Design, Observation, Surprise! A Test of Emergence

Edmund M. A. Ronald
Centre de Mathématiques Appliquées
Ecole Polytechnique
91128 Palaiseau Cedex, France
eronald@cmapx.polytechnique.fr.

Moshe Sipper
Mathieu S. Capcarrère
Logic Systems Laboratory
Swiss Federal Institute of Technology
IN-Ecublens, CH-1015 Lausanne, Switzerland
{name.surname}@epfl.ch.

Abstract The field of artificial life (Alife) is replete with documented instances of *emergence*, though debate still persists as to the meaning of this term. We contend that, in the absence of an acceptable definition, researchers in the field would be well served by adopting an emergence certification mark that would garner approval from the Alife community. Toward this end, we propose an *emergence test*, namely, criteria by which one can justify conferring the emergence label.

Keywords
artificial life, emergence, Turing test

1 Introduction

When a bank's accounting program goes seemingly independent and does its own thing, the programmer scratches his head, sighs, and prepares for doing overtime with the debugger. But when a society of agents does something surprising, Alife researchers may solemnly document this "emergent behavior," and move on to other issues without always seeking to determine the cause of their observations. Indeed, overly facile use of the term emergence has made it controversial. Arkin recently observed that:

> Emergence is often invoked in an almost mystical sense regarding the capabilities of behavior-based systems. Emergent behavior implies a holistic capability where the sum is considerably greater than its parts. It is true that what occurs in a behavior-based system is often a surprise to the system's designer, but does the surprise come because of a shortcoming of the analysis of the constituent behavioral building blocks and their coordination, or because of something else? ([1], p. 105).

Altogether, it seems the emergence tag has become a great attention grabber, thanks to the striking behaviors demonstrated in artificial life experiments. We do not think, however, that emergence should be diagnosed ipso facto whenever the unexpected intrudes into the visual field of the experimenter; nor should the diagnosis of emergence immediately justify an economy of explanation. Such abuse and overuse of the term eventually will devalue its significance, and bring work centered on emergence into disrepute. Therefore, we contend that, in the absence of an acceptable definition, researchers in the field would be well served by adopting an emergence certification mark that would garner approval from the Alife community.

Motivated by this wish to standardize the tagging task, we propose an *emergence test*, namely, criteria by which one can justify conferring the emergence label [24]. Our criteria are motivated by an examination of published work in the field of Alife.

The emergence test is presented in the next section and followed in Section 3 by a host of case studies demonstrating its applicability. Finally, in Section 4, we discuss our

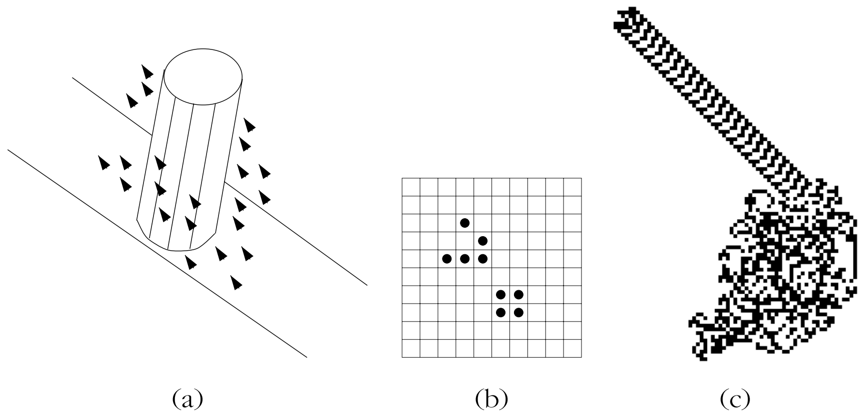


Figure 1. Examples of emergence. (a) A flock of simulated birds parts smoothly when faced with an obstacle, and "flows" around it—to then reunite again (after Reynolds [23]). (b) Two Game-of-Life patterns, known as a "glider" and a "block." (c) The trail created by the highway-constructing Langton ant.

test in the light of previous work, and establish that it indeed articulates and clarifies existing views on the matter of emergence.

2 An Operant Definition of Emergence for Alife Researchers

2.1 Examples of Emergence in the Alife Literature

Before presenting the emergence test itself, we outline below a sampling of work in the field, which serves to delineate the scope of our investigation. The reader is urged to keep these in mind while perusing the emergence test, described in Section 2.4.

- Emergence of flocking behavior in simulated birds, from a set of three simple steering behaviors (Figure 1a) [23].
- Emergence of wall-following behavior in an autonomous, mobile robot, from the simultaneous operation of two simple behavior systems: obstacle avoidance and wall seeking [29].
- Emergence of cooperation in the iterated prisoner's dilemma, from the application of simple game strategies [3].
- Emergence of self-replicating structures from simple basic components [10, 19, 27].
- Emergence of a menagerie of patterns in the Game of Life (e.g., gliders, spaceships, puffer trains) from simple, local rules (Figure 1b) [4].
- Emergence of team behavior (foraging, flocking, consuming, moving material, grazing) in autonomous, mobile robots, from simple rules [2].
- Emergence of social structures and group behaviors in the artificial society of "Sugarscape," from the interactions of individuals (agents) [14].
- Emergence of a "highway" created by the artificial Langton ant, from simple movement rules (Figure 1c) [30].
- Emergence of complex behaviors in machines known as Braitenberg vehicles, from simple internal structures and mechanisms [7].

- Emergence of a nest structure in a simulated wasp colony, from the interactions taking place between individual wasps [31].
- Emergence of a solution to a character-recognition problem in artificial neural networks, from the interactions of the individual neurons.
- Emergence of a solution to the density problem in cellular automata, from simple, local interactions (Figure 2) [8, 26, 28].
- Minsky's theory, according to which mind emerges from a society of myriad, mindless components [21].

2.2 Artificial Life as a Science of the Artificial

As seen from the examples in the previous subsection, we are restricting the scope of our study of emergence to instances of artificial life. Alife is a constructive endeavor: Some researchers aim at evolving patterns in a computer; some seek to elicit social behaviors in real-world robots; others wish to study life-related phenomena in a more controllable setting, while still others are interested in the synthesis of novel lifelike systems in chemical, electronic, mechanical, and other artificial media. Alife is an experimental discipline, fundamentally consisting of the observation of run-time behaviors, those complex interactions generated when populations of man-made, artificial creatures are immersed in real or simulated environments. Published work in the field usually relates the conception of a model, its instantiation into real-world or simulated objects, and the observed behavior of these objects in a collection of experiments.

The field of artificial life thus quintessentially exemplifies a science of the artificial, as it accords with the four indicia given by Herbert Simon in his influential work, *The Sciences of the Artificial* ([25], p. 8):

1. Artificial things are synthesized (though not always or usually with full forethought) by man.
2. Artificial things may imitate appearances in natural things while lacking, in one or many respects, the reality of the latter.
3. Artificial things can be characterized in terms of functions, goals, adaptation.
4. Artificial things are often discussed, particularly when they are being designed, in terms of imperatives as well as descriptives.

We agree with Simon in that the artificial can—and should—be treated differently from the natural. In this spirit, our emergence test below is aimed only at the artificial, and in particular at artificial life.

2.3 Preliminary Remarks on the Ontology of Emergence

Before formulating a definition of emergence, let us make some preliminary remarks:

- Emergence is not a specific thing; for example, like a stone in your pocket.
- Emergence is not one specific behavior, known to occur at some moment in time.
- Neither is emergence a well-defined category, like the stones on some particular beach.
- The category of all emergent behaviors is of interest; yet debate will occur on which behaviors should be included, and which should be excluded.

- Hence it is each observer who decides to include or exclude a given behavior in his own category of emergent behaviors.

The description of a phenomenon as emergent is contingent, then, on the existence of an observer; being *a visualization constructed in the mind of the observer*, emergence can be described as a *concept*, like beauty or intelligence. Such concepts are slippery.

2.4 Formulating the Emergence Test

The difficulties we face in adopting a definition of the concept of emergence are reminiscent of the complications faced by early Artificial Intelligence (AI) researchers in defining intelligence. Nonetheless, where the equally elusive concept of intelligence is concerned, Alan Turing found a way to cut the Gordian knot, by means of an *operant definition* that is useful *within the limited context of man-machine interaction* [32]. Debate concerning the concept of intelligence is unlikely to subside in the foreseeable future, and the same, we believe, holds for emergence. We deem, however, that viewing the world through Turing colored glasses might improve our vision as regards the concept of emergence—at least where modern-day Alife practice is concerned.

The Turing test focuses on a human experimenter's incapacity at discerning human from machine when holding what we now would call an Internet chat session. Our emergence test centers on an observer's avowed incapacity (amazement) to reconcile his perception of an experiment in terms of a global world view with his awareness of the atomic nature of the elementary interactions.

Assume that the scientists attendant upon an Alife experiment are just two: a system designer and a system observer (both of whom in fact can be one and the same), and that the following three conditions hold:

1. *Design*: The system has been constructed by the designer, by describing *local* elementary interactions between components (e.g., artificial creatures and elements of the environment) in a language $\mathcal{L}_1$.
2. *Observation*: The observer is *fully aware* of the design, but describes *global* behaviors and properties of the running system, over a period of time, using a language $\mathcal{L}_2$.
3. *Surprise*: The language of design $\mathcal{L}_1$ and the language of observation $\mathcal{L}_2$ are distinct, and the causal link between the elementary interactions programmed in $\mathcal{L}_1$ and the behaviors observed in $\mathcal{L}_2$ is *non-obvious* to the observer—who therefore experiences surprise. In other words, there is a cognitive dissonance between the observer's mental image of the system's design stated in $\mathcal{L}_1$ and his contemporaneous observation of the system's behavior stated in $\mathcal{L}_2$.

   When assessing this clause of our test one should bear in mind that as human beings we are quite easily surprised (as any novice magician will attest). The question reposes rather on how *evanescent* the surprise effect is; that is, how easy (or strenuous) it is for the observer to bridge the $\mathcal{L}_1$–$\mathcal{L}_2$ gap, thus reconciling his global view of the system with his awareness of the underlying elementary interactions. One can draw an analogy with the concept of intelligence and the Turing test: the chatty terminal at first might appear to be carrying on like an intelligent interlocutor, only to lose its "intelligence certificate" once the tester has pondered upon the true nature of the ongoing conversation.

The above three clauses relating design, observation, and surprise describe our conditions for diagnosing emergence, that is, for accepting that a system is displaying emergent behavior. Some of the above points deserve further elaboration, or indeed invite debate. Before treating these issues in Section 4, we wish to demonstrate the application of our test to several cases.

3 Administering the Emergence Test: Case Studies

In this section we administer the emergence test to eight examples (some of which are taken from Section 2.1), thus demonstrating its application. Each example ends with a "test score," a diagnosis constituting our own assertion as observers of whether we are indeed surprised, that is, of whether emergent behavior is indeed displayed—or not.

3.1 Emergence of a Nest Structure in a Simulated Wasp Colony, from the Interactions Taking Place between Individual Wasps [31]

- *Design*: The design language $\mathcal{L}_1$ is that of local wasp interactions, including movement on a three-dimensional cubic lattice and placement of bricks. A wasp's decision is based upon a local configuration of bricks, which lie in its "visual" field. Actions to be taken are prewired under the form of a lookup table with as many entries as there are stimulating configurations.
- *Observation*: The observation language $\mathcal{L}_2$ is that of large-scale geometry, as employed to describe nest architectures.
- *Surprise*: While fully aware of the underlying wasp interaction rules, the observer nonetheless marvels at the sophistication of the constructions and at their striking similarity to naturally occurring nests.
- ? *Diagnosis*: Emergent behavior is displayed by the nest-building wasps.

3.2 Emergence of a "Highway" Created by the Artificial Langton Ant, from Simple Movement Rules [30]

- *Design*: The design language $\mathcal{L}_1$ is that of single moves of a simple, myopic ant. The ant starts out on the central cell of a two-dimensional, rectangular lattice, heading in some selected direction. It moves one cell in that direction and looks at the color of the cell it lands on—black or white. If it lands on a black cell, it paints it white and turns 90 degrees to the left; if it lands on a white cell, it paints it black and turns 90 degrees to the right. These simple rules are iterated indefinitely.
- *Observation*: The observation language $\mathcal{L}_2$ is that of global behavioral patterns, extended over time and space (i.e., tens of thousands of single ant moves, spanning thousands of cells). Specifically, the ant was observed to construct a "highway," that is, a repeating pattern of fixed width that extends indefinitely in a specific direction (Figure 1c).
- *Surprise*: While fully aware of the very simple ant rules, the observer is nonetheless surprised by the appearance of a highway.
- ? *Diagnosis*: Emergent behavior is displayed by the highway-constructing ant.

3.3 Emergence of a Menagerie of Patterns in the Game of Life (e.g., Gliders, Spaceships, Puffer Trains), from Simple, Local Rules [4]

- *Design*: The Game of Life is played out on a two-dimensional, rectangular lattice, each cell of which can be colored either white or black (as with the Langton ant). The design language $\mathcal{L}_1$ is that of local color changes; this language is employed to delineate the rules according to which a cell changes its color in light of its immediate surrounding cells.
- *Observation*: The observation language $\mathcal{L}_2$ is that of global behavioral patterns, including such observed structures as gliders, spaceships, and puffer trains (Figure 1b).
- *Surprise*: While fully aware of the simple color-transformation rules, the observer is nonetheless amazed by the appearance of this bestiary of critters.
- ? *Diagnosis*: Emergent behavior is displayed by numerous instantiations of the Game of Life.

3.4 Emergence of Complex Behaviors in Machines Known as Braitenberg Vehicles, from Simple Internal Structures and Mechanisms [7]

- *Design*: The design language $\mathcal{L}_1$ is that of simple internal structures and mechanisms (sensors, actuators, and computational devices).
- *Observation*: The observation language $\mathcal{L}_2$ is that of global behavioral patterns, to which Braitenberg playfully ascribed such anthropomorphic terms as "fear," "aggression," and "love."
- *Surprise*: While fully aware of the vehicles' simple internal workings, the observer is nonetheless amazed by the appearance of lifelike behaviors. This is true for vehicles 3 through 14; on the contrary, the behaviors of vehicles 1 and 2 can be straightforwardly divined by the observer.
- ? *Diagnosis*: Emergent behavior is displayed where vehicles 3 through 14 are concerned, while vehicles 1 and 2 are non-emergent.

3.5 Minsky's Theory, According to which Mind Emerges from a Society of Myriad, Mindless Components [21]

- *Design*: The design language $\mathcal{L}_1$ is that of simple (putative) processes, which Minsky calls agents.
- *Observation*: The observation language $\mathcal{L}_2$ is the common language of discourse used to describe intelligent, human behavior (examples from Minsky's book are nonverbal reasoning, language learning, and humor [21]).
- ? *Surprise and Diagnosis*: Mind is an emergent phenomenon, *par excellence*, since the observer always marvels at its appearance.

3.6 Emergence of Flocking Behavior in Simulated Birds, from a Set of Three Simple Steering Behaviors [23]

- *Design*: The design language $\mathcal{L}_1$ is that of local bird interactions, the three rules being: *separation*—steer to avoid crowding local flockmates; *alignment*—steer toward the average heading of local flockmates; *cohesion*—steer to move toward the average position of local flockmates. A bird's decision is based upon its nearby neighbors, that is, those that are in its "visual" field.

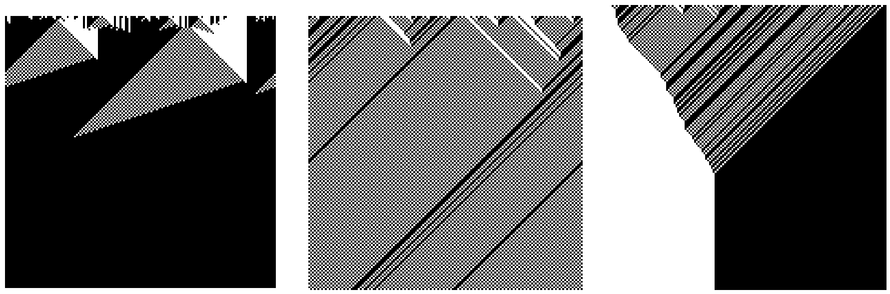
- *Observation*: The observation language $\mathcal{L}_2$ is that of flocking behaviors, such as the flock's parting smoothly when faced with an obstacle, and "flowing" around it—to then reunite again (Figure 1a).
- *Surprise*: While fully aware of the underlying bird interaction rules, the observer nonetheless marvels at the lifelike flocking behaviors.
- ? *Diagnosis*: The flocking behavior exhibited by the artificial birds was considered a clear case of emergence when it first appeared in 1987. However, one now could maintain that it no longer passes the emergence test, since widespread use of this technique in computer graphics has obviated the element of surprise. This example demonstrates that the diagnosis of emergence is contingent upon the sophistication of the observer.

3.7 Emergence of Wall-Following Behavior in an Autonomous, Mobile Robot, from the Simultaneous Operation of Two Simple Behavior Systems: Obstacle Avoidance and Wall Seeking [29]

- *Design*: The design language $\mathcal{L}_1$ is that of simple robot behaviors, including—in this case—obstacle avoidance and wall seeking.
- *Observation*: The observation language $\mathcal{L}_2$ is that of more elaborate robot behaviors, consisting—in this case—of wall following.
- *Surprise*: Steels wrote that "Wall following is emergent in this case because the category 'equidistance to the (left/right) wall' is not explicitly sensed by the robot or causally used in one of the controlling behavior systems" ([29], p. 92).
- ? *Diagnosis*: Steels diagnosed emergence in this case as it accords with his own definition, namely, that a behavior is emergent if it necessitates the use of new descriptive categories that are not needed to describe the behavior of the constituent components [29]. While thus alluding to the language dichotomy rendered explicit by our definition (i.e., the existence of two distinct languages—that of design and that of observation), we maintain that the surprise element is missing: The wall-following behavior can be quite readily deduced by an observer aware of the two underlying simpler behaviors. We thus conclude that emergent behavior is *not* displayed by the wall-following robot.

3.8 Emergence of a Solution to the Density Problem in Cellular Automata, from Simple, Local Interactions [8, 26, 28]

This final example is delineated in more detail than the previous ones so as to demonstrate a number of interesting points concerning our test. The example is based on the model known as cellular automata (CA), originally conceived by Ulam and von Neumann in the 1940s to provide a formal framework for investigating the behavior of complex, extended systems [27, 33]. CAs have been widely used over the years in many fields of inquiry, including physics, biology, and computer science; in particular, they figure prominently in Alife research. We first describe briefly the workings of a CA, followed by the presentation of a specific problem, known as density classification, which has received much attention in the CA literature. We delineate three CA solutions



(I) (II) (III)

Figure 2. CA solutions to three versions of the density classification problem. The one-dimensional grid is of size $N = 149$ cells. White squares represent cells in state 0, black squares represent cells in state 1. The two-dimensional images shown above are so-called space-time diagrams, a common method of depicting the behavior of one-dimensional CAs over time. In the images, the horizontal axis depicts the configuration of states at a certain time t, and the vertical axis depicts successive time steps (time thus increases down the page). The initial density in all three examples is > 0.5.

to this problem, concluding that the first two pass the emergence test while the last one does not.

Cellular automata are discrete, dynamical systems that perform computations in a distributed fashion on a spatially extended grid. A cellular automaton consists of an array of cells, each of which can be in one of a finite number of possible states, updated synchronously in discrete time steps according to a local, identical interaction rule [26]. The *state* of a cell at the next time step is determined by the current states of a surrounding neighborhood of cells. This transition is usually specified in the form of a *rule table*, delineating the cell's next state for each possible neighborhood configuration. The cellular array (grid) is n-dimensional, where $n = 1, 2, 3$ is used in practice.

CAs are one of the prime models used to study emergent behavior and computation. One oft-cited problem involving (putative) emergent computation is for a CA to determine the global density of bits in an initial state configuration. This problem, known as density classification, has been studied intensively over the past few years [26]. In a recent paper, Sipper, Capcarrère, and Ronald [28] described two previous versions of the problem along with their CA solutions, and then went on to show that there exists yet a third version—which admits a simple solution. Below, we summarize their results, after which we will administer the emergence test to all three versions.

3.8.1 Version I

In the original statement of the problem [22], a one-dimensional, two-state CA (meaning that each cell can be in one of two states, 0 or 1) is presented with an arbitrary initial configuration of states (the input). The CA then should converge in time to a state of all 1s if the initial configuration contains a density of 1s > 0.5 (i.e., a majority of 1s), and converge to all 0s if this density < 0.5 (i.e., a majority of 0s); for an initial density of 0.5, the CA's behavior is undefined (Figure 2(I)). The final configuration is considered as the output of the computation. Spatially periodic boundary conditions are used, resulting in a circular grid. It has been proven that no perfect CA solution exists for this problem version, though high-performance CAs have been designed by hand as well as found by means of artificial evolution [26] (these CAs do not perform perfect classification, i.e., they misclassify some of the initial configurations; the CA solution demonstrated in Figure 2(I), known as the GKL rule, in fact does not classify correctly all initial configurations).

3.8.2 Version II

Capcarrère, Sipper, and Tomassini [8] showed that a perfect, one-dimensional, two-state CA density classifier does exist, upon defining a different output specification (again, periodic boundary conditions are assumed). This CA is demonstrated in Figure 2(II): Upon presentation of an arbitrary initial configuration, the N-cell grid relaxes to a limit-cycle, within $\lceil N/2 \rceil$ time steps, that provides a classification of the initial configuration's density of 1s. If this density > 0.5 (respectively, < 0.5), then the final configuration consists of one or more blocks of at least two consecutive 1s (0s), interspersed by an alternation of 0s and 1s; for an initial density of exactly 0.5, the final configuration consists of an alternation of 0s and 1s. The computation's output is given by the state of the consecutive block (or blocks) of same-state cells: If the same-state cells are in state 1 (respectively, 0), then this signifies a majority of 1s (0s) in the input; if there is but an alternation of 0s and 1s, then this signifies that the input contains an equal number of 0s and 1s.

3.8.3 Version III

More recently, Sipper, Capcarrère, and Ronald [28] described yet another modification of the original problem (version I), with a different output specification, as well as fixed boundary conditions, rather than the periodic ones previously assumed. These two modifications give rise to a simple density classifier, demonstrated in Figure 2(III): The CA of size N (boundary cells excluded) converges in at most $N - 1$ time steps to a configuration $0^\alpha 1^\beta$, where α, β denote the number of 0s and 1s at time step 0, respectively; $\alpha, \beta \in \{0, \ldots, N\}$, $\alpha + \beta = N$. For N odd, the density classification of the input is attained by considering the middle cell's final state: 0 signifies a majority of 0s in the input, 1 signifies a majority of 1s; for N even, we consider the two middle cells: 00 signifies a majority of 0s in the input, 11 signifies a majority of 1s, 01 signifies equality, and 10 is impossible. This example in fact initiated our current study of emergence, ultimately culminating in the proposed emergence test.

3.8.4 Emergent or Not?

As CA researchers, we observed the three versions depicted in Figure 2, experiencing unease with respect to version III—as though we were "cheating." And yet, why do the first two solutions seem complex and emergent, whereas the third one seems simple and non-emergent?[1] Density is a global property of a configuration (the 1s can be distributed throughout the grid), whereas a CA relies solely on local interactions; this property holds true for all three versions of the problem. Moreover, the three CA solutions presented are all legal, in that they violate none of the CA principles of operation. In short, we are faced here with three perfectly valid solutions (albeit an approximate one where version I is concerned).

The crux of the matter lies in the surprise phase of the emergence test: we maintain that there is no (lasting) surprise where version III is concerned, whereas we are surprised by versions I and II. The language of design $\mathcal{L}_1$ and the language of observation $\mathcal{L}_2$ are obviously distinct in all three cases, and yet the causal link between the elementary interactions programmed in $\mathcal{L}_1$ and the behaviors observed in $\mathcal{L}_2$ is rather straightforward for version III, whereas it is non-obvious for versions I and II. An observer even but slightly versed in CA dynamics will divine quickly the workings of the version-III CA (whose rule is fully specified in [28]): basically, one can picture it as a horizontally-oriented, transparent tube full of tennis balls, where each ball represents a state of 1 (the absence of a ball represents a state of 0). If one places the tube in the

1 We note in passing that this "feeling" is also supported by the complexity of the proofs concerning the behaviors of CA versions I and II, versus the simplicity of the proof of CA III's behavior [8, 28].

vertical position, the balls will roll down, with classification of the initial density then given by simply observing the absence or presence of balls in the central section of the tube. The CA rule in question implements this metaphor (the fixed boundary conditions are crucial in that they stop the balls dead at the tube's end). It is the quickly fading surprise effect experienced with the version-III solution that evoked in us the uneasy feeling that something was not "right." Armed with the emergence test, we can now pinpoint the source of our uneasiness: Version III is non-emergent (no lasting surprise) whereas versions I and II are (surprise!). (We emphasize again that version III—while non-emergent—is "right" in the sense that it is a bona fide solution, which perfectly accords with the CA principles.) This example illustrates the subtleties involved in our test. Our conclusions regarding this CA example are recapitulated below within the emergence-test framework:

- *Design*: The design language $\mathcal{L}_1$ is that of local CA rules.
- *Observation*: The observation language $\mathcal{L}_2$ is that of global behavioral patterns. Specifically, the patterns of interest involve those that represent a solution to the density classification problem.
- *Surprise*: While fully aware of the underlying CA rules, the observer is nonetheless puzzled by the intricate behaviors of versions I and II, whereas version III straightforwardly yields its "secret."
- ? *Diagnosis*: Emergent behavior is displayed where versions I and II are concerned, while version III is non-emergent.

4 Discussion

Our conception of the emergence test builds on a number of ideas that have been addressed over the years in the literature; the relationship between our test and these works is described in this section. While a number of researchers have remarked upon some or other ingredient of our definition, to the best of our knowledge nobody has yet put together all the constituent elements into an emergence test such as ours. Moreover, those elements that have been discussed in the literature are highly intermingled—which has rendered it impossible for us to separate the discussion below into totally orthogonal clauses; indeed, untangling this vicious circle of reasoning is, in our opinion, one of the major contributions of our test.[2]

4.1 The Operant Nature of the Test

In this we have drawn our inspiration from Turing, who—concerning intelligence—opted for an operant, informal, "social" definition, deliberately eschewing rigor. Turing's definition has served the AI community well, and it is still considered one of the seminal works in the field, almost half a century after its publication [32].

4.2 Emergence as a Property of Artificial Systems

In his book, *Emergence: From Chaos to Order*, Holland wrote that "Emergence occurs in systems that are generated" ([17], p. 225). Reviewing Holland's book, Mallot also opined that "In this context [the construction of artificial systems], the problem of emergence may actually be a genuine one" [20]. As noted in Section 2.2, we have chosen to limit

2 This section is by no means intended to serve as a review on the subject of emergence. Rather, we have cited what we believe to be major works in this area that tie in with our emergence test. For a good critical review, the reader is referred to Bonabeau, Dessalles, and Grumbach [5].

the scope of our test to the sciences of the artificial, and in particular to artificial life; this restriction is embodied in clause (1) of the test.

4.3 The Existence of an Observer

Artificial systems are constructed to be beheld—usually one does not build one's system, then walk away nonchalantly without ever looking back. Hence, there exists an observer ipso facto (who need not necessarily be the constructor himself), a fundamental aspect that has not escaped researchers in the field. In a paper discussing emergence and artificial life, Cariani wrote that "The interesting emergent events that involve artificial life simulations reside not in the simulations themselves, but in the ways that they change the way we think and interact with the world" ([9], p. 790). He goes on to say that "computer simulations are catalysts for emergent processes in our own minds ..." ([9], p. 790).

Another author, Emmeche, in an introductory monograph on artificial life, examines the case for emergence "in the eye of the beholder" ([13], p. 145). Also, Crutchfield, in an article devoted to the subject of emergence, asks: "But for whom has the emergence occurred? More particularly, to whom are the emergent features 'new'? ... The newness in both cases is in the eye of an observer ..." ([12], p. 517).

Bonabeau, Dessalles, and Grumbach, in an article presenting a conceptual framework for characterizing emergent phenomena, noted the difference between "actors: interacting agents with *local perception* and the ability to *act locally*" and "spectators: one or several entities sensitive to the emergent phenomenon, and possessing *global perception*" ([6], p. 348). They wrote that "the emergent aspect of a phenomenon is related to the point of view of an observer of this phenomenon: it is not intrinsic to the phenomenon, but related to the global system (phenomenon + observer)" ([6], pp. 348–349).

Holland brings up the issue of the observer circuitously, when writing that "The whole is more than the sum of the parts in these generated systems. ... Said another way, there are regularities in system behavior that are not revealed by direct inspection of the laws satisfied by the components" ([17], p. 225). One may ask *direct inspection by whom?* Why, by the observer of course![3] Clearly, the existence of an observer is a sine qua non for the issue of emergence to arise at all.

We wish to point out that one can make a case for the analogy between the concepts of emergence and complexity, as regards the presence of a baffled observer. For example, Kolen and Pollack, considering the highly formal notion of computational complexity, wrote: "Computational complexity, often used to separate cognitive behaviors from other types of animal behavior, will be shown to be dependent upon the observation mechanism as well as the process under examination" ([18], p. 254).

4.4 The Language Dichotomy

A number of authors have alluded to the existence of a language of observation as distinct from the language of design. In his paper on behavior-oriented artificial intelligence, Steels put forward a definition of emergence, writing that "A behavior is emergent if new categories are needed to describe this underlying regularity that are not needed to describe the behaviors (i.e., the regularities) generated by the underlying behavior systems on their own" ([29], p. 89).

In a book, entitled *Frontiers of Complexity*, Coveney and Highfield, upon discussing the behavior of collective systems of simple, interacting units, wrote that "Their inter-

3 Holland also cites a passage from Gell-Mann's book, *The Quark and the Jaguar* [15], which brings up indirectly the role of the observer: "In an astonishing variety of contexts, apparently complex structures or behaviors emerge from systems characterized by simple rules." ([17], p. 238). Gell-Mann's use of the qualifier "apparently" suggests that the quality in question necessitates a judgment call—that is, an observer.

actions lead to coherent collective phenomena, so-called emergent properties that can be described only at higher levels than those of the individual units" ([11], p. 7).

Holland emphasized the distinct nature of these two languages, noting that one can "converse" in the language of observation without resorting to the language of design: "When a macrolaw can be formulated, the behavior of the whole pattern can be described without recourse to the microlaws (generators and constraints) that determine the behavior of its components" ([17], p. 227).

4.5 The Observer's Reasoning Abilities

Writing on intelligence as an emergent behavior, Hillis contends that "The emergent behaviors exhibited by these systems are a consequence of the simple underlying rules defined by the program. Although the systems succeed in producing the desired results, their detailed behaviors are beyond our ability to analyze and predict" ([16], pp. 188–189). Hillis thus can be seen to allude to the observer's reasoning abilities.

As regards our test, the extent of the observer's reasoning abilities does indeed influence the diagnosis of emergence. To render a diagnosis established by a single judge more credible, we could replace one judge with a collective; moreover, membership of such an emergence jury could be restricted to the suitably qualified. On the jury issue, Turing also noted that "A number of interrogators could be used, and statistics compiled to show how often the right identification was given" [32]. Judgments often will come with a statute of limitations—a phenomenon might be reclassified from emergent to non-emergent with the progress of science (as with the artificial-flock example in Section 3). Here again, the analogy with intelligence: Tasks that were once considered intelligent—such as doing sums—nowadays are considered to be but a job for dullards.

Minsky, in his book on the emergence of mind—*The Society of Mind* [21]—nicely illustrates the role of the observer's reasoning abilities. He provides two sets of examples, which he denotes subjective and objective, to which we might refer within our framework as emergent and non-emergent, respectively. Minsky's first set of examples (subjective or emergent) includes such questions as "What makes a drawing more than just its separate lines?" while the second set of examples (objective or non-emergent) includes such questions as "What makes a tower more than separate blocks?" Minsky goes on to explain that the development of the observer's reasoning abilities nullifies the emergence quality where questions of the second type are concerned: "To explain how walls and towers work, we just point out how every block is held in place by its neighbors and by gravity. ... These explanations seem almost self-evident to adults. However, they did not seem so simple when we were children. ... We regard such knowledge as 'obvious' only because we cannot remember how hard it was to learn" ([21], p. 27).

4.6 Surprise

The surprise element also has received attention in a number of works. We noted in Section 1 Arkin's view: "[W]hat occurs in a behavior-based system is often a surprise to the system's designer..." ([1], p. 105). Minsky wrote that:

> We're often told that certain wholes are "more than the sum of their parts." We hear this expressed with reverent words like "holistic" and "gestalt," whose academic tones suggest that they refer to clear and definite ideas. But I suspect the actual function of such terms is to anesthetize a sense of ignorance. We say "gestalt" when things combine to act in ways we can't explain, "holistic" when we're caught off guard by unexpected happenings and realize we understand less than we thought we did ([21], p. 27).

By bringing the observer's *emotion* of surprise into play, our emergence test widens the focal beam of discussion, now shining both on the *system's behavior* as well as on the experimenter and her *internalized expectations*. This relates to Cariani's nutshell description of emergence relative to a model as "the deviation of the behavior of a physical system from an observer's model of it" ([9], p. 779). An author subscribing to said deviation-from-model view would wish to document her a priori expectations before diagnosing emergence and abandoning attempts at explanation. Our emergence test then might be reformulated as *Design (Expectations), Observation, Surprise*.

4.7 Non-Obviousness

A key element of our test is the non-obviousness experienced in the surprise phase by the observer. The study of complex systems is revealing common causes of non-obviousness. Known categories to date include:

1. computational undecidability (e.g., in the Game of Life and cellular automata)
2. self-organizing phenomena
3. sensitivity to initial conditions, known as chaos (e.g., as in weather patterns, and in predator-prey oscillations)

5 Summary

To summarize, the three clauses of our emergent test are grounded in previous work: The design clause expresses our wish to restrict the test to artificially constructed systems, the observation clause reflects the necessity of there being an observer for emergence to arise at all, and the surprise clause embodies both the deliberation and the emotion implied by human judgments of value.

Acknowledgments

We thank Bernard Besnier, Dario Floreano, Daniel Mange, Eduardo Sanchez, Marco Tomassini, and the anonymous reviewers for their helpful remarks and suggestions.

From *The Adventure of the Dancing Men*, by Arthur Conan Doyle:

> He wheeled round upon his stool, with a steaming test-tube in his hand and a gleam of amusement in his deep-set eyes.
> "Now, Watson, confess yourself utterly taken aback," said he.
> "I am."
> "I ought to make you sign a paper to that effect."
> "Why?"
> "Because in five minutes you will say that it is all so absurdly simple."
> "I am sure that I shall say nothing of the kind."
> "You see, my dear Watson"—he propped his test-tube in the rack and began to lecture with the air of a professor addressing his class. . .
> . . .
> "How absurdly simple!" I cried.

References

1. Arkin, R. C. (1998). *Behavior-Based Robotics*. Cambridge, MA: MIT Press.
2. Arkin, R. C. (1998). *Behavior-Based Robotics* (pp. 359–420). Cambridge, MA: MIT Press.

3. Axelrod, R. (1984). *The Evolution of Cooperation*. New York, NY: Basic Books.
4. Berlekamp, E. R., Conway, J. H., & Guy, R. K. (1982). *Winning Ways for Your Mathematical Plays, Volume 2*, (pp. 817–850). New York, NY: Academic Press.
5. Bonabeau, E., Dessalles, J. L., & Grumbach, A. (1995). Characterizing emergent phenomena (1): A critical review. *Revue Internationale de Systémique, 9*, 327–346.
6. Bonabeau, E., Dessalles, J. L., & Grumbach, A. (1995). Characterizing emergent phenomena (2): A conceptual framework. *Revue Internationale de Systémique, 9*, 347–371.
7. Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge, MA: MIT Press.
8. Capcarrère, M. S., Sipper, M., & Tomassini, M. (1996). Two-state, r = 1 cellular automaton that classifies density. *Physical Review Letters, 77*, 4969–4971.
9. Cariani, P. (1992). Emergence and artificial life. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial Life II, Volume X, SFI Studies in the Sciences of Complexity* (pp. 775–797). Redwood City, CA: Addison-Wesley.
10. Chou, H. H., & Reggia, J. A. (1997). Emergence of self-replicating structures in a cellular automata space. *Physica D, 110*, 252–276.
11. Coveney, P., & Highfield, R. (1995). *Frontiers of Complexity: The Search for Order in a Chaotic World*. London: Faber and Faber.
12. Crutchfield, J. P. (1994). Is anything ever new? considering emergence. In G. Cowan, D. Pines, & D. Melzner (Eds.), *Complexity: Metaphors, Models, and Reality* (pp. 515–537). Reading, MA: Addison-Wesley.
13. Emmeche, C. (1994). *The Garden in the Machine: The Emerging Science of Artificial Life*. Princeton, NJ: Princeton University Press.
14. Epstein, J. M., & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brookings Institution Press.
15. Gell-Mann, M. (1994). *The Quark and the Jaguar: Adventures in the Simple and the Complex*. New York, NY: Freeman.
16. Hillis, W. D. (1988). Intelligence as an emergent behavior; or, the songs of eden. *Daedalus, Journal of the American Academy of Arts and Sciences, 117*, 175–189.
17. Holland, J. H. (1998). *Emergence: From Chaos to Order*. Reading, MA: Addison-Wesley.
18. Kolen, J. F., & Pollack, J. B. (1995). The observers' paradox: Apparent computational complexity in physical systems. *Journal of Experimental and Theoretical Artificial Intelligence, 7*, 253–277.
19. Koza, J. R. (1994). Artificial life: Spontaneous emergence of self-replicating and evolutionary self-improving computer programs. In C. G. Langton (Ed.), *Artificial Life III*, volume XVII of *SFI Studies in the Sciences of Complexity* (pp. 225–262). Reading, MA: Addison-Wesley.
20. Mallot, H. (1998). Life is like a game of chess: Review of *Emergence: From Chaos to Order* by John Holland. *Nature, 395*, 342.
21. Minsky, M. (1986). *The Society of Mind*. New York: Simon and Schuster.
22. Packard, N. H. (1988). Adaptation toward the edge of chaos. In J. A. S. Kelso, A. J. Mandell, & M. F. Shlesinger (Eds.), *Dynamic Patterns in Complex Systems* (pp. 293–301). Singapore: World Scientific.
23. Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21*, 25–34.
24. Ronald, E. M. A., Sipper, M., & Capcarrère, M. S. (1999). Testing for emergence in artificial life. In D. Floreano, J.–D. Nicond, and F. Mondada (Eds.), *Proceedings of the Fifth European Conference on Artificial Life (ECAL'99)* (pp. 13–20). Heidelberg: Springer-Verlag.
25. Simon, H. A. (1981). *The Sciences of the Artificial* (2nd ed). Cambridge, MA: MIT Press.

26. Sipper, M. (1997). *Evolution of Parallel Cellular Machines: The Cellular Programming Approach*. Heidelberg: Springer-Verlag.

27. Sipper, M. (1998). Fifty years of research on self-replication: An overview. *Artificial Life, 4*, 237–257.

28. Sipper, M., Capcarrère, M. S., & Ronald, E. (1998). A simple cellular automaton that solves the density and ordering problems. *International Journal of Modern Physics C, 9*, 899–902.

29. Steels, L. (1994). The artificial life roots of artificial intelligence. *Artificial Life, 1*, 75–110.

30. Stewart, I. (1994). The ultimate in anty-particles. *Scientific American, 271*, 104–107.

31. Theraulaz, G., & Bonabeau, E. (1995). Coordination in distributed building. *Science, 269*, 686–688.

32. Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460.

33. von Neumann, J. (1966). *Theory of Self-Reproducing Automata*, ed. A. W. Burks. Champaign, IL: University of Illinois Press.