Representations of Complexity: How Nature Appears in Our Theories

J. J McDowell
Emory University

In science we study processes in the material world. The way these processes operate can be discovered by conducting experiments that activate them, and findings from such experiments can lead to functional complexity theories of how the material processes work. The results of a good functional theory will agree with experimental measurements, but the theory may not incorporate in its algorithmic workings a representation of the material processes themselves. Nevertheless, the algorithmic operation of a good functional theory may be said to make contact with material reality by incorporating the emergent computations the material processes carry out. These points are illustrated in the experimental analysis of behavior by considering an evolutionary theory of behavior dynamics, the algorithmic operation of which does not correspond to material features of the physical world, but the functional output of which agrees quantitatively and qualitatively with findings from a large body of research with live organisms.

Key words: behavior dynamics, cellular automata, complexity theory, emergence, evolutionary theory, neural networks

---

This article is about a particular class of theory known as *complexity theory* (McDowell, 2013a; McDowell & Popa, 2009), and a particular instance of that class known as the *evolutionary theory of behavior dynamics* (McDowell, 2004). Complexity theories are characterized by algorithmic mechanisms that generate outputs that can be compared to empirical measurements and phenomena (McDowell & Popa, 2009). A *condicio sine qua non* of a good complexity theory is that its output agrees with empirical measurements. But one can also ask whether the algorithmic mechanism of the theory, which is usually a set of fairly simple

rules, also corresponds to the natural world. In other words, must nature appear *in* the theory, as well as appearing as a result of its algorithmic operation? Consider, for example, a cellular automaton, which is an instance of complexity theory in which a set of simple rules is applied to rows of squares, or cells. Each cell can exist in one of two states, such as black or white, and the rules of the automaton specify how the states of the cells change from row to row. Different sets of rules constitute different cellular automata. The result of a cellular automaton's operation is an image, a pattern of black and white cells in the consecutive rows of cells the automaton generates. McDowell and Popa (2009) used specific mappings from cellular automata to study the applicability of these finite state machines to behavior. They found that at least one cellular automaton generated outcomes that corresponded to the behavior of live organisms in limited ways (cf. McDowell, 2013a). But the algorithmic mechanism of this account of behavior is completely abstract; it consists of cells that change state according to rules. Nature, one might say, is not in the

---

I thank Jack Marr for pointing out James Clerk Maxwell's reliance on physical analogy in the development of his theory of electromagnetism, and for the cartoon metaphor for artificial neural networks. I also thank him for his question at the conference about what it means to say that the brain carries out computations, a question that I attempt to answer in this article. Nick Calvin and Andrei Popa made many helpful comments on an earlier version of this paper, for which I am grateful.

Address correspondence to J. J McDowell, Department of Psychology, Emory University, Atlanta, Georgia 30322 (e-mail: psyjjmd@emory.edu).

theory. Must nature be in a theory for it to count as a good theory?

As will be seen in the discussion that follows, consideration of this question raises at least two related questions. The first is: What does it mean to have two or more complexity theories with different algorithmic mechanisms that account for the same phenomena? Accounting for the same phenomena means that the outputs generated by the theories' distinct operations correspond to the same features of the natural world with comparable fidelity. The second related question is: If nature does not appear in a complexity theory, then in what sense, if any, can its algorithmic operation be said to correspond to the natural world? It will be best to consider these issues in the context of a concrete example. A good exemplar of complexity theory in the experimental analysis of behavior is the evolutionary theory of behavior dynamics (McDowell, 2004). This theory will be explained briefly in the following section, and then the evidence supporting it will be reviewed in some detail to establish that it is a good complexity theory, which means that the results of its algorithmic operation agree with a large and varied set of behavioral measurements.

THE EVOLUTIONARY THEORY OF BEHAVIOR DYNAMICS

In the evolutionary theory, a population of potential behaviors evolves under the selection pressure of consequences from the environment. One behavior is emitted at random from this population at each tick of time. The result is a stream of emitted behaviors that can be recorded and analyzed just as if it were the behavior of a live organism. Behaviors in the population of potential behaviors are represented by integers that are expressed in both decimal (phenotypic) and binary (genotypic) forms. The binary forms are strings

of zeros and ones, or bit strings, and one can conceive of breaking them into segments that can be recombined with segments of other bit strings. After a behavior is emitted, pairs of parent bit strings are recombined in this way to produce child bit strings that populate the next generation of potential behaviors. If the emitted behavior results in benefit to the organism, then parent bit strings with phenotypes (i.e., decimal integer values) closer to the emitted behavior, which are fitter parents, are more likely to produce offspring than those with behaviors with phenotypes farther from the emitted behavior. If the emitted behavior does not result in benefit to the organism, then random pairs of parent bit strings are recombined to produce child behaviors for the next generation. After a new generation of behaviors is produced, it is subjected to a small amount of mutation by flipping random bits in the bit strings of the new population. This process can then be repeated indefinitely. A random behavior is emitted from the population, pairs of parent behaviors recombine to produce a new population of child behaviors (with the parents chosen on the basis of their fitness if the emitted behavior resulted in benefit to the organism or at random if it did not), and then this new population undergoes a small amount of mutation.

The evolutionary theory implements Darwinian principles of variation and selection in the form of simple rules applied to a population of potential behaviors. Mutation and random recombination of parent genotypes cause behavioral phenotypes to vary across their permissible range. In contrast, fitness-based recombination of genotypes causes the phenotypes to become relatively concentrated near the phenotype that produced benefit to the organism. In a constant environment, like that arranged by a single or concurrent schedule of random-interval (RI) reinforcement, these opposing forces

of variation and selection reach a dynamic equilibrium, at which point a roughly constant rate of behavior is generated.

The processes of selection, reproduction, and mutation just described constitute the algorithmic operation of the evolutionary theory in its entirety. The theory is simple enough to explain with a few diagrams drawn on the back of an envelope. Yet a substantial body of research has shown that it generates behavior in a variety of experimental environments that is quantitatively and qualitatively indistinguishable from the behavior of live organisms. Perhaps the most significant finding is that, over a range of moderate mutation rates, the theory generates steady-state behavior that conforms to every empirically valid equation of matching theory (McDowell, 2013b) exactly, and without systematic error (McDowell, 2004; McDowell & Caron, 2007; McDowell, Caron, Kulubekova, & Berg, 2008; McDowell & Popa, 2010). In other words, the steady states described by the matching theory equations emerge from (Holland, 1998) and are maintained by the evolutionary dynamics of the theory. In addition, the theory generates specific values for the exponents in these equations that vary around 0.8, which is the value that is typically found in experiments with live organisms (McDowell, 2013b; McDowell et al., 2008). The theory also produces biased responding in asymmetrical choice situations that is indistinguishable from that observed in similar experiments with live organisms (McDowell et al., 2008). When both rate and magnitude of reinforcement are varied, the evolutionary theory generates behavior over a range of moderate mutation rates that conforms exactly, and without systematic error, to a bivariate form of the power function matching equation (Baum & Rachlin, 1969) which Cording, McLean, and Grace (2011) concluded accurately

described the behavior of live organisms (McDowell, Popa, & Calvin, 2012). In McDowell et al.'s (2012) experiments, fits of the bivariate matching equation to behavior generated by the theory yielded exponents on the reinforcement-rate ratio and the reinforcer-magnitude ratio that varied around 0.8 and 0.6 respectively, which are values that are essentially identical to those reported in Cording et al.'s review of experiments with live organisms.

The behavior of virtual organisms animated by the evolutionary theory has also been studied in more complicated two-alternative environments, in which sets of different unsignaled reinforcement-rate ratios are arranged sequentially and in an unpredictable order within sessions (e.g., Davison & Baum, 2000). In these environments, the virtual organisms' behavior reaches a dynamic equilibrium characterized by specific quantitative properties that develop during the period that each reinforcement-rate ratio is in effect. These include (a) preference (as measured by the logarithm of the response-rate ratio) that increases from zero and reaches an asymptotic value rapidly, (b) an exponent from fits of the power function matching equation that increases to an asymptotic value rapidly and also (c) increases more rapidly and reaches a higher asymptote the greater the overall rate of reinforcement, and (d) preference that shifts between components on a reinforcer-by-reinforcer basis according to a variety of well-defined patterns that depend on whether previous sequences of reinforcers came from the same or from the other component (Kulubekova & McDowell, 2013). Remarkably, these same four features have been observed in the behavior of live organisms responding in these environments, including the detailed patterns of preference shift that occur after each reinforcer delivery (Davison & Baum, 2000).

In addition to its quantitative agreement with the behavior of live organisms, the evolutionary theory also generates behavior that agrees qualitatively with many features of live organism behavior. For example, virtual organisms animated by the theory produce cumulative records on single-alternative RI schedules that show roughly constant response rates with brief periods of response acceleration and deceleration, pausing, and other irregularities, rendering them essentially indistinguishable from cumulative records produced by live organisms. Interresponse-time distributions generated by the theory on single-alternative schedules are also similar to those produced by live organisms (Kulubekova & McDowell, 2008). On concurrent schedules, the theory generates changeover patterns (McDowell et al., 2008, 2012) and responsiveness to changeover delays (Popa & McDowell, 2010) that are indistinguishable from those observed in experiments with live organisms.

The evidence supporting the evolutionary theory shows that it is a good complexity theory in the sense that its output agrees quantitatively and qualitatively with many measurements and observations in the natural world. But does nature appear in the theory? That is, does the theory's algorithmic operation correspond to material reality?

NATURE IN THE THEORY

Clearly, the answer to this question is no. Populations of potential behaviors do not exist in the natural world. Behaviors are not bit strings, they do not recombine to produce offspring, and they do not undergo mutation. Hence, nature cannot reasonably be said to appear in the evolutionary theory. But how are we to understand a successful complexity theory that does not incorporate nature in its algorithmic operation?

To answer this question, it will be helpful to consider some of the

controversies that arose in the development of quantum mechanics in the early 1900s. By almost any measure, quantum mechanics is the most successful scientific theory ever developed, and yet nature does not appear in this theory on any reasonable understanding of space and time in the physical world. Quantum theory does not provide what physicists call *Anschaulichkeit*, which is an intuitive picture or visualization of physical reality. The theory has rules, and following the rules produces outcomes that agree with physical measurements to many places of decimals, but it does not provide *Anschaulichkeit*. Werner Heisenberg, for one, accepted this feature of quantum theory with equanimity. The abstract of his famous 1925 paper, in which he introduced his successful matrix mechanics, consisted of the following single sentence: "The present paper seeks to establish a basis for theoretical quantum mechanics founded exclusively upon relationships between quantities which in principle are observable" (Heisenberg, 1925/2007, p. 261). Contemporary theoretical physicists Brian Cox and Jeff Forshaw (2012) note that

This is an important step, because Heisenberg is saying that the underlying mathematics of quantum theory need not correspond to anything with which we are familiar. The job of quantum theory should be to predict directly observable things, such as the colour of the light emitted from hydrogen atoms. It should not be expected to provide some kind of satisfying mental picture for the internal workings of the atom, because this is not necessary and it may not even be possible. (pp. 12ff)

However, other physicists, including Erwin Schrödinger and Albert Einstein, disagreed. Schrödinger believed that his wave mechanics, which was an alternative to Heisenberg's matrix mechanics, was related to the distribution of electric charge inside the atom (Cox & Forshaw). In his 1926 paper, in which he compared his

theory to Heisenberg's, Schrödinger commented in a footnote,

My theory was inspired by L. de Broglie and by brief but infinitely far-seeing remarks of A. Einstein. I was absolutely unaware of any genetic relationship with Heisenberg. I naturally knew about his theory, but because of the to me very difficult-appearing methods of transcendental algebra and the lack of *Anschaulichkeit*, I felt deterred by it, if not to say repelled. (quoted and translated in Moore, 1992, p. 211)

Indeed, Schrödinger rejected in no uncertain terms what was to become the dominant view in quantum mechanics, which was a view without *Anschaulichkeit*: "Bohr's standpoint, that a space-time description is impossible, I reject *a limine*. ... If [atomic research] cannot be fitted into space and time, then it fails in its whole aim and one does not know what purpose it really serves" (quoted in Moore, 1992, p. 226). Heisenberg, in turn, commented on Schrödinger's wave theory in a letter to Wolfgang Pauli, and he did not mince words: "The more I think about the physical portion of the Schrödinger theory, the more repulsive I find it. ... What Schrödinger writes about the visualizability of his theory 'is probably not quite right,' in other words it's crap" (quoted and translated in Cassidy, 1993, p. 215).

It turns out that Heisenberg was right. Max Born later showed that Schrödinger's wave equation could not be related to the distribution of electric charge in the atom, but instead had to be interpreted as a probabilistic representation of the location of a quantum of electric charge (Cox & Forshaw, 2012).

With the demise of Schrödinger's *Anschaulichkeit*, quantum physics was left with two different theories that dealt with the same phenomena, and no *Anschaulichkeit*. The Heisenberg and Schrödinger mechanics were (and are) functional theories in the sense that they operate like mathematical functions, producing answers that can be compared to

measurements in the natural world. In 1926 Schrödinger proved that the two theories gave the same answers, that is, predicted the same physical measurements (Moore, 1992).

Further efforts were made by Einstein and others to establish *Anschaulichkeit* in quantum theory, but none succeeded (van Dongen, 2010). Joining in the search were philosophers of science, including Hilary Putnam, who subscribed to a kind of mathematical realism, what he called an ontological commitment to mathematical entities (Hill, 1992; Putnam, 1985). This mathematical realism gave rise to a form of scientific realism. Putnam believed, or at least did at one time (Baghramian, 2008), that a successful functional theory, like quantum mechanics, *must* be a realist theory in the sense that its operation *must* correspond to the operation of the natural world; otherwise the theory would not be so successful. He famously said that if this were not the case, then the success of such a theory would be a miracle (Putnam, 1985). Hence, rather than requiring that the connection between a theory's operation and the operation of the natural world be demonstrated, Putnam was happy to assert it as a *fait accompli*, given the theory's functional success. Needless to say, Putnam's view had many critics (e.g., Rorty, 1979). One problem with it is that it seems incapable of dealing with the existence of two or more successful functional theories that entail different mathematical entities.

In the 1940s, Richard Feynman (1985/2006) developed yet another theory of quantum phenomena. It was based on ideas of Paul Dirac and has been referred to as the Dirac-Feynman path integral formalism. Amazingly, Feynman's theory gives the same answers as Heisenberg's and Schrödinger's theories, but via a different mathematical route. Also like Heisenberg's and Schrödinger's theories, Feynman's theory lacks

Anschaulichkeit. Some 20 years after Schrödinger found himself repelled by Heisenberg's stark, unvisualizable theory, Feynman commented that physicists have

learned to realize that whether they like a theory or they don't like a theory is *not* the essential question. Rather, it is whether or not the theory gives predictions that agree with experiment. It is not a question of whether a theory is philosophically delightful ... or perfectly reasonable from the point of view of common sense. The [path integral formalism is] absurd from the point of view of common sense. And it agrees fully with experiment. (p. 10)

Indeed, physicists have settled on the Bohr-Heisenberg-Feynman perspective, having come to terms with the Schrödinger-Einstein discontent regarding the absence of *Anschaulichkeit* (Cox & Forshaw, 2012). And interestingly, there are three good theories of quantum phenomena, all without *Anschaulichkeit*. Which theory a physicist uses in a particular instance is a matter of preference and suitability to the problem at hand (Feynman, 1985/2006). The existence of multiple theories that accurately describe the same phenomena gives rise to interesting and important philosophical questions in both theoretical physics and theoretical behavior analysis, which will be considered below.

ARTIFICIAL NEURAL NETWORKS

As proved to be the case for theoretical physics, in behavior analysis it may not be necessary, or even possible, to develop a complexity theory about behavior that incorporates nature in its algorithmic operation. However, general acceptance of this point of view in behavior analysis may not be won easily, as it was not won easily in physics. Consider, for example, Feynman's somewhat confrontational tone in the passage just quoted, in which he insisted that his own theory was absurd and yet

agreed fully with experiment. Evidently he was trying to persuade his readers (and perhaps also himself) that *Anschaulichkeit* really wasn't necessary, indeed that it could simply be tossed aside. He was even more insistent in one of his Robb Memorial Lectures at the University of Auckland in 1979. In that lecture, the ever-colorful Feynman admonished his audience, basically, to take it or leave it:

You don't like it [quantum theory without *Anschaulichkeit*]? Go somewhere else! To another universe, where the rules are simpler, philosophically more pleasing. ... I can't help it, okay? ... I'm not going to tell you it's something like a ball bearing on a spring. It isn't. So I'm going to tell you what it really is like, and if you don't like it, that's too bad, okay? (Dick P, 2009)

One objection to adopting a similar point of view in behavior analysis is based on the observation that a biological organism is a very different sort of entity than an atom. Hence, *Anschaulichkeit* may really be possible in a complexity theory of behavior. An obvious place to look for it is in the brain or nervous system. And to pursue this possibility, it might be helpful to state complexity theories about behavior in the form of artificial neural networks, the operation of which could be analogous to brain functioning. The network would generate results that could be compared to behavioral measurements, and its algorithmic operation could be compared to the operation of the brain. Hence, nature would appear in the theory, at least in principle.

Artificial neural network theories of behavior have been proposed and studied by a number of researchers, including Donahoe and colleagues (e.g., Donahoe, Burgos, & Palmer, 1993; Donahoe & Palmer, 1994; Donahoe, Palmer, & Burgos, 1997) and Edelman and colleagues (e.g., Edelman, 1987; Tononi, Sporns, & Edelman, 1992). These theories are usually quite complicated, consisting

of many artificial neural units with various activation properties, and with many parameters and rules by which the synapses between the units operate, increase or decrease in strength, and so on. In addition, the structures of the networks can vary in many ways, including type of connectivity, number of layers, number and nature of input and output units, and the inclusion or not of subnetworks. A glance at the appendix of Donahoe et al.'s (1993) paper, for example, reveals the complicated array of vectors, parameters, weights, thresholds, probability distributions, activation functions, learning functions, and so on in their theory.

Evidently, neural network theories do not have the virtue of simplicity. But one could argue that they must be complicated because the brain is complicated. The latter of course is true, but the brain is in fact far more complicated than any artificial neural network theory (Marr, 1997, 2000), making the theories very rough approximations of brain function at best. In addition, various features of the theories do not correspond to properties and operating characteristics of real nervous tissue in any straightforward way (Kehoe, 1989). For example, what do vectors of activations, vectors of excitatory weights, and activation-decay parameters (Donahoe et al., 1993) correspond to in the brain? Neural network theories give the appearance of representing brain structure and function, but careful consideration reveals them to be more like cartoons of the brain. The networks look more or less like the real thing because they incorporate some obvious features of nervous systems, such as a network structure. But this resemblance is superficial. The networks often omit important details of nervous system structure and functioning, and those that are included may work differently in the artificial network than in the real one. In addition, as noted above, structural and operational

details may appear in artificial neural networks (e.g., full connectivity, as in many of Donahoe and Palmer's, 1994, networks) that do not correspond in any reasonable way to the structure or operation of a real brain.

It may be that the operation of the nervous system can be described only by statements of facts about neuro-anatomical structures along with neurophysiological and neurochemical events and processes, such as one finds in a textbook of neuroscience. If so, then behavioral theories stated in the form of neural networks are in fact functional theories (Kehoe, 1989). The results of their operation can be compared to measurements, but they cannot be said to incorporate a reasonable representation of nature in their algorithmic operation. One might object to this characterization by saying that no artificial neural network theory can be expected to incorporate every possible detail of nervous system functioning. For example, there will never be molecules in such a theory, or neurotransmitter receptors on cell bodies, or vesicles that release neurotransmitter into synaptic clefts, and so on. But that is precisely the point. These missing details are statements of fact that are needed to describe the operation of a real brain.

This does not mean that theories stated in the form of neural networks are bad theories. On the contrary, they can be good functional theories if they produce results that correspond to experimental measurements. But their algorithmic operation cannot be taken to represent the actual operation of the material world on any reasonable understanding of how the brain really works. Interestingly, a theorist who develops artificial neural networks may be like Schrödinger, whose work was guided by the belief that he was describing real waves of electric charge. No doubt the work of many neural network theorists (see Donahoe et al., 1993) is guided by the belief that they are describing the real

operation of the nervous system. In both cases, it is possible for the resulting theories to be good and comprehensive (e.g., Schrödinger's definitely was) even though the guiding beliefs may ultimately prove untenable. A similar story can be told of James Clerk Maxwell's important theoretical unification of the phenomena of electricity, magnetism, and light some half century before the quantum era. Maxwell's work was guided by a realist analogy, which included the existence of a luminiferous aether (Harman, 2001). But the aether lost its existential credibility thanks to the famous experiments of Michelson and Morley, and so robbed Maxwell's theory of *Anschaulichkeit*. His equations nevertheless remain as an important functional theory of electromagnetic phenomena.

COMPUTATIONAL EQUIVALENCE

Physicists have learned to live with the absence of *Anschaulichkeit* in quantum theory. Some even seem to celebrate it. Niels Bohr is quoted as saying to Wolfgang Pauli after hearing Pauli lecture on a new field theory, "We are all agreed that your theory is crazy. The question that divides us is whether it is crazy enough to have a chance of being correct" (quoted in Wolfle, 1959, p. 66). In theoretical behavior analysis, the idea that the operation of neural network theories might provide a connection with physical reality is probably a false hope. Does this mean that behavior analysts also must be content with complexity theories whose outputs correspond to the natural world, but whose internal workings do not? This concession actually may not be necessary, because there is a sense in which the algorithmic operation of a good complexity theory in behavior analysis, or in physics for that matter, does in fact correspond to the natural

world, no matter how crazy the algorithmic workings may be. To understand how this might be so, it will be helpful to take a brief detour to consider what it means to have multiple theories that account for the same phenomena.

Suppose we had two functional theories in behavior analysis, such as the evolutionary theory of behavior dynamics and an artificial neural network theory, that both accounted for the same behavioral phenomena. We do not have this in behavior analysis, but let us suppose for the sake of argument that we do. This would be analogous to having three quantum theories that account for the same quantum phenomena. How are we to understand such a state of affairs? Well, one could simply say that the theories are computationally equivalent (Wolfram, 2002), meaning that they give the same answers even though they use different computational methods to arrive at them.

Computational equivalence in this sense can be illustrated more concretely with Feynman's (1985/2006) interesting beans-in-jars method of calculation. If one asks "what is 3 times 15?," the answer can be obtained by the method of multiplication that is taught in elementary school: Line up the two numbers on a piece of paper with the 15 on top, then look up the product of the two rightmost digits in a multiplication table, write the right digit of this product below the existing right digits, carry the 1 over to the next calculation, and so on, resulting in a final answer of 45. An alternative method is to get three jars, fill each with 15 beans, and then dump all the beans onto a table and count them, giving an answer of 45. The second method is logically simpler because it involves only counting, but of course it is more physically cumbersome because it involves jars and beans. Nevertheless, both methods give the same answer, and hence are computationally equivalent. Matrix mechanics,

wave mechanics, and the path integral formalism are computationally equivalent in this sense. As noted earlier, the particular theory a physicist uses is a matter of preference and suitability to the problem at hand (Feynman). The same state of affairs must hold for two behavioral theories that account for the same phenomena with equal fidelity; the two theories must be computationally equivalent.

But if a theory without *Anschaulichkeit* agrees with experimental measurements, then there is another computational equivalence that must exist, namely, that between the theory and the natural world. If both the successful theory and nature give the same answers, then they must be computationally equivalent in the same sense that two theories that give the same answers must be computationally equivalent. Specifically, the material processes that operate in nature must be carrying out the same computations that are carried out by the successful theory, even though the way they are doing so may be quite different (cf. beans in jars to the rules of arithmetic). This point of view is not discussed in physics, but it provides an interesting perspective on *Anschaulichkeit* inasmuch as it makes a statement about the connection between the algorithmic operation of a successful theory and physical reality. The connection does not entail the theory's algorithmic workings themselves, but rather the computations those workings carry out. In this sense, then, nature *is* in a successful theory after all, even if its algorithmic workings cannot reasonably be said to correspond to material processes in nature. In their discussion of behavioral theories, Staddon and Bueno (1991) argued, "if our ... theory really works well as an explanation for behavioral data, it must reflect something true about the neural processes that underlie those data" (p. 6). According to the present point of view, the truths reflected in

such a theory are the computations the neural processes carry out. A principle of computational equivalence also can be seen as a replacement for Putnam's principle, which might be called *assertive realism*. It is not the algorithmic operation of a successful theory that *must* be present in the natural world, it is the set of computations the theory carries out. But a serious problem remains. In what sense can the natural world, or more specifically in the present context, the brain or nervous system, be said to compute anything? Nature is not a computer, and neither is the brain, notwithstanding the popular assumption to the contrary among cognitive psychologists. But it may be that computations are *emergent* outcomes of the material operation of the natural world, and that these computations are what a successful theory captures.

Emergent computation can be illustrated by considering one of the elementary cellular automata, the class of finite state machines discussed earlier. Recall that each cell in an automaton's row of cells can exist in one of two states, such as 1 (black) or 0 (white). The state of each cell in a subsequent row is determined by its state in the current row and by the states of its left and right neighbors in the current row. There are four possible neighbor conditions for a white cell and the same four possible neighbor conditions for a black cell; hence, every cellular automaton has eight possible neighbor conditions. These are listed in the first three columns of Table 1. The "Self" column refers to the cell under consideration; possible states of its left and right neighbors are listed in the first and third columns. Every elementary cellular automaton has these same eight neighbor conditions. The "Next self" column in Table 1 specifies the rule by which a specific cellular automaton operates. It gives the state of the "Self" cell in the next row of cells for every possible neighbor

TABLE 1
Cellular Automaton Rule 152

Neighbor condition			Next self
Left	Self	Right	
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

condition of that cell in the current row of cells. For example, according to the rules in Table 1, a white cell in the current row of cells that has two white neighbors (first row of the table) will remain white in the next row of cells, but if a white cell in the current row of cells has a black left neighbor and a white right neighbor (fifth row of the table), then it will become black in the next row of cells. A cellular automaton is named by converting the binary string in the “Next self” column (from bottom to top) into a decimal integer. In this case, the binary number, 10011000, translates into the decimal integer, 152; hence this automaton is referred to as a Rule 152 cellular automaton, or simply as Rule 152.

Figure 1 shows the time evolution of a Rule 152 cellular automaton for three different inputs or initial conditions. The initial conditions in the left, middle, and right portions of the figure are 6, 10, and 14 consecutive black cells in the first (top) row. The second row of cells was produced by applying the rules in Table 1 to the cells in the first row; the third row was produced by applying the rules in the table to the cells in the second row, and so on, for the 32 rows (or time-steps) shown in the figure. The material operation of this cellular automaton is fully specified by the rules in Table 1. Although there is no information in this table about any computation that may emerge from the application of its rules, it turns out that the rules carry out the computation, $n/2$, where n is the number of consecutive black cells in the initial condition. This can be seen in the steady-state output of the cellular automaton, which consists of diagonal shooters, the number of which equals half the number of adjacent black cells in the initial condition. This calculation is an emergent property of the automaton’s operation. It was discovered empirically, that is, by running the cellular automaton. Taking $n/2$ as our functional theory of this automaton’s

Figure 1. Time evolution of cellular automaton Rule 152 given initial conditions of 6 (left), 10 (center), and 14 (right) adjacent black cells in the first row. Rule 152 “divides by two” when the initial condition is an even number of adjacent black cells. The result is displayed as a set of diagonal black lines, or shooters, the number of which gives the result of the calculation. Adapted from McDowell and Popa’s (2009) Figure 3.

operation, we can predict the steady-state outcome of any set of initial conditions (consisting of an even number of adjacent black cells) from the theory using the rules of arithmetic, even though the automaton itself follows the rules in Table 1. Notice that the latter rules do not appear in our theory, although the computation that emerges from applying them does. Hence, we may say that the theory is computationally equivalent to the material processes described by the Table 1 rules.

This cellular automaton has another important property, which was alluded to above. The computation that emerges from its operation cannot be predicted by examining the automaton itself, that is, by examining the rules in Table 1, even though these rules constitute a complete specification of the automaton's material operation. To learn how the automaton functions, it is necessary to conduct experiments, which in this case means running the automaton with various initial conditions. Given various sets of initial conditions and the corresponding experimental results, it was possible to formulate the functional theory, $n/2$, of the automaton's operation. Further experiments, using odd numbers of consecutive black cells as initial conditions, show that Rule 152 actually carries out a slightly more complicated computation, namely, $(n + n \bmod 2)/2$, where n is the number of consecutive black cells in the initial condition. This reduces to $n/2$ for even n . Interestingly, Rule 152 is the cellular automaton McDowell and Popa (2009) successfully mapped onto the behavior of live organisms in limited ways using initial conditions that consisted of cells colored black or white at random (but cf. McDowell, 2013a). Another interesting point of information is that whereas Rule 152 carries out a relatively simple computation, a different cellular automaton, Rule 110 (with "Next self" column, 01101110), is capable of

universal computation. In other words, Rule 110 can carry out any calculation whatsoever, including logical operations such as AND, NOR, and XOR, as an emergent result of applying the eight simple rules in its table to appropriate initial conditions (McDowell & Popa, 2009).

Let us follow the analogy afforded by Rule 152 a little further. Our complete theory of its operation, given an initial condition of n consecutive black cells, is $(n + n \bmod 2)/2$. With this theory it is possible to accurately predict the steady-state outcome of the cellular automaton for any initial condition of consecutive black cells using the rules of arithmetic. But, as we know, the cellular automaton itself uses the rules listed in Table 1. Would it be possible to somehow incorporate these rules into the theory and therefore obtain a better or more complete theory? This seems unlikely. It is probably the case that the only way to talk reasonably about these rules is simply to state them as facts about the operation of the cellular automaton, which is accomplished by the listing in Table 1.

We now have an understanding of the properties of, and the relations between, the functional theory, $(n + n \bmod 2)/2$, and the material processes the theory is about (the application of the rules in Table 1 to rows of cells). The same properties and relations hold for functional theories of behavior and the material nervous system processes that generate them. Let us summarize these for the behavioral case. A good functional theory will generate the same behavioral output as a live nervous system, although surely by means of different computational machinery. It seems unlikely that actual nervous system functioning can be incorporated into such a theory, although this functioning can be spoken about by stating facts about neuroanatomy, neurophysiology, and neurochemistry. And, just

Figure 2. Material processes and functional theories about their operation. The functional theories do not incorporate the processes themselves, but they do incorporate the emergent computations the processes carry out. The question mark in the center indicates that an atom may not be the sort of material process about which facts can be stated in ordinary terms of space and time.

as was the case for the cellular automaton, it seems unlikely that the way the nervous system generates behavior can be determined just by examining facts about it. Instead, experiments are required that set conditions (e.g., properties of specific environments) on which the nervous system acts to generate behavior. Findings from these experiments may then lead to functional complexity theories, such as the evolutionary theory of behavior dynamics. If a functional complexity theory of behavior is successful, then its algorithmic operation must be computationally equivalent to the material operation of the nervous system. Putting this final, important, point in a different way, it must be the case that a successful functional theory of behavior describes emergent computations carried out by the operation of the brain, just as $(n + n \bmod 2)/2$ describes emergent computations carried out by the operation of the rules in Table 1. The brain in this analogy is as a greatly expanded Table 1 that consists of facts about brain structure and function. For the cellular automaton, Table 1 is known. The brain's Table 1 must be discovered, which is the task of neuroscience.

CONCLUSION

To recapitulate, in science we study material processes that are described by various versions of a generic Table 1, the details of which may or may not be known fully. These material processes are illustrated in the top row of Figure 2 for the three cases discussed in this article: a Rule 152 cellular automaton, an atom, and a brain. The functioning of these processes must be determined by experiments that activate them. The experiments may then lead to functional theories that capture the emergent computations the processes generate, as illustrated in the bottom row of Figure 2. Although such theories may not incorporate the processes themselves in their algorithmic workings, they do make contact with material reality (top row of Figure 2) by incorporating the emergent computations the processes carry out. This is the sense in which nature can be said to appear in the functional theories. Applying these insights specifically to our case study of the evolutionary theory of adaptive behavior, we may conclude that if the theory is correct, then it must be the case that the computations that generate selection, reproduction, and

mutation in the theory are also emergent outcomes of the normal functioning of the nervous system (McDowell, 2010).

Finally, it is interesting to consider more generally how a functional theory about the natural world may be developed, and how this developmental course applies specifically to our case study. There are many possible starting points. In the case of behavior, for example, one might begin with the idea that the brain produces behavior like a cellular automaton produces a pattern of black and white cells, or with the idea that the brain produces behavior like an artificial neural network produces an output, or with the idea that the brain works like a population of oscillators to generate behavior (Pringle, 1951), and so on. In his seventh Messenger lecture at Cornell University in 1964, Feynman said that the development of a functional theory or law in science begins with a guess (Feynman Chaser, 2008; Microsoft Corporation, 2009). Although such a guess can be made at random, in most cases it probably is not. Instead, the scientist probably has some idea, based on intuition or prior knowledge, or perhaps on experimental measurements, that a particular guess might prove fruitful. In behavior analysis, the guess that behavior is the result of processes of variation and selection dates back at least as far as Thorndike (Donahoe, 1999; Henriques, 2003), and has been discussed extensively for 60 years or more (e.g., Edelman, 1987; Hayek, 1952a, 1952b; Pringle, 1951; Skinner, 1981; Staddon & Simmelhag, 1971). The popularity of this guess is no doubt due to the fact that it tells a plausible and compelling story about adaptive behavior, a story that fits well with other laws and principles in biology (Skinner, 1981; Staddon & Simmelhag, 1971). According to Feynman (Feynman Chaser, 2008; Microsoft Corporation, 2009), the process of theory development continues by working

out the consequences of the guess and then comparing them to experimental measurements. The consequences of the guess that behavior is the result of processes of variation and selection are worked out by the evolutionary theory, and the comparison of these consequences with experiment so far has been favorable. Furthermore, based on the principle of computational equivalence, we may conclude at least provisionally that the nervous system must generate behavior by emergent processes of variation and selection, which are captured by the algorithmic workings of the evolutionary theory.

REFERENCES

- Baghramian, M. (2008). "From realism back to realism": Putnam's long journey. *Philosophical Topics*, 36, 17-35.
- Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior*, 12, 861-874.
- Cassidy, D. (1993). *Uncertainty: The life and science of Werner Heisenberg*. New York, NY: Freeman. Retrieved from <http://math.ucr.edu/home/baez/photon/anschaulichkeit.htm>
- Cording, J. R., McLean, A. P., & Grace, R. C. (2011). Testing the linearity and independence assumptions of the generalized matching law for reinforcer magnitude: A residual meta-analysis. *Behavioural Processes*, 87, 64-70.
- Cox, B., & Forshaw, J. (2012). *The quantum universe: Everything that can happen does happen*. Boston, MA: Da Capo Press.
- Davison, M., & Baum, W. M. (2000). Choice in a variable environment: Every reinforcer counts. *Journal of the Experimental Analysis of Behavior*, 74, 1-24. doi: 10.1901/jeab.2000.74-1
- DickPPPPP. (2009, April 26). *You don't like it? Go somewhere else! by Richard Feynman, the QED Lecture at University of Auckland*. Retrieved from <http://www.youtube.com/watch?v=iMDTcMD6pOw>
- Donahoe, J. W. (1999). Edward L. Thorndike: The selectionist connectionist. *Journal of the Experimental Analysis of Behavior*, 72, 451-454.
- Donahoe, J. W., Burgos, J. E., & Palmer, D. C. (1993). A selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior*, 60, 17-40.
- Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Boston, MA: Allyn & Bacon.

- Donahoe, J. W., Palmer, D. C., & Burgos, J. E. (1997). The S-R issue: Its status in behavior analysis and in Donahoe and Palmer's *Learning and Complex Behavior*. *Journal of the Experimental Analysis of Behavior*, 67, 193–211.
- Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York, NY: Basic Books.
- Feynman, R. P. (2006). *QED: The strange theory of light and matter*. Princeton, NJ: Princeton University Press. (Original work published 1985)
- Feynman, Chaser. (2008, July 13). *Feynman chaser—the key to science*. Retrieved from <http://www.youtube.com/watch?v=b240PGCMwV0>
- Harman, P. M. (2001). *The natural philosophy of James Clerk Maxwell*. Cambridge, UK: Cambridge University Press.
- Hayek, F. A. (1952a). *The counter-revolution of science: Studies on the abuse of reason*. London, UK: The Free Press.
- Hayek, F. A. (1952b). *The sensory order: An inquiry into the foundations of theoretical psychology*. Chicago, IL: University of Chicago Press.
- Heisenberg, W. (2007). Quantum-theoretical re-interpretation of kinematic and mechanical relations. In B. L. van der Waerden (Ed.), *Sources of quantum mechanics* (pp. 261–276). Mineola, NY: Dover. (Original work published 1925) Retrieved from <http://math.ucr.edu/home/baez/photon/anschaulichkeit.htm>
- Henriques, G. (2003). The tree of knowledge system and the theoretical unification of psychology. *Review of General Psychology*, 7, 150–182.
- Hill, C. S. (Ed.). (1992). *The philosophy of Hilary Putnam*. Fayetteville, AR: University of Arkansas Press.
- Holland, J. H. (1998). *Emergence*. Cambridge, MA: Perseus.
- Kehoe, E. J. (1989). Connectionist models of conditioning: A tutorial. *Journal of the Experimental Analysis of Behavior*, 52, 427–440.
- Kulubekova, S., & McDowell, J. J. (2008). A computational model of selection by consequences: Log survivor plots. *Behavioural Processes*, 78, 291–296.
- Kulubekova, S., & McDowell, J. J. (2013). Computational model of selection by consequences: Patterns of preference change on concurrent schedules. *Journal of the Experimental Analysis of Behavior*, 100, 147–164.
- Marr, M. J. (1997). The eternal antithesis: A commentary on Donahoe, Palmer, and Burgos. *Journal of the Experimental Analysis of Behavior*, 67, 232–235.
- Marr, M. J. (2000). What is the net worth? Some thoughts on neural networks and behavior. *Mexicana de Análisis de la Conducta*, 26, 273–287.
- McDowell, J. J. (2004). A computational model of selection by consequences. *Journal of the Experimental Analysis of Behavior*, 81, 297–317.
- McDowell, J. J. (2010). Behavioral and neural Darwinism: Selectionist function and mechanism in adaptive behavior dynamics. *Behavioural Processes*, 84, 358–365.
- McDowell, J. J. (2013a). Calculi of complexity: How phenomena emerge from rules. A review of *Complexity: A Guided Tour* by Melanie Mitchell. *Journal of the Experimental Analysis of Behavior*, 99, 234–244.
- McDowell, J. J. (2013b). On the theoretical and empirical status of the matching law and matching theory. *Psychological Bulletin*, 139, 1000–1028.
- McDowell, J. J., & Caron, M. L. (2007). Undermatching is an emergent property of selection by consequences. *Behavioural Processes*, 75, 97–106.
- McDowell, J. J., Caron, M. L., Kulubekova, S., & Berg, J. P. (2008). A computational theory of selection by consequences applied to concurrent schedules. *Journal of the Experimental Analysis of Behavior*, 90, 387–403.
- McDowell, J. J., & Popa, A. (2009). Beyond continuous mathematics and traditional scientific analysis: Understanding and mining Wolfram's *A New Kind of Science*. *Behavioural Processes*, 81, 343–352.
- McDowell, J. J., & Popa, A. (2010). Toward a mechanics of adaptive behavior: Evolutionary dynamics and matching theory statics. *Journal of the Experimental Analysis of Behavior*, 94, 241–260. doi: 10.1901/jeab.2010.94-241
- McDowell, J. J., Popa, A., & Calvin, N. T. (2012). Selection dynamics in joint matching to rate and magnitude of reinforcement. *Journal of the Experimental Analysis of Behavior*, 98, 199–212. doi: 10.1901/jeab.2012.98-199
- Microsoft Corporation. (2009). *Project Tuva*. Retrieved from <http://research.microsoft.com/apps/tools/tuva/index.html#>
- Moore, W. (1992). *Schrödinger: Life and thought*. Cambridge, UK: Cambridge University Press. Retrieved from <http://math.ucr.edu/home/baez/photon/anschaulichkeit.htm>
- Popa, A., & McDowell, J. J. (2010). The effect of Hamming distances in a computational model of selection by consequences. *Behavioural Processes*, 84, 428–434.
- Pringle, J. W. S. (1951). On the parallel between learning and evolution. *Behaviour*, 3, 174–215.
- Putnam, H. (1985). *Mathematics, matter and method. Philosophical papers, Volume 1* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton, NJ: Princeton University Press.
- Skinner, B. F. (1981). Selection by consequences. *Science*, 213, 501–504.
- Staddon, J. E. R., & Bueno, J. L. O. (1991). On models, behaviorism and the neural basis of learning. *Psychological Science*, 2, 3–11.

- Staddon, J. E. R., & Simmelhag, V. (1971). The "superstition" experiment: A reexamination of its implications for the principles of adaptive behavior. *Psychological Review*, 78, 3-43.
- Tononi, G., Sporns, O., & Edelman, G. M. (1992). Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex*, 2, 310-335.
- van Dongen, J. (2010). *Einstein's unification*. Cambridge, UK: Cambridge University Press.
- Wofle, D. (Ed.). (1959). *Symposium on basic research*. Washington, DC: American Association for the Advancement of Science. Retrieved from http://en.wikiquote.org/wiki/Niels_Bohr
- Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.