**Strengthening Weak Emergence**

Nora Berenstain

Forthcoming in *Erkenntnis*

*(please cite final version)*

**Abstract**:I offer an improved version of Bedau’s influential (1997) account of weak emergence in light of insights from information theory. Bedau analyzes weak emergence in terms of the non-derivability of a system’s macrostates from its microstates except by simulation. However, non-derivability alone does not guarantee that a system’s macrostates are weakly emergent. Rather, it is non-derivability plusthe algorithmic compressibility of the system’s macrostates that makes them weakly emergent. I argue that the resulting information-theoretic picture provides a metaphysical account of weak emergence rather than a merely epistemic one.

**1. Introduction**

Weak emergence provides a potentially useful conceptual framework to make sense of chaotic phenomena and systems whose behaviors can be modeled with chaos theory. Chaos theory, as a branch of mathematics, models dynamical systems whose behaviors demonstrate sensitive dependence on initial conditions. While the behavior of chaotic systems often appears random, it is actually determined by underlying rules and patterns. A chaotic complex system is thus usually understood as a nonlinear system whose path through phase space appears random while depending on deterministic rules or laws. Some examples of systems that may demonstrate chaotic behavior are stock markets (Hsieh 1991), the movement of krill herds (Saremi, Mirajalili, and Mirajalili 2014), weather patterns such as turbulence, cardiac arrhythmias (Oestreicher 2007), the motion of a magnetic pendulum over a plane containing two or more magnets, and the fractal development of landscape features such as rivers and clouds (Lovejoy 1982). Chaos theory is an interdisciplinary area of study, as systems demonstrating chaotic properties span a remarkably wide disciplinary range.

The concept of weak emergence is useful for making sense of chaotic phenomena because it aims to capture the relation between the behaviors of complex systems and the foundational physical states from which they arise. It originated as a notion that could tie biological systems to processes of computation and the limits of computability. Bedau (2013) suggests that the discipline of synthetic biology, for instance, can be understood as the practice of engineering weak emergence and that this explains why synthesis is so important for predicting life’s emergent properties. Bedau (2003) has also argued that weak emergence illuminates the notions of complexity invoked in biology and psychology and that it sheds light on the methodologies of modeling approaches such as a neural networks and agent-based models. The models in complexity science are so bound up with presumptions of emergence that Bedau suggests that, “one could fairly call the whole enterprise the science of emergence” (2003, 1). Weak emergence is a fitting concept for complexity science since it aims to bridge the metaphysical gap between physical systems and the computer simulations that model them, and it does so by applying the notion of computational derivability to physical systems.

The concept of weak emergence comes from Bedau’s influential (1997) work, in which he purports to offer a realist yet simultaneously metaphysically ‘innocent’ picture of emergence. One of his goals is to avoid the strong emergentists’ metaphysical commitments to such mysterious things as emergent substances and primitive high-level causal entities. Strong emergentists usually accept the idea that if a phenomenon is strongly emergent then it reflects the presence of a new metaphysical substance that is not unified with the underlying physical stuff which gave rise to it. Epstein (1999), who does not accept the view, describes it as an attempt to “preserve a ‘mystery gap’ between the macro and the micro.” In order to avoid a metaphysical commitment to strong emergence, Bedau aims to capture the conflicting intuitions that emergent phenomena are somehow both independent of yet constituted by the underlying processes that create them. His account identifies macrostates of a system as weakly emergent if they cannot be derived from the system’s initial state and its dynamics except by simulation. While Bedau intended his account of weak emergence to be a metaphysical one, because of the central role that ‘non-derivability’ plays in the account, some such as Wilson (2013) have objected that weak emergence can only be understood as epistemic and not as metaphysical

This paper provides a metaphysical interpretation of weak emergence after offering a revision to Bedau’s picture. The revision I suggest addresses the fact that Bedau’s original account allows for macro-level randomness to count as weakly emergent. Bedau identifies an important connection between weak emergence and nonlinear systems whose macrostates cannot be predicted except by simulation. However, he misidentifies the non-derivability of the system’s behavior as the singular source of its weak emergence. I argue that Bedau must add a criterion of algorithmic compressibility to his picture in order to produce a substantive, metaphysically realist account of weak emergence. Rather than the non-derivability of a system’s macrostates alone guaranteeing that they are weakly emergent, it is the non-derivability *plus* the algorithmic compressibility of the system’s macrostates that make them weakly emergent. This revised account avoids the problem of counting macro-level randomness as weak emergence.

After explaining the need for such a revision, I defend the claim that the resulting account of weak emergence can and should be interpreted as metaphysical. The metaphysical interpretation of weak emergence offered here draws from two separate philosophical approaches rooted in an information-theoretic approach to ontology. The first is Dennett’s (1991) account of real patterns. Like Bedau, Dennett tries to make sense of seemingly emergent objects that arise within the cellular automaton of Conway’s Game of Life. The criterion of algorithmic compressibility is the focus of Dennett’s view. He claims that his picture is neither reductionist nor emergentist and takes it to be a middle road between realism and anti-realism. However, his view backslides into instrumentalism, as he defines facts about algorithmic compressibility in such a way that they are inherently interest-relative. Ladyman & Ross’s (2007) information-theoretic structural realism draws from Dennett’s Real Patterns picture but rejects the interest-relativity of algorithmic compressibility. Ladyman & Ross add a criterion of projectibility and a criterion of non-redundancy in order to develop Dennett’s view into a realist criterion of ontological reality.

I draw on these accounts in order to respond to objections that a picture of weak emergence expressed in information-theoretic terms can be understood only as epistemological rather than metaphysical. I argue that adding a criterion of compressibility to Bedau’s picture fits with his goal of offering a metaphysical account of weak emergence. I explore formulations of algorithmic compressibility in terms of Kolmogorov complexity, which is universal and neither purely interest-relative nor relative to a choice of computer. The resulting objective, universal definition of algorithmic compressibility provides the basis for an account of weak emergence that is in line with the goals of Bedau’s original project and can be understood as metaphysical.

**2. Bedau’s Weak Emergence**

Weak emergence connects complex systems and their foundational physical states by conceiving of features of physical systems in terms of the computations designed to simulate them. Accounts of weak emergence must avoid pitfalls on both the left and the right: 1) loss of metaphysical innocence, and 2) not being metaphysical to begin with. The first is an issue because the loss of metaphysical innocence is taken to make the phenomenon of weak emergence incompatible with physicalism. If an account of weak emergence is not metaphysically weak enough, it will end up saddling its proponents with the same ontological commitments as strong emergence, the avoidance of which was an initial motivation for the view. Or it might be so weak as to not be metaphysical at all and end up only picking out our cognitive or epistemic limits with respect to complex systems, as Wilson (2013) and Theurer (2014) have argued.

Bedau’s goal is to produce a picture of weak emergence that is metaphysically innocent, consistent with materialism, and scientifically useful. He intends his picture to capture the two main criteria that are commonly taken to characterize emergent phenomena but are also seemingly in tension with one another. These are that “emergent phenomena are somehow *autonomous* from underlying processes” and that they are “somehow *constituted by,* and *generated from*, underlying processes” (1997, 375). While these two criteria may be in tension and “raise the specter of illegitimately getting something for nothing,” Bedau emphasizes that an account of weak emergence must explain away the apparently illegitimate metaphysics that haunted British Emergentism. He takes his view to fulfill these criteria, and he illustrates the view with an analysis of Conway’s Game of Life.

This famous cellular automaton takes place on a two-dimensional infinite grid and evolves according to four deterministic update rules. The grid has infinitely many columns and rows, each of which are infinite in both directions. In the Life world, time is discrete. Each cell on the grid is in one of two positions – on or off, black or white, alive or dead. The initial conditions of the cells are determined randomly. Once determined, each cell updates its state according to the four rules that govern the Life universe:

1. Any live cell with fewer than two live neighbors dies.
2. Any live cell with two or three live neighbors stays alive.
3. Any live cell with more than three live neighbors dies.
4. Any dead cell with exactly three live neighbors becomes a live cell.

Each cell has eight neighbors, which include the four cells next to it and the four cells catty-corner to it. The four update rules and the initial conditions of the grid are the only things that determine the future states of the Life grid. The update rules are linear and deterministic. Yet what we see emerge on the grid is a fascinating array of complex patterns and seemingly autonomous entities. *Traffic lights* blink on and off. *Glider guns* set up shop in a stable location on the grid and produce *gliders* that move diagonally across the grid until they collide with *eaters* and are destroyed. The *pulsar*, a large oscillator composed of four mutually stabilizing quadrants, flickers through its three cell arrangements in a never-ending cycle. The *snacker*, a period 15 oscillator, throws off domino sparks that can convert blocks into gliders and flips the orientations of passing *toads*. The Life universe is characterized by a vast array of busy yet stable entities that materialize from the small finite set of update rules applied to a randomly selected set of initial conditions.

Bedau emphasizes that while the system of the Life universe is relatively simple, there is no way to predict the future states of these emergent entities without simply letting the game run and watching what happens. He writes, “With few exceptions, it is impossible without simulation to derive the behavior of any macrostate in a Life configuration even given complete knowledge of the configuration.” (386) This feature of the Life universe provides the foundation for Bedau’s notion of weak emergence, which he defines as follows (1997, 378):

Where system S is composed of micro-level entities having associated microstates, and where microdynamic D governs the time evolution of S’s microstates,

Macrostate P of S with microdynamic D is weakly emergent iff P can be derived from D and S’s external conditions but only by simulation.

Bedau characterizes external conditions as those that are “outside” the system. The only external conditions in a closed system with a deterministic microdynamic are the initial conditions. If the system is open, then contingencies regarding the flux of parts and states through the system also amount to external conditions.[[1]](#footnote-1) When the system’s microdynamic is nondeterministic, every accidental effect counts as an external condition. Thus, a macrostate of some system is weakly emergent iff it cannot be derived from the system’s initial conditions plus update rules except by simulation. Non-derivability without simulation—understood as derivability only by simulation—is the key notion in this picture of weak emergence.

Bedau’s claim that this account of weak emergence is metaphysical rests on his interpretation of the modal terms in his definition as metaphysical rather than epistemological. “For P to be weakly emergent what matters is that *there is* a derivation of P from D and S’s external conditions and *any* such derivation is a simulation. It does not matter whether anyone has discovered such a derivation . . . Our need to use a simulation is due neither to the current contingent state of our knowledge nor to some specifically human limitation” (379). He emphasizes the objective nature of the facts about whether or not there is a derivation by simulation of the system’s macrostates from its external conditions and microdynamics. For Bedau, this is what makes his account a metaphysical one.

For Wilson (2013, 2019), Bedau’s reasons are unconvincing. Wilson questions why *non-derivability without simulation* shouldbe taken to characterize a metaphysical rather than merely epistemological account of emergence. Wilson takes Bedau’s criterion to be compatible with ontological reduction of a system’s macrostates and believes that this compatibility precludes his account of weak emergence from being metaphysical. I return to this objection after presenting an objection to and improvement upon Bedau’s account.

**3. The Objection from Randomness**

Bedau offers what he takes to be necessary and sufficient conditions for weak emergence. However, the conditions he offers are in fact not sufficient. Bedau does not satisfactorily delineate which algorithmically incompressible macrostates his view counts as weakly emergent. Imbert (2005) for instance, notes that Bedau’s equation of weak emergence with non-derivability except by simulation results in a number of what he calls ‘deceptive properties’—those that are built out of any conjunction of microproperties—being counted as weakly emergent. A similar issue that deserves attention is that, without a criterion of algorithmic compressibility, Bedau’s picture allows random patternlessness at the macro level to count as weakly emergent so long as it is not derivable from the initial conditions of the microstates plus the microdynamics except by simulation.[[2]](#footnote-2) In the Game of Life and in other systems, there are many such macro states that are not derivable from microstates and microdynamics except by simulation. But these macrostates exhibiting random patternlessness are not those that we aim to capture with a picture of weak emergence. They do not exhibit interesting self-organizing behavior, nor do they call for explanation. That these states are considered weakly emergent under Bedau’s account shows that it is too broad.

Humphreys (2008) considers two types of cases that can motivate the objection from randomness. The first involves a cellular automaton transforming from a random initial state to a random final state with each state in between also being random. Humphreys emphasizes that “as long as the computational connections between the instances are of the appropriately complex kind,” then such a case will show that, given Bedau’s account, “a process within which all states are random can be a candidate for producing weakly emergent states” (591). The second type of case involves a cellular automaton moving from an initially ordered state into a final random state. Humphreys notes that “an appeal to the motivations behind self-organizing systems as producers of emergent states might in this case incline as toward insisting that this should not count as a case of emergence.” While Humphreys notes that such objections are motivated by considerations beyond those that are purely formal or internal to Bedau’s conception, they are still compelling given that Bedau is trying to capture a type of phenomenon that arises in the physical world rather than only in models.

How ought Bedau’s account be strengthened in order to rule out as weakly emergent problematic properties and macrostates that are purely random? In the next sections, I clarify the distinction between non-derivability and algorithmic incompressibility in order to answer this question. I discuss how Dennett’s notion of real patterns can be invoked to capture what is interesting and important about the weakly emergent macrostates of complex systems. Finally, I demonstrate how a criterion of algorithmic compressibility can be introduced in order to strengthen Bedau’s account and rule out random macrostates from those that count as weakly emergent.

**3.1 Incompressibility and Non-derivability**

What exactly is algorithmic incompressibility? First, it is important to note that non-derivability without simulation is distinct from algorithmic incompressibility. Algorithmic incompressibility entails non-derivability but non-derivability does not entail algorithmic incompressibility. In this section, I explain what these information-theoretic properties are and how they relate to Dennett’s concept of real patterns.

According to Dennett, a pattern is real only if there is a compressible algorithm of it. And a pattern has a compressible algorithm if and only if there is a description of it that is shorter than its point-by-point bitmap. Dennett (1991, 32) invokes Chaitin’s definition of mathematical randomness to elucidate his idea: “A series (of dots or numbers or whatever) is random if and only if the information required to describe (transmit) the series accurately is incompressible; nothing shorter than the verbatim bitmap will preserve the series.” Complete randomness is algorithmic incompressibility; anything with a description shorter than the bitmap is a real pattern.

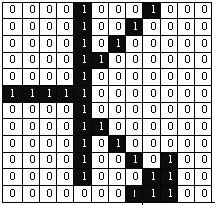


Figure 1. A patterned, non-random bitmap

Consider the above figure. Suppose we label the rows A-L from top to bottom and label the columns 1-12 from left to right. The bitmap description of the grid would be 144 bits long and consist of a sequence A1-0, A2-0, A3-0, A4-0, A5-1, . . . and so on. But there are many ways to reproduce this exact grid using algorithms that are shorter than the bitmap. Another way we can give a description shorter than the bitmap is by using a program that gives all cells a value of 0 unless otherwise noted and explicitly identify the cells that should receive a value of 1. Because there are 28 black cells, this description would take only 29 bits and would thus be more efficient than the bitmap. There are numerous ways that we can compress the information contained in this grid. The fact that there are more white cells than black cells in the grid guarantees that it can be produced by strings shorter than the bitmap. That it can be produced by strings shorter than the bitmap means that the data it contains are compressible, since the amount of information necessary to transmit or store the pattern can be reduced. Thus, algorithmic compressibility as a property of a system’s macrostate can be understood as follows: A system’s macrostate is algorithmically compressible when it can be fully reconstructed from a description whose string is shorter than the macrostate’s bitmap. Ladyman, Lembert, and Wiesner explain algorithmic compressibility thusly:

It is easiest to think in terms of binary strings and to associate every object (numbers, vectors, lists, etc.) with a string that describes it. . . Any repetition of digits or symmetry in the string will allow the programme that outputs the string to be shorter than the string itself. The more compressible the string, the shorter the programme. A random sequence is maximally incompressible while a string of all zeroes, say, is maximally compressible. (2013, 13)

The fact that the grid pictured above is algorithmically compressible reflects a further fact about it, namely that the layout of the grid is nonrandom. The only type of grid for which there are no descriptions shorter than the bitmap available are those that have a completely random distribution of black and white cells. Dennett explains, “A series is not random––has a pattern––if and only if there is some more efficient way of describing it” (1991, 32). Algorithmic compressibility is a property that is on a gradient, as is randomness. One pattern can be more or less compressible than another, and it can be more or less random. More randomness corresponds to more noise and less compressibility; more compressibility corresponds to less noise and less randomness.[[3]](#footnote-3) The only states that are not algorithmically compressible are those that are completely random.

**3.2 Algorithmic Compressibility and Weak Emergence**

How do randomness and non-randomness, compressibility and incompressibility relate to non-derivability in general? And how are they relevant to Bedau’s account of weak emergence in particular? Non-derivability except by simulation entails neither algorithmic compressibility nor incompressibility. That a system’s macrostates can be derived only from its microstates and microdynamics by simulation is compatible with those macrostates being algorithmically compressible *or* incompressible. It is thus not the case that only interesting and puzzling macrostates such as those containing blinkers and gliders fit Bedau’s criteria of weak emergence. Random macrostates of cellular automata such as the Life universe can also be derived from the initial conditions plus the update rules—but only by simulation.

The issue, as noted in the previous section, is that these states do not exhibit interesting patterns or behaviors that call for explanation. What is special about weakly emergent macrostates is that they are *real patterns* that are not derivable from their initial conditions and microdynamics except by simulation. There are plenty of random sequences and states that also cannot be predicted from their initial conditions and microdynamics except by simulation. But weakly emergent macrostates can be produced by an algorithm that describes them in less information than their bitmap. This is what makes weak emergence interesting and worth theorizing. The macrostates arising from the microstates of weakly emergent systems are not merely random configurations but codifiable patterns.

Since weak emergence is widespread and found in many different systems and phenomena, it should not be conceptually limited by an a priori commitment to its rarity. However, it should apply only to genuine phenomena or, in other words, to non-random patterns of information. Requiring that a system’s macrostates must themselves be algorithmically compressible in order to be weakly emergent captures the sense that something has gone wrong when random macrostates are classed as weakly emergent. This requirement, that a system’s macrostate be describable in a string shorter than the bitmap, is still a very low threshold to meet. There is no requirement about the level of compressibility a state must exhibit, just that it must exhibit some compressibility.[[4]](#footnote-4) Only purely random sequences are ruled out.

Note that this account of weak emergence as non-derivability + algorithmic compressibility also answers Theurer’s (2014) objection that computational complexity cannot provide a basis for emergence because it renders complete randomness maximally complex. Theurer draws on Ladyman, Lembert, and Wiesner’s (2013) work, which argues that “Kolmogorov complexity is not an appropriate measure of complexity in a physical system because it yields measures that are maximal for completely random strings. Shannon complexity suffers from the same problem” (Theurer 2014). However, by adding the criterion that weakly emergent phenomena meet a minimal threshold of algorithmic compressibility, this concern is rendered null. While completely random patterns would still be considered maximally complex by this measure, that would not be enough to make them weakly emergent on the view proposed here.

Weakly emergent phenomena are interesting in part because we can predict facts about their future behavior. We cannot make such predictions by attending solely to the microstates and update rules, but we can predict their future behavior if we attend to the macro-scale at which such phenomena arise. This illustrates another reason that purely random macrostates should not be taken to be weakly emergent—nothing about them can be predicted at the macro-scale. As Dennett emphasizes, “Where utter randomness or patternlessness prevails, nothing is predictable” (1991, 30).

We can follow Dennett in asking, “When are the elements of a pattern real and not merely apparent?” What is it, for Dennett, that makes a pattern real? Dennett takes facts about algorithmic compressibility to be inherently interest-relative, as different human interests and goals allow for different levels of tolerance for noise. Depending on our purpose, we might be fine with sending a low-quality image so long as we know that the recipient will be able to make out what is depicted; other times we want to send a sharp image with no blurriness. Our standards for how much inaccuracy and inexactness we are willing to tolerate differ depending on the agent, the situation, and the purpose of the communication. For Dennett, there is no way to adjudicate among competing versions of the same real pattern except by appealing to human interests. Because Dennett builds interest-relativity into his definition of real patterns, his view backslides into instrumentalism. This leaves his account of realness to be swallowed up by the quicksand of relativism. Dennett’s account of real patterns therefore cannot provide a foundation for a metaphysically realist conception of weak emergence. A notion of algorithmic compressibility understood in terms of real patterns will not be able to provide us with a metaphysical account of weak emergence so long as we invoke a relativistic understanding of what it is for a pattern to be real.

It is when the elements of a pattern are real and cannot be predicted from the microstates and dynamics except by simulation that we should say they are weakly emergent. How can we make metaphysical sense of this requirement that a pattern be real and avoid the backslide to interest-relativity? In their account of *information-theoretic structural realism*, Ladyman and Ross (2007) build on Dennett’s picture but reject the interest-relativity of algorithmic compressibility. They avoid the quicksand by adding criteria of both projectibility and informational non-redundancy. This allows them to develop Dennett’s notion of real pattern into the foundation for their information-theoretic realist ontology: to be is to be a real pattern.

For Ladyman and Ross, a pattern is real iff:

(i) it is projectible under at least one physically possible perspective

(ii) it encodes information about at least one structure of events or entities *S* where that encoding is more efficient, in information-theoretic terms, than the bit-map encoding of *S*, and where for at least one of the physically possible perspectives under which the pattern is projectible, there exists an aspect of *S* that cannot be tracked unless the encoding is recovered from the perspective in question. (2007, 226)

The criterion of projectibility allows real patterns to be distinguished from merepatterns. For instance, Ladyman and Ross consider that many curves can be drawn through the past price data on a stock market. They note that all curves that link accurate data points are patterns in that they are records of past values and the mathematical relations that hold among them. But, they note, “what the financial economist or stock analyst wants to know is which of these patterns can be reliably projected forward—which ones generalize to the unobserved cases” (228). It is this predictive power and generalizability that the criterion of projectibility aims to capture.

Condition (ii) provides the criterion of algorithmic compressibility and combines it with the requirement of informational non-redundancy. The requirement that real patterns must be informationally non-redundant prevents the ontology from incorporating numerous versions of the same basic pattern, which vary according to degrees of precision. These are the sort of pattern that Dennett allows for by making his criterion interest-relative. Thus, Dennett’s slip into instrumentalism is ruled out. Condition ii) also precludes reification of patterns that freeload on other real patterns but do not carry their own ability to predict or explain additional phenomena. These are the two types of patterns that Ladyman and Ross identify as those that they specifically aim to exclude with such a criterion: “those that are mere artifacts of *crude* measurement that bring no informational gain—the cases Dennett wrongly lets in—but also putative patterns that are just arbitrary concatenations of other patterns” (231). An instance of the latter would be the classic mereology example of the object composed of Socrates’ nose and the Eiffel Tower. Such an object would not be considered a real pattern, because identification of it supports no generalizations beyond those already supported by reification of the two conjuncts on their own. Thus, if we assume Ladyman and Ross’s criterion of projectibility in our definition of real pattern, we can use the notion of a real pattern to define algorithmic compressibility in a way that is metaphysically robust. This blocks the move to instrumentalism and produces a candidate for a metaphysically realist account.

**4. A Metaphysical Weak Emergence**

Using the interpretation of algorithmic compressibility as non-interest-relative that comes out of Ladyman and Ross’s view, we can replace Bedau’s initial definition of weak emergence with one that limits weakly emergent macrostates of a system to those that are algorithmically compressible:

On this account, we can understand weak emergence as follows:

Where system S is composed of micro-level entities having associated micro-states, and where microdynamic D governs the time evolution of S’s microstates, macrostate P of S with microdynamic D is weakly emergent iff P *is algorithmically compressible* and can be derived from D and S’s external conditions only by simulation.

On this view, weakly emergent macrostates are characterized by their non-derivability except by simulation *plus* their algorithmic compressibility. This prevents completely random macrostates from being categorized as weakly emergent.

Can this improved version of weak emergence understood in information-theoretic terms withstand Wilson’s critique that the resulting picture is not genuinely metaphysical? Wilson (2013) claims that a picture of weak emergence that is compatible with ontological reduction cannot be metaphysical. She questions why *non-derivability without simulation* shouldbe taken to characterize a metaphysical rather than merely epistemological account of emergence. Wilson doubts that such a property can provide a metaphysical foundation for weak emergence, because non-derivability seems compatible with the ontological reducibility of a system’s macrostates to its microstates. Wilson assumes that if a system’s macrostates are ontologically reducible to its microstates then there is no sense in which they can be metaphysically weakly emergent. Like many others (e.g. Epstein 1999), she also assumes that a rejection of ontological reduction entails the existence of strong emergence.

Does a rejection of ontological reduction entail strong emergence, or is it possible to deny that weakly emergent entities and phenomena are ontologically reducible while still steering clear of the spooky metaphysics associated with strong emergence? The denial that a rejection of ontological reduction entails strong emergence is the strategy that Ladyman and Ross (2007) use in developing their account of *rainforest realism*. Their view is an information-theoretic ontology whose name references the authors’ rejection of the puritanical aesthetic at the heart of Quinean desert ontology. Ladyman and Ross take their picture to navigate between the Scylla and Charybdis of reduction and emergence. Since they aim to take the special sciences seriously, they are committed to allowing entities (or, in their terms, real patterns) from special-science domains into their ontology. They deem this necessary to avoid instrumentalism about the special sciences, which they take to be entailed by reductionism. Ladyman and Ross place their view somewhere between reduction and emergence. Their reason for this is that their picture retains two of Kim’s (1999) five central doctrines of emergent phenomena, namely *unpredictability* and *unexplainability/irreducibility*, while jettisoning the rest.[[5]](#footnote-5) Non-fundamental real patterns that are the purview of the special sciences are not predictable from exhaustive information concerning their basal conditions, neither are they explainable in terms of nor reducible to them.

Though Ladyman and Ross take their criteria to constitute an information-theoretic principle of ontology, we can also use their criteria to simply characterize the phenomenon of metaphysical weak emergence. While Ladyman and Ross reject emergence, the spirit of their view is very much in line with the motivations for weak emergence. Like Bedau, Ladyman and Ross wish to take seriously the methods and domains of the special sciences. They do not want to foreclose on the reality of the special sciences’ laws and objects because of a theoretical commitment to reductionism. Yet they also do not take their rejection of reductionism to entail a commitment to strong emergence. If there are indeed more than the two metaphysical options of ontological reduction and strong emergence, then the improved account of weak emergence I have outlined here ought to be understood as both metaphysical and compatible with materialism.

Let us return to Wilson’s challenge in light of what has been suggested above. She writes, (2019, 162):

Bedau also suggests that emergence on his account is properly metaphysical since the non-linear phenomena at issue typically instantiate macro-patterns. But first, incompressibility isn’t either necessary or sufficient for the occurrence of a macro-pattern, so even if the occurrence of a macro-pattern is in some sense a metaphysical fact, the connection to Bedau’s official account of emergence is unclear.

Second, and more importantly, the mere presence of a macro-pattern isn’t enough to establish ontological and causal autonomy. . . Why think that there is anything metaphysically as opposed to epistemologically real about the phenomenon of macro-patterns? Bedau never addresses or answers this question; hence an appeal to macro-patterns doesn’t clearly establish that his account of emergence is properly metaphysical, either by the lights of blocking ontological reduction (as per the criterion of appropriate contrast) or by any other lights.

Consider Wilson’s first charge—that incompressibility is neither necessary nor sufficient for weak emergence and thus, its connection to Bedau’s account remains unclear. Here we are in agreement, but this objection is not devastating. There is indeed some conceptual confusion evident in both Bedau’s original account and in some of the literature surrounding it regarding how exactly his conception of weak emergence is related to the notions of algorithmic compressibility and incompressibility.[[6]](#footnote-6) I suggest that this confusion arises from the following fact: That, *once we establish* that phenomena must reach a certain minimal threshold of algorithmic compressibility in order to qualify as weakly emergent, *then* there is a positive correlation between algorithmic incompressibility and weak emergence, meaning that the more incompressible the phenomenon is the more emergent it is. But that correlation only holds for those phenomena that possess *some minimal level* of algorithmic compressibility and are *not* maximally incompressible—as maximal incompressibility is equated with complete randomness. If we try to employ a universal principle that directly equates algorithmic incompressibility with weak emergence (because the former is positively correlated with non-derivability except by simulation), then we wind up with an untenable view that counts completely random noise as weakly emergent phenomena. This is, as it stands, what Bedau’s view does. However, if we invoke a criterion of minimal algorithmic compressibility (as I suggest we do), in order to ensure that we only count *real phenomena* as weakly emergent, then the problem is avoided. This is why it is useful to bring in the notion of real patterns. We do not want to say that a pattern is weakly emergent *unless that pattern is real*. Weak emergence is a property of real phenomena, not randomly disordered unstructured noise whose behavior cannot be derived from initial conditions *simply because* it is completely random.

This offers another way to make sense of Wilson’s objection. Barring an invocation of a minimal criterion of algorithmic compressibility, there really is nothing metaphysical about the definition of emergence offered in Bedau’s account. It counts plenty of random patterns as weakly emergent when they aren’t even real. With the additional criterion suggested here, however, these random (non-)patterns are ruled out from the class of weakly emergent phenomena and what remains are a class of patterns interestingly linked to the underlying micro-constituents that produce them by the fact they are non-derivable from them except by simulation or iteration.

To answer Wilson’s second charge, let us consider why we might think there is something metaphysically real about the phenomenon of macro-patterns. One background assumption that goes some way toward taking the account of weak emergence outlined here to be plausibly metaphysical is that there is an objective compressibility measure. This is the assumption that there is a single most accurate way of defining algorithmic compressibility that captures the true sense in which an object or pattern is compressible or incompressible and does not just come down to a choice of what language we use to program our Turing machine. If there is a universal objective measure of compressibility then it is plausible that the non-derivability of weakly emergent macrostates *is* a reflection of some moderate level of ontological independence if we understand their non-derivability to stem in part from their information-theoretic properties.

Information theory’s practical origin in communication theory means that it is primarily concerned with the goal of finding a description that is good *on average*. The descriptive complexity of an event is often taken to be given by the number of bits needed to describe it by a Shannon code, where a description is satisfactory if it satisfices some threshold of precision even while allowing for some error. This might lead one to think that compressibility is a broadly instrumentalist property and one that is relative to human interests. Indeed, Dennett (1991) makes such an assumption when formulating his definition of real patterns. However, the foundation that Kolmogorov complexity provides for information theory allows us to formulate the notion of the universally shortest description (Cover & Thomas 2006). Kolmogorov went further than Shannon’s broadly pragmatic account and defined the algorithmic or descriptive complexity of an object as the length of the shortest binary computer program that describes the object. Kolmogorov observed that, according to this definition, the complexity of an object is independent of the choice of computer. “It is an amazing fact that the expected length of the shortest binary computer description of a random variable is approximately equal to its entropy. Thus, the shortest computer description acts as a universal code, which is uniformly good for all probability distributions” (Cover & Thomas 2006, 463).

How can this understanding of descriptive complexity shed light on the possibility of non-derivability providing a metaphysical foundation for weak emergence? It is a consequence of the fact that there is no algorithm for the halting problem that Kolmogorov complexity is non-computable. The only way to find the shortest program to describe an event is trial and error:

One of the consequences of the nonexistence of an algorithm for the halting problem is the noncomputability of Kolmogorov complexity. The only way to find the shortest program in general is to try all short programs and see which of them can do the job. However, at any time some of the short programs may not have halted and there is no effective (finite mechanical) way to tell whether or not they will halt and what they will print out. Hence, there is no effective way to find the shortest program to print a given string. (Cover & Thomas 2006, 483)

That trial and error is the only way to tell whether a short program will halt reflects the noncomputability of Kolmogorov complexity and follows from the lack of an algorithm for the halting problem. There is no algorithm that tells us how to determine a system’s Kolmogorov complexity, though there is a single, objective answer. Similarly, weakly emergent systems’ macrostates cannot be derived except by simulation, even though there is often a determinate fact of the matter about what a given (closed) weakly emergent system’s macrostates will be.

The nature of computational processes gives us important insights into epistemology. But their import is not solely epistemological. Humphreys (2009) emphasizes that humans are not able to specify and interpret every step of the computational processes that produce the outputs of computer simulations, which are themselves knowledge-producing. He refers to this as the “essential epistemic opacity of the computational process leading from the abstract model underlying the simulation to the output” (2009, 618). Because of this feature of computation processes, he argues that the increasingly central role that computer simulations play in science ought to move us away from an anthropocentric conception of epistemology. What we can know now transcends human cognitive abilities.

How then should we understand whether there can be contexts in which it is appropriate to interpret properties such as non-derivability as metaphysical? On the relationship between in/compressibility and non-derivability of macro-states from micro, Wilson asserts that “these are epistemic notions, with no clear bearing on the nature of metaphysical emergence.” But this is not quite as straightforward as Wilson suggests. Sometimes, our inability to derive macropatterns from micro ones is simply a reflection of our limited cognitive capacities or our under-developed technologies. This type of non-derivability can be accurately identified as merely epistemic and should not be taken as evidence of anything interestingly metaphysical at play. In other instances, however, our inability to derive some facts about a system from others may be reflective of interesting metaphysical features of the system in question. Consider, for instance, how we might interpret the fact that we cannot derive all truths about the arithmetic of natural numbers from any consistent set of axioms that algorithmically produce it as an output. Godel’s incompleteness theorem, in other words, is usually not—and should not—be understood as merely a reflection of the finiteness of human minds. It is derivable in principle (it is a theorem after all) and thus likely reflects something metaphysically interesting and important about the nature of certain formal systems.

Or consider another example that is commonly deployed in philosophy—the naturalistic fallacy. The fact that we ‘cannot derive an *ought* from an *is*’ is rarely taken to be reflective of a mere human cognitive limitation; it is far more often employed as evidence of the (metaphysical) independence of the realms of the normative and the empirical. My purpose here is not to comment on the accuracy of this inference, but to direct attention to the methodological availability of making a metaphysical inference from what could be construed as an epistemic limitation. After all, the metaphysical independence of the normative and the descriptive would entail that we cannot derive statements of the former from statements of the latter. Of course, that is not the only possible explanation of the naturalistic fallacy or of what we might term the ‘non-derivability’ of the normative from the descriptive. Some epistemic constraints are explained by metaphysical facts; others are not.

Our job in this case is to tease apart whether the epistemic constraints on the derivability of the macro from the micro in weakly emergent systems are due to something interestingly metaphysical, an underlying feature of the world that explains and entails the constraints. Since the epistemic constraints themselves are derivable from information-theoretic facts and the a priori notions of computational complexity that underly information theory, I suggest that they are indeed due to a metaphysical feature of the world. When we prove a priori the existence of epistemic limitations that hold for all possible observers, that strongly suggests the presence of a metaphysical constraint that is responsible for producing the epistemic one. We do not look at Cantor’s diagonalization theorem and say, “if we had more powerful cognitive capacities, we would be able to put the natural numbers and the real numbers in a one-to-one correspondence.” Rather, we take the theorem to show that it *is not possible* to put the natural numbers and the real numbers in a one-to-one correspondence—a metaphysical conclusion. Since Cantor’s diagonalization is a priori and holds for all possible observers, we consider it metaphysical. I suggest we do the same with respect to an information-theoretic understanding of the relationship between weakly emergent phenomena and the micro-patterns that underly them. Baker (2010, 2) gestures at a similar point when he explains,

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. . . 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.

Baker emphasizes the parallel with the standard way of mathematically defining irrational numbers as those numbers that cannot be expressed as the ratio of two integers. While that definition may use language that references “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” (Baker 2010, 2). In other words, though we may define irrational numbers as those that cannot be expressed or constructed as the fraction of two integers, we do not take this constraint to be about the inadequacy of our modes of expression or our construction methods. We take it to reflect the nature of irrational numbers themselves, whatever they might be. The modal constraint expressed in the definition is a metaphysical one rather than an epistemic one.

The fact that computations now push up against the limits of what can be known tells us something about the nature of these limits themselves. While computational limits produce many constraints on what we can know and how we can know, this does not mean that facts about the limits of computation are merely epistemic. Insofar as the limits of computability and computation are features of the world that are independent of facts about the way we cognize the world, they deserve to be considered metaphysical. Indeed, that facts about computability entail limits of knowledge that apply to all possible knowers reflects their metaphysical nature. These metaphysical facts about computability provide an information-theoretic explanation for why weakly emergent macrostates can only be derived from microstates by simulation.

**5. Conclusion**

A definition of weak emergence that includes a minimal information-theoretic requirement of algorithmic compressibility is a necessary amendment to Bedau’s (1997) account, as Bedau’s initial failure to rule out incompressible macrostates from his definition allows random patternless macrostates to count as weakly emergent. The information-theoretic property of algorithmic compressibility and Dennett’s concept of real patterns can be used to clarify the intuitive notion of weak emergence that Bedau’s picture aims to capture. I have argued that an account of weak emergence that combines *non-derivability* with a minimal threshold of *algorithmic compressibility* is plausibly understood as metaphysical. Weak emergence ought to be understood as a thesis about the information-theoretic relations between microstates and macrostates. These relations can all be characterized in terms of compressibility and computability and should be understood in objective realist terms. A picture of weak emergence understood in terms of metaphysical properties of compressibility, computability, and derivability can itself be understood as metaphysical.

**Acknowledgments**

Thanks to the following people for their valuable feedback on earlier versions of this paper: Shamik Dasgupta, Nina Emery, Joshua Watson, Jessica Wilson, and two particularly helpful anonymous reviewers.

**Works Cited**

Baker, A. (2010). Simulation-based definitions of emergence. *Journal of Artificial*

*Societies and Social Simulation*, 13(1).

Bedau, M. (1997). Weak emergence. *Philosophical Perspectives* 11: 375-99.

Bedau, M. (2003). Downward causation and the autonomy of weak emergence. Principia 6

(1):5-50.

Bedau, M. (2011). Weak emergence and computer simulation. In P. Humphreys &

C. Imbert (eds.), [*Models, Simulations, and Representations*](http://philpapers.org/rec/HUMMSA-2). Routledge. pp 91-114.

Bedau, M. (2013). Weak emergence drives the science, epistemology, and

metaphysics of synthetic biology. *Biological Theory* 8 (4):334-345.

Chaitin, G.. (1996). How to run algorithmic information theory on a computer:

Studying the limits of mathematical reasoning. *Complexity* 2 (1):15-21.

Cover and Thomas. (2006). *Elements of Information Theory*. Wiley-Interscience.

Dennett, D. (1991). Real patterns. *Journal of Philosophy* 88 (1):27-51.

Epstein, J. (1999). Agent-based computational models and generative social science.

*Complexity* 4 (5): 41–60.

Grünwald, P. D. & Vitányi, P. M. B. (2003). Kolmogorov complexity and

information theory. With an interpretation in terms of questions and answers. *Journal of Logic, Language and Information* 12 (4):497-529.

Hovda, P. (2008). Quantifying weak emergence. *Minds and Machines* 18 (4):461-473.

Hsieh, D. A. (1991). Chaos and Nonlinear Dynamics: Application to Financial Markets.

*The Journal of Finance* 46 (5): 1839-77.

Humphreys, P. (2008). Computational and conceptual emergence. *Philosophy of*

*Science* 75 (5): 584-594.

Humphreys, P. (2009). The philosophical novelty of computer simulation

methods. *Synthese* 169 (3):615 - 626.

Imbert, C. (2005). Why diachronically emergent properties must also be salient. In

C. Gershenson, D. Aerts & B. Edmonds (eds.), Worldviews, Science, and Us: Philosophy and Complexity. World Scientific. pp. 99--116.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies* 95.1-2:3-36.

Ladyman, J., Lambert, J. & Wiesner, K. (2013). What is a complex

system? *European Journal for Philosophy of Science* 3 (1):33-67.

Ladyman, J. & Ross, D. with Spurrett, D. & Collier, J. (2007) *Every Thing Must Go.*

*Metaphysics Naturalised.* Oxford: Oxford University Press.

Lovejoy, S. (1982). Area-Perimeter Relation for Rain and Cloud Areas. *Science,*

*216* (4542), 185-187 DOI: [10.1126/science.216.4542.185](http://dx.doi.org/10.1126/science.216.4542.185)

Oestreicher, C. (2007). A history of chaos theory. *Dialogues in Clinical Neuroscience*

9 (3): 279-89.

Saremi, S., Mirajalili, S. M., and Mirajalili, S. (2014). Chaotic

krill herd optimization algorithm. *Procedia Technology* 12: 180-85.

Symons, J. (2008). Computational Models of Emergent Properties. *Minds and*

*Machines* 18 (4):475-491.

Theurer, K. L. (2014). Complexity-based Theories of Emergence: Criticisms and

Constraints. *International Studies in the Philosophy of Science* 28 (3):277-301.

Wilson, J. (2019). Between Scientism and Abstractionism in the Metaphysics of

Emergence. Routledge Handbook of Emergence (ed. Sophie Gibb, Robin Hendry, Thomas Lancaster, eds.): 157-176.

Wilson, J. (2013). Nonlinearity and Metaphysical Emergence. In Stephen

Mumford & Matthew Tugby (eds.), [*Metaphysics and Science*](http://philpapers.org/rec/MUMMAS).

1. Of course, there are limits to how far an analogy between the weak emergence demonstrated in a cellular automaton and the relationship between microstates and macrostates of a physical system in the real world can go. As Symons emphasizes, “CA simulations provide qualitative analogies to the natural systems in question rather than measurable quantities. These analogies can be extremely useful and may give us some insight into methods for controlling the phenomenon in question. However, in practice and in application, generalizations that we derive from these models will rest entirely on the analogy between the system under consideration and the simulation. This is one of the reasons that the success of a computational models is generally judged not by its predictive power, but by the degree to which it imitates the known behavior of a target system” (2008, 482). Thus, it is not a foregone conclusion whether the sort of weak emergence that Bedau outlines will have cognates in, for instance, the realm of mental phenomena. That will depend on whether the relevant relationship between the candidate phenomenon and its system’s microstates are present. [↑](#footnote-ref-1)
2. Baker (2010, 3) offers another way to understand the notion of *non-derivability except by simulation* present in Bedau’s work. He suggests that the notion that Bedau is working with involves equating simulation with “’iterating the microdynamic’ of the given system. Thus, a weakly emergent property or behavior in the Game of Life would be one that can be derived only by iterating the update rules time step by time step. [↑](#footnote-ref-2)
3. When ‘compressibility’ appears unmodified, take it to be shorthand for ‘algorithmic compressibility.’ [↑](#footnote-ref-3)
4. One could, of course, draw an arbitrary compressibility threshold higher than the bare minimum that must be met for a system’s states to be counted as weakly emergent. Another option is to simply take weak emergence to be a matter of degree, as Hovda (2008) does. [↑](#footnote-ref-4)
5. They reserve commitment to other doctrines, such as causal efficacy, in part because they do not take them to be necessary to make sense of science nor demanded or even warranted by contemporary scientific theory or practice. [↑](#footnote-ref-5)
6. For instance, Wilson (2013, 214) initially describes Bedau’s picture as follows: “The absence of analytic or otherwise ‘compressible’ means of predicting the evolution of such systems means that the only way to find out what this behavior will be is by ‘going through the motions’: Set up the system, let it roll, and see what happens. It is this feature—namely, *algorithmic incompressibility*—that serves as the basis for Bedau’s account of weak emergence.” Wilson takes the account to be one that identifies weak emergence with algorithmic incompressibility. While the notion of algorithmic incompressibility does not explicitly appear in Bedau’s (1997) original account, the fact that Bedau (2013) generally uses ‘incompressible’ interchangeably with the notion of non-derivability except by simulation makes Wilson’s interpretation of Bedau (1997) a plausible one. As I argue, however, such a direct identification of weak emergence with algorithmic incompressibility is too coarse-grained to capture the relationship between these two notions that is needed for a metaphysical account of weak emergence. [↑](#footnote-ref-6)