The Mathematical Universe, arXiv: 0704.0646v2
11. Shapiro, S.: Philosophy of Mathematics: Structure and Ontology. Oxford University Press, Oxford (1997)