Cognitive Cellular Automata

Pete Mandik

Department of Philosophy
William Paterson University of New Jersey
300 Pompton Road
Wayne, NJ 07470
mandikp@wpunj.edu

Abstract

In this paper I explore the question of how artificial life might be used to get a handle on philosophical issues concerning the mind-body problem. I focus on questions concerning what the physical precursors were to the earliest evolved versions of intelligent life. I discuss how cellular automata might constitute an experimental platform for the exploration of such issues, since cellular automata offer a unified framework for the modeling of physical, biological, and psychological processes. I discuss what it would take to implement in a cellular automaton the evolutionary emergence of cognition from non-cognitive artificial organisms. I review work on the artificial evolution of minimally cognitive organisms and discuss how such projects might be translated into cellular automata simulations.

1. Introduction

We know that some physical processes suffice to support mental processes. What we don't know is what general conditions result in physical processes giving rise to mental processes. Not all physical processes give rise to mental processes. There were periods in which there were only physical processes and not mental processes. One strong hunch is that life is an intermediary step, that mind emerged only after life did. Questions of cosmic evolution are difficult to settle experimentally since we are hardly in the position to start the whole universe over and see what happens the next time around. Computer simulations allow us to circumvent this limitation, or at least simulate a circumvention. Much work in the field of artificial life is the pursuit of such computer assisted cosmological experimentation.

In 2000, several prominent artificial life researchers published their co-authored list of 14 "open problems in artificial life". Of special interest to the current article is their open problem number 11: "Demonstrate the emergence of intelligence and mind in an artificial living system" (Bedau et al., 2000 p. 365). Not only do the authors pose the problem, but they give what strikes me as excellent advice towards its solution:

> To make progress, one must have a method to detect intelligence and mind when they are present in a system. Consciousness is the most difficult aspect of mind to detect, and initial progress is certain to be somewhere else. A more tractable aspect of mind to detect is meaning, that is, internal states that have semantic or representational significance for the entity and that influence the entity's behavior by means of their semantic content (*ibid.* pp 372-373).

Progress along these recommended lines toward the solution of problem 11 will also involve work of relevance to what they identify as open problem number 10: "Develop a theory of information processing, information flow, and information generation for evolving systems." Among their remarks on information, one in particular strikes me as especially significant:

> Firstly, there appear to be two complementary kinds of information transmission in living systems. One is the conservative hereditary transmission of information through evolutionary time. The other is transmission of information specified in a system's physical environment to components of the system, possibly mediated by the components themselves, with the concomitant possibility of a combination of information processing and transmission. The latter is clearly also linked with the generation of information (to be discussed last). Clarifying the range of possibilities for information transmission, and determining which of those possibilities the biosphere exploits, is a fundamental enquiry of artificial life (*ibid.*, p. 372).

As I read the quoted passage, the first kind of information transmission is that which passes from parent to offspring in virtue of reproduction. This is information transmission that traverses generations. The second kind of information transmission is from the environment to the organism. This is something that can happen over multiple generations as populations adapt and evolve. But the transmission of information from environment to organism can also take place within the lifetime of a single organism and this is especially evident in creatures capable of sensory perception and memory. Both perception and memory are amenable to information-theoretic analyses: perception involves the transmission of a signal across space and memory involves the transmission of a signal across time.

The search for mind will be guided by the search for entities that have states with "semantic or representational significance". The earliest instances of such states will be ones that constitute the pick-up by organisms of information about their environments via sensory components. Slightly more sophisticated instances will involve the retention and processing of that information over time via mechanisms of memory and computation. These forms of information transmission and processing—the ones that constitute the earliest instances of cognition—will emerge in the course of the evolution of organisms that are not themselves in possession of anything cognitive, but may nonetheless be understood in informational terms as follows. The pre-cognitive forbears of cognizers, the non-cognitive "mere" organisms from which cognitive organisms evolve, can be characterized in terms of the transmission of information from parents to offspring via inheritance and the acquisition of novel information at the species level. Non-cognitive or "mere" organisms are not capable of the acquisition of information except by inheritance: novel information is acquired only at the species level over evolutionary time. In contrast, cognitive organisms are the ones capable of the acquisition of novel information in their own lifetime.

These remarks help to suggest a method for addressing open problem number 11: develop a method for evolving artificial organisms in ways such that we (1) are able to detect which of the various kinds of information transmission are present in the system and (2) manipulate factors such as environments and fitness functions to encourage the evolution of the modes of information transmission distinctive of cognitive activity. It is the goal of this paper to explore the feasibility of implementing (1) and (2) in a cellular automaton.

Cellular Automata (CA's) comprise a class of dynamical systems defined as (i) n-dimensional arrays of cells in which (ii) each cell can be in one of k states at a given time t, and (iii) what state some cell C is in at t is a function (which can be either deterministic or pseudo-random) of what states a designated set of other cells (C's neighborhood) were in at time t-1. CA's are typically spatially and temporally discrete: cells are the discrete units of space in CA's and the smallest units of time are called "generations" or "ticks". (However, continuous CA's have been investigated (Wolfram, 2002)).

CA's have a long history in the study of artificial life. One of the earliest uses of CA's was von Neumann's exploration of mechanical self-reproduction and his construction a CA pattern that implemented a universal constructor (von Neumann, 1966). One particularly famous CA is John Horton Conway's Game of Life (Berlekamp et al., 1982). Life is (i) a 2-dimensional array in which (ii) each cell C can be in one of 2 states (ON or OFF), (iii) as a deterministic function (The Rule) of the states of the eight cells diagonally and orthogonally adjacent to C.

> **The Rule:** If at time t C has exactly two neighbors that are ON, C will be in the same state at t+1 as in t. If at time t C has exactly three neighbors that are ON, C will be ON at time t+1. If the number of C's neighbors that are ON at time t is less than two or greater than three, C will be OFF at t+1.
>
> Watching the graphical display of a computer applying The Rule reveals many interesting phenomena:
>
> Shimmering forms pulsate, grow, or flicker out of existence. "Gliders" slide across the screen. Life fields tend to fragment into a "constellation" of scattered geometric forms suggestive of a Miró painting. (Poundstone 1985: 24)

The pulsating and gliding patterns observed in Life are excellent examples of macro-level phenomena that 'emerge' from micro-level interactions. For example, at the lowest level are the cells which change state but do not move. At higher levels are patterns such as gliders which move. Thus is movement an emergent property of the Life universe insofar as it is a property had by higher but not lower levels of organization. Might a cellular automaton such as life be an illuminating basis for the study of how mind emerges from matter in our own universe? Cellular automata are useful tools for studying physical phenomena as well as studying the emergence of higher level properties. Perhaps some of the higher level properties that CA's can support include those considered cognitive. One piece of evidence in favor of this hypothesis is that cellular automata are capable of sustaining universal computation. Conway proved the existence of cellular automata capable of supporting universal computation. (Berlekamp et al., 1982. See also Rendell 2001).

On the assumption of the computational theory of mind, we know then that cellular automata are capable of supporting cognition. What has not been shown yet is the existence of a cellular automaton that, in the course of its evolution, will give rise to cognition. That is, what has not been shown is the implementation of cognition in a cellular automaton without it being built in by hand. One suspicion, based on the way things happened in our universe, is that a cellular automaton will give rise to mind only if it first gives rise to life. While cellular automata have been used to study artificial life, no instance yet has been put forward in which a cellular automaton that supports life also gives rise to cognition. However, artificial life researchers do have somewhat of a handle on the problem of how to start with simulated instances of biological phenomena and wind up via evolution by natural selection with instances of cognitive phenomena. However, the systems that they were working with were not cellular automata. They were not even systems whereby the biological properties were emergent from non-biological physical properties. What I propose is to conjoin the insights of the emergence of life from cellular automata with the insights of the emergence of cognition from artificial life and sketch out a plan for devising a system wherein the emergence of cognition can arise—all the way from the bottom, as it were—with the emergence of life as an intermediary. Such a cellular automaton would start off with merely physical properties which later give rise to biological properties, which in turn give rise to psychological properties. No such cellular automaton has been constructed nor do I have a fully worked out plan of how such a construction would proceed. Many of the main obstacles are that we aren't quite sure what we would be looking for. How could we tell whether mind emerged or not? What would make one cellular automaton a better choice for such a thing than another? What I hope to provide are sketches of materials that would put us closer to such answers.

The sketch of the remainder of the paper is as follows. In section 2 I discuss projects that have involved implementations of computation in cellular automata. I will focus on both how far these projects go toward and how far they fall short of the goal of implementing cognition. Central to my concerns of the insufficiency of computation for cognition is that natural instances of cognition are all embodied and embedded, that is, in the service of the functioning and guidance of a natural organism's body in its environment. I will sketch how these concerns arose during the history of functionalist theories of mind associated with philosophical interest in the notion of Turing machines. In section 3 I flesh out further a characterization of cognition that more closely relates it to biological function. To help spell this out I describe artificial life research that involves evolving simple cognitive systems from non-cognitive replicators. In section 4 I describe attempts to emulate life in cellular automata, especially the life-like properties of self-replication and evolution by natural selection. In section 5 I present speculations as to how the insights from the kinds of work described in sections 3 and 4 might be combined to give rise to the emergence, in a cellular automaton, of cognition from a biological (albeit synthetic) substrate.

2. Computation, cognition, and cellular automata

My main goals in this section are to describe why it's a big deal, but only part of the story, that cellular automata support universal computation. I begin by addressing why it's a big deal. What follows is a very brief and biased history of AI from Turing to connectionism to embedded, embodied, evolutionary approaches most associated with artificial life.

In the history of cellular automata research, three key discoveries involve applications of some notion or other of universality: von Neumann proved the existence of cellular automata capable of supporting universal construction (von Neumann, 1966). Conway proved the existence of cellular automata capable of supporting universal computation (Berlekamp et al., 1982). Wolfram proved the existence of universal cellular automata: cellular automata capable of supporting emulations of any other cellular automaton (Wolfram, 2002). Of particular interest to the current section is Conway's result. It will be useful first to sketch the significance of Turing's notion of computation for the study of cognition.

There's nothing new to the suggestion that people are nothing but machines and further, that a crucial step in understanding how we are machines will involve construing reasoning as the effecting of a mechanical procedure. Such views are at least as old as Hobbes' *Leviathan*. What is crucial about Turing's contribution is not simply the suggestion that cognition may be regarded as a mechanical, but in his suggestion of how such a mechanization can be subjected to a mathematical treatment. This is crucial because, as Rosenthal points out, "Post-Galilean physics demands that all physical properties be described in mathematical terms" (2006, p. 12). Alan Turing's contributions were monumental in fleshing out Hobbes' suggestion in mathematical terms. Turing's 1950 "Computing machinery and intelligence" advanced the proposition that a suitably programmed digital computer could think. However, Turing's crucial link to a mathematization of cognition via computation was his explication of what computation consists in in his 1936 "On Computable Numbers, with an Application to the Entscheidungsproblem". Turing's suggestion constituted a way of thinking of mathematical answers being arrived at in terms of collections of mechanistic steps or effective procedures. Famously, Turing spelled this out in terms of an abstraction referred to as a 'Turing machine'. A Turing machine consists in the machine proper which can be in one of a finite number of states and an infinite tape of cells upon which symbols may be written and from which symbols may be read. The machine goes into various states including reading a symbol, writing a symbol and shifting to the left or right as specified by the machine's look-up table which contains specifications of the transition functions from state to state. Different Turing machines are distinguished by their different look-up tables. Universal Turing machines have look-up tables that allow them to emulate any other Turing machine. Part of what makes such universality possible is the possibility of encoding on a Turing machine's tape a representation of some other Turing machine's look-up table.

While mathematical interest in Turing machines involves the way in which the notion of computing may be defined in terms of the actions of Turing machines, cognitive scientific interest in Turing machines involves the way in which the notion of cognition, or more broadly, the having of mental states, may be defined in mechanistic terms. This has been especially interesting to cognitive scientists and philosophers of mind who advocate the view of the mind known as functionalism. One early philosophical treatment of minds in terms of the kinds of machines Turing hypothesized was due to Putnam (1967) and merits quoting at some length. Putnam proposed

> … the hypothesis that pain, or the state of being in pain, is a functional state of a whole organism....I shall assume the notion of a Probabilistic Automaton has been generalized to allow for "sensory inputs," and "motor outputs"—that is, the Machine Table specifies, for every possible combination of a "state" and a complete set of "sensory inputs," an "instruction" which determines the probability of the next "state," and also the probabilities of the "motor outputs." (This replaces the idea of the Machine as printing on a tape.) I shall also assume that the physical realization of the sense organs responsible for the various inputs, and of the motor organs, is specified, but that the "states" and the "inputs" themselves are, as usual, specified only "implicitly"--i.e., by the set of transition probabilities given by the Machine Table....A Description of S where S is a system, is any true statement to the effect that S possesses distinct states S1, S2 ..., Sn which are related to one another and to the motor outputs and sensory inputs by the transition probabilities given in such-and-such a Machine Table (p. 199).

The question arises as to what machine tables suffice to specify minds, for surely not just any finite state machine suffices to implement cognition. Thus is computation insufficient for cognition since not just any computation will implement cognition. Questions arise as to what constraints on computation are needed to implement cognitions. We get a prominent set of answers to these sorts of questions from so-called analytic functionalists. Of course, analytical functionalists are not without there own problems, but it will be useful to briefly review their place in the history of these discussions. What makes analytical functionalists *functionalists* is their belief that what makes something a mental state is the role that it plays in a complex economy of causal interactions. What makes analytical functionalists *analytical* is their belief that which roles are essential is to be discovered by a consultation of common sense knowledge about mental states. Thus, as Braddon-Mitchell and Jackson (2000) write:

> We distinguish the roles that matter for having a mind, and matter for being in one or another mental state, by drawing on what is common knowledge about mental states. We extract the crucial functional roles from the huge collection of what is pretty much common knowledge about pains, itches, beliefs, desires, intentions, and so on and so forth, focusing on what is most central to our conception of what a pain is, what a belief is, what it is to desire beer and all the rest. And we can group the common knowledge into the three clauses distinctive of the functionalist approach. The input clauses will contain sentences like 'Bodily damage causes pain' and 'Chairs in front of people in daylight cause perceptions as of chairs'; the output clauses will contain sentences like 'Pain causes bodily movement that relieves the pain and minimizes damage' and 'Desire for beer causes behaviour that leads to beer consumption'; the internal clauses will contain sentences like 'Perception as of beer in front of one typically causes belief in beer in front of one' and 'Belief that if *p* then *q* typically causes belief that *q* on learning *p*'.
>
> …

> What then is a given mental state *M*, according to the common sense functionalist story? It is the state that plays the *M* role in the network of interconnections delivered by common knowledge about the mind. The network in effect identifies a role for each mental state, and thus, according to it, a mental state is simply the state that occupies the role definitive of it (pp 45-47).

There are three serious related problems that arise for analytical functionalism. The first problem is that analytical functionalism appears to be committed to the existence of analytical truths and various philosophers inspired by Quine have been skeptical of analytical truths. As Prinz (2006) succinctly sums up this Quinean skepticism, the objection is that "[r]oughly, definitions of analyticity are either circular because they invoke semantic concepts that presuppose analyticity, or they are epistemically implausible, because they presuppose a notion of unrevisability that cannot be reconciled with the holistic nature of confirmation" (p. 92). There are two main ways in which analytic functionalism seems saddled with belief in analytic truths. The first concerns the nature of psychological state types such as beliefs and desires. Analytical functionalism is committed to there being analytic truths concerning the necessary and sufficient conditions for being a belief. The second concerns the meaning of mental representations. The most natural theory of meaning for the analytic functionalist to adopt is that what makes one's belief about, say, cows, have the meaning that it does, is the causal relations it bears to all other belief states. However, it is likely that no two people have all the same beliefs about cows. Thus, on pain of asserting that no one means the same thing when they think about cows, the functionalist cannot allow that every belief one has about cows affects the meaning of one's cow thoughts. In order to allow that people with divergent beliefs about cows can both share the concept of cows, that is, both think about the same things when they think about cows, the analytic functionalist seems forced to draw a distinction between analytic and synthetic beliefs, eg., a distinction between beliefs about cows that are constitutive of cow concepts and beliefs that are not. But if Quinean skepticism about the analytic/synthetic distinction is correct, no such distinction is forthcoming.

The second problem arises from worries about how minds are implemented in brains. Many so-called connectionists may be seen to agree with analytical functionalists that mental states are defined in terms of networks. However, many connectionists may object that when one looks to neural network implementations of cognitive functions, it is not clear that the sets of nodes and relations postulated by common sense psychology will map on to the nodes and relations postulated by a connectionist architecture (see, e.g. Ramsey, et al., 1991). The question arises of whether folk-psychological states will smoothly reduce to brain states or be eliminated in favor of them. (I will not discuss further the third option that folk-psychological states concern a domain autonomous from brainstates.)

A third problem arises from worries about the evolution of cognition. If a mind just is whatever the collection of folk psychological platitudes are true of, then there seem not to be any simple minds, for a so called simple mind would be something that the folk psychological platitudes were only partially true of in the sense that only some proper subset of the platitudes were true of it. However a very plausible proposal for how our minds evolved is from simpler minds. It counts against a theory that it rules out a priori the existence of simpler minds than ours for it leaves utterly mysterious what the evolutionary forebears of our minds were. This third problem is especially pertinent to artificial life researchers and the 11th problem.

One promising solution to these three problems involves appreciating a certain view concerning how information-bearing or representational states are implemented in neural networks and how similarities between states in distinct networks may be measured. Models of neural networks frequently involve three kinds of interconnected neurons: input neurons, output neurons, and neurons intermediate between inputs and outputs sometimes referred to as "interneurons" or "hidden-neurons". These three kinds of neurons comprise three "layers" of a network: the input layer, the hidden layer, and the output layer. Each neuron can be, at any given time, one of several states of activation. The state of activation of a given neuron is determined in part by the states of activations of neurons connected to it. Connections between neurons may have varying weights which determine how much the activation of one neuron can influence the other. Each neuron has a transition function that determines how its state is to depend on the states of its neighbors, for example, the transition function may be a linear function of the weighted sums of the activations of neighboring neurons. Learning in neural networks is typically modeled by procedures for changing the connection weights. States of a network may be modeled by state spaces wherein, for example, each dimension of the space corresponds to the possible values of a single hidden neuron. Each point in that space is specified by an ordered n-tuple or vector. A network's activation vector in response to an input may be regarded as its representation of that input. A state of hidden-unit activation for a three-unit hidden layer is a point in a three-dimensional vector-space. A cluster of points may be regarded as a concept (Churchland 1989).

Laakso and Cottrell (1999, 2000) propose a method whereby representations in distinct networks may be quantified with respect to their similarity. Such a similarity measure may apply even in cases where the networks in question differ with respect to their numbers of hidden units and thus the number of dimensions of their respective vector spaces. In brief the technique involves first assessing the distances between various vectors within a single network and second measuring correlations between relative distances between points in one network and points in another. Points in distinct networks are highly similar if their distinct relative distances are highly correlated.

Regarding the analytic/synthetic distinction related worries, the Laakso and Cottrell technique allows one to bypass attributions of literally identical representations to distinct individuals and make do instead with objective measures of degrees of similarity between the representations of different individuals. Thus if I believe that a cow once ate my bother's hat and you have no such belief, we may nonetheless have measurably similar cow concepts. This is no less true of our psychological concepts such as our concepts of belief and concepts of desire. The so-called common-sense platitudes of folk psychology so important to analytic functionalism may very well diverge from folk to folk and the best we can say is that each person's divergent beliefs about beliefs may be similar. And similarity measures are not restricted to the concepts that constitute various folk theories, we may additionally make meaningful comparisons between various folk theories and various scientific theories. This last maneuver allows us to retain one of the key insights of analytic functionalism mentioned earlier: that we are in need of some kind of answer to the question 'how do you know that your theory is a theory of belief?" The answer will be along the lines of "because what I'm talking about is similar to beliefs."

Regarding the question of simple minds, if there are no analytic truths, then there is no *a priori* basis (if any) for drawing a boundary between the systems that are genuine minds and those that are not. Similarity measurements between simple minds and human minds would form the basis for a (mind-body?) continuum along which to position various natural and artificial instances. How useful for understanding human minds will be the study of systems far away on the continuum? We cannot know *a priori* the answer to such a question.

The crucial aspects of the above discussion concern (i) how analytic functionalism helped to connect the crucial notions from Turing’s work to a philosophical account of the nature of mental states and (ii) the remaining problems left unsolved by analytic functionalism cry out for evolved and embodied solutions.

We are in a position now to more fruitfully return to the main questions of this paper concerning the implementation of cognition in a cellular automaton. We are also in a position, now, to see why, in spite of being a big deal, the demonstration of universal computation in cellular automata is only a proper part of the story of how to implement cognition. The earliest evolved minds probably won't themselves be universal Turing machines or neural networks equivalent to universal Turing machines. While there have been successful implementations of computation in cellular automata, they have shown at best, necessary conditions for minds not sufficient ones being implemented in cellular automata. As I will flesh out in further detail in the next section, the quest for cognitive cellular automata needs to be informed by recognition of the importance of embodiment and evolution.

3. Features of early natural minds and their artificial analogs

Minds arise naturally only after organisms have. While some have attributed mentality to organisms as simple as bacteria and paramecia, the less controversial examples of early cognizers were not only multicellular but in possession of a nervous system. The earliest nervous systems have very little going on between input and output. The evolution of nervous systems has witnessed an increase in the complexity of what goes on in between inputs and outputs. There is, however, a basic commonality between the simplest and the most complex organisms and it has to do with the nature of their sensory motor periphery. In all of these organisms, environmental energies must be encoded into sensory information for later use by central systems. Further, central results must be conveyed to behavioral outputs whereby motor commands are decoded by motor effectors. In both cases—from input to central and from central to output—we have a code that can be accounted for in terms of information.

The mathematical notions of information developed by Claude Shannon in his mathematical theory of communication may be especially useful in detecting the emergence of information-using cognitive systems in a cellular automaton. It will be useful to briefly review some of Shannon's core concepts. The first is that of the self-information or surprisal of a message. If we view a message as one of several possible events then the more improbable the event is, the more self-information (or surprisal) its occurrence has. The related notion of entropy may be defined as the average self-information of messages in an ensemble of possible messages. Entropy may thus be viewed as an amount of uncertainty that one of several messages will occur. Mutual information or transinformation of X relative to Y is a measure of how much information about Y is carried by X. Mutual information between X and Y may be defined as the sum of the entropy of X and the entropy of Y minus the joint entropy of X and Y. The capacity of an information channel can be expressed as the maximum transinformation between channel input and output.

Transinformation is of interest to the current discussion for two reasons. The first is that sensory and motor systems may be regarded as information channels. The second is that Langton (1990) used average transinformation as a measure of complexity in a cellular automaton. Cellular automata with high average transinformation correspond to cellular automata that Wolfram labels as class IV automata (automata that exhibit a high degree of complexity as opposed to either quiescence or noise).

Greater cognitive complexity is due not only to greater amounts of transinformation in input and output channels. More complex nervous systems can be characterized in terms of advancements of memory. For simplest organisms a given stimulus will eventuate in a given response. However, when memory is on the scene, a given stimulus might result in different responses depending on what past stimuli have impacted on memory. Memory too is amenable to an informational analysis, as memory may be regarded as the transmission of a signal through time (Feynman 1989).

Much of the mind can be considered as information flow: perception, intention, memory. But what about thought? A natural suggestion for how to model thought scientifically is as computation. What does computation have to do with information? In some circles "computation" and "information processing" are near synonymous, but this equivalence is not particularly illuminating. However, much of the connection between information and computation can be spelled out when we consider the Turing machine abstraction. There is the memory, which can be explicated as above in terms of information. There is the reading and the writing, both of which are informational. There is the execution of the program, which itself is informational, there is a flow of information from the commands to their execution.

Perhaps a more important connection for the current paper is the connection between information and the evolution of the earliest minds. The evolution of the earliest minds will involve the evolution of the first organisms able to pick up and utilize information. One of the simplest adaptive behaviors that is nonetheless sufficiently complex to be regarded as cognitive is taxis: the movement toward or away from a stimulus. Positive phototaxis and negative phototaxis—the movement toward and away from light, respectively—are perhaps too simple to merit regarding as examples of visually guided action. Nonetheless, the accomplishment of such taxes does involve the transduction and processing of information.

To appreciate the informational demands that taxes pose on organisms, it will be illuminating to consider some examples of how real organisms accomplish positive chemotaxis. Positive chemotaxis is often the way in which organisms are guided toward a nutrient source. Positive chemotaxis involves navigating a chemical gradient and the motion problem a motile creature positioned somewhere in the gradient must solve is how to orient its body in the direction of the greatest concentration. A creature with multiple sensors located on different parts of its body is in a position to solve this problem by comparing the activity in the different sensors. Level of activity in a single sensor may be assumed to correspond to the level of chemical concentration at that location. If so, then greater activity on the right than on the left indicates that orienting up the gradient will be accomplished by turning to the right. However, the demands of chemotaxis are a bit more daunting for creatures that are restricted to a single sensor or, what amounts to effectively the same thing, have multiple sensors insufficiently far apart to register any appreciable differences in the local chemical concentration. This latter situation describes the position that the chemotactic bacterium *E. coli* is in. While it has multiple sensors, the sensors are insufficiently far apart and the gradients navigated too diffuse for there to be any significant difference in the activation of the different sensors. One natural hypothesis for how to solve the problem of single-sensor chemotaxis is to utilize some sort of memory. If a creature moves to a location that results in a different level of activity in the sensor, and a memory record is stored of the previous sensor activity, then the creature is in a position to compare current to past sensor activity and compute whether its motions have been taking it up the gradient. Koshland (1977) has tested the hypothesis that *E. coli* make use of memory.

E. coli move by alternating between "runs" and "tumbles". During runs, their various cilia curl around each other forming an impromptu tail allowing for relatively straight motion. During tumbles, the cilia separate and point out it various directions making the motion of the bacterium much more erratic. When the bacterium is moving from a region of low concentration to high, this increases the likelihood that it will go into run mode and when moving from high to low this increases the likelihood of going into tumble mode.

Koshland tested the hypothesis that the bacteria were utilizing memory to ascertain the change in sensor activity. The researcher reasoned that if the bacterium were not utilizing memory then its reaction to a chemical solution with a uniform concentration would not differ as a function of what the concentration was of the solution it was in previously. Conversely, if memory is employed then a given concentration would result in a different behavior if it was higher than a previous concentration than if it is lower. More specifically, the bacterium if moved from low to high, would be more likely to go into run mode and, if moved from high to low, would be more likely to go into tumble mode. This is indeed what the experimenters found.

Single celled organisms are not the only organisms that perform what is effectively single-sensor chemotaxis. The nematode *C. Elegans* is in a similar position. *C. elegans* have a pair of sensors on their head and another pair on their tail. However, *C. elegans* can perform positive chemotaxis even when the tail sensors are removed. Further, when navigating chemical gradients on the flat surfaces of Petri dishes, the worms lie on their side, so the pair of head sensors is perpendicular to the gradient. And even if they weren't perpendicular to the gradient, the head sensors are too close together for any appreciable difference in activity to be utilized (Feree and Lockery 1999).

How might *C. elegans* navigate their gradients? As with *E. coli*, the suggestion that they are implementing some kind of memory mechanism is a natural one. One possible implementation of memory would involve recurrent neural networks (networks that have feed-back as well as feed-forward connections). Recurrence allows for a kind of memory in a neural network since it allows for neural activation to remain even after the input activation is no longer present. Anatomical studies of the networks that mediate between *C. elegans* sensors and muscles show evidence of recurrence (Dunn et al., 2004). Another neural implementation of memory would involve a delay circuit.

In several publications (Mandik 2003, Mandik 2002, Mandik et al in press) I describe artificial life experiments I have conducted concerning the kinds of positive chemotaxis exhibited by the real world creatures described above. The simulations used the Framsticks 3-D Artificial Life software (Komosinski 2000) to show the evolvability in an artificial life simulation of minimally cognitive creatures. The simulations reviewed below involve the implementations of memory in a neural network: a feed-forward implementation and a recurrent implementation. The feed-forward implementation involves passing a sensory signal through two channels one of which delays the signal. The two channels allow for a comparison of sensory signals concerning concentrations at two different times, the delayed signal constituting a memory of a past concentration. The second, recurrent, implementation takes advantage of the way in which signals in a recurrent network may cycle through the network triggering a series of oscillating activations in the neurons that eventually decay and thus constitute a form of short-term memory.

In Mandik (2003) the artificial life simulations of the delay-based memory implementations was explored. A population of creatures was created and evolution was allowed to act upon certain aspects of the creatures' nervous systems. More specifically, the creatures' bodies and neural topologies were set by hand with initial weights of neural connections also set. Evolution was allowed to work only on neural weights by allowing mutations to occur only in the neural weights. In the experiments involving memory described in Mandik (2003), a population of walking creatures was designed and then subjected to evolutionary pressures to see if the creatures could evolve to find food by positive chemotaxis. The creatures' environments had various food sources scattered about and each creature was born with a finite initial store of energy which, if not replenished with food, would reach zero over time and thus kill the creature. Creatures had a single sensor and a steering muscle which, when flexed, could alternately steer the creature to the left or right. One group of creatures had a single feed forward connection leading from the sensor to the steering muscle. Another group had an additional route from the sensor to the steering muscle which involved passing the sensory signal through a chain of neurons thus delaying the arrival of the signal at the steering muscle. Both groups were subjected to evolution by Darwinian selection whereby fitness was defined in terms of total horizontal distance traversed in the lifetime of the creature. Such a fitness function was picked to encourage the evolution of creatures that not only move, but extend their life spans by moving toward food. Populations in the two groups initially were equally poor at finding food, but after 200 million steps of running the program, the creatures with memory buffers showed a fitness significantly superior to the creatures without.

Mandik et al. (forthcoming) explored similar artificial life simulations but with recurrent instead of feed-forward implementations of memory. In these experiments creatures in three different groups were compared. In the first group, creatures had only feed-forward connections from a single sensor to steering muscles. In the second group there were both feed-forward and feed back connections thus implementing recurrence. In the third group the creatures had no sensors. As in the previous simulations, creatures' morphologies and neural topologies were set by hand and mutations were allowed to only the connection weights. Fitness was once again defined as lifetime horizontal distance. The three groups initially had similarly poor performance but over time (240 million steps) one group—the recurrence group—vastly out performed the other groups. Especially surprising was the finding that the single-sensor feed-forward group fared as poorly as the creatures with no sensors as all. Having a single sensor without a memory mechanism conferred no advantages over being blind. As in the previous experiment, having information about past as well as present local concentration allows the artificial life-forms to solve the problem posed by single-sensor chemotaxis.

It is worth discussing how far the above simulations go in answering the open question # 11 concerning the demonstration of "the emergence of intelligence and mind in an artificial living system": First off, the initial populations in the Framsticks simulations were not in the possession of either perceptual or mnemonic sensory states. They were, however, both initially and throughout the duration of the simulation, capable of evolution by natural selection and thus arguably satisfied conditions for being a living system. Through evolution some of the creatures acquired the ability to use not only information about the current chemical concentration but also the past. The utilization of this information constitutes the computation of a food-increasing heading for the creature. What merits the ascription of such mental states to the creature? The answer is the similarity of the theory that explains their behavior to the theory that explains the behavior of humans and non-humans solving a similar problem (Mandik et al., forthcoming).

As already mentioned, we can contrast life and mind in terms of information transmission. The virtue of the informational view for the current project is that it paves the way for the automatic detection of minds in a cellular automaton. If these insights are to guide the construction of cognitive cellular automata, it will be helpful to look at work that has implemented life in cellular automata.

4. Life in cellular automata

This is neither the time nor place to attempt to settle controversies concerning what all and only living organisms have in common. This issue can be sidestepped somewhat by asking not what properties are criterial of living organisms, but what properties were crucially had by the first instances of life. To ask these sorts of questions is not necessarily to ask easy questions, but it does focus things in a manner useful to the current discussion. While the question of the origin of life from non-living matter is vexing indeed, there has been work done on the matter. Theories of the origin of life from non-living matter—abiogenesis—fall roughly into the following groups: reproduction first, metabolism first, and both reproduction and metabolism at the same time. According to Dyson (1999), much attention has been paid to the reproduction-first approach because reproduction is so much easier to define than metabolism. This sort of bias is no less present among researchers in artificial life, and many artificial life projects simulate self-reproduction without bothering to simulate metabolism. In the current paper I will side with the reproduction group and while this may ultimately not be the correct theory of abiogenesis (who's in a position to know at this stage of the game?) it will be instructive to see how far the project of cognitive cellular automata can go if the initial assumptions side with the reproduction-first view. It will suffice, then, for the purposes of this paper, to characterize as living any systems capable of self-reproduction. Especially interesting examples of self-reproducing systems will be those with just a right amount of slop in the system to give rise to evolution by natural selection, since Darwinian evolution occurs wherever the mechanisms of the *variable* inheritance of fitness are in place.

The earliest work concerning simulations of life in cellular automata tackled life's capacity for self-reproduction. There was a time when self-reproduction seemed especially difficult for any purely mechanical system to achieve and it was posited that possession of *élan vital* was necessary to pull off such a fancy trick. As mentioned previously, von Neumann (1966) constructed a CA that demonstrated the capacity for mechanical self-reproduction. Von Neumann constructed a cellular automaton and a pattern within it that constituted a universal constructor. It could construct any pattern possible in that CA space, including itself. The resulting constructor was quite complex and subsequent cellular automata researchers sought implementations of self-reproduction that were neither so complex nor universal constructors. One early success along these lines was the work of Langton (1984). Langton's self-reproducing patterns were implemented in a cellular automaton wherein each cell had eight possible states and each cell had five neighbors. The self-reproducing pattern looked like a rectilinear lower-case "q", the tail of which would lengthen over time. When the length of the growing tail advanced in an increment equal to the width of the body—the square loop—the tail's growth would proceed along a right angled bend until another such increment of length was achieved. After three such bends and four new sides were created, the result is a loop similar to the body of the original pattern. The new loop then grows its own tail and the tail of the parent loop dissolves. The parent loop, after birthing its sole offspring, goes into a quiescent state and remains as subsequent generations spread into the cellular automaton space with one offspring each. In each of these patterns, the tail and the sides of the loop are a three-layer structure consisting of two outer sheath layers and one inner core layer. The cells in the sheath layer are in a single state whereas cells in the core layer change and different states "move" around the loop and down the tail. Different circulating core states have different functions, of particular interest are two different gene states: one for straight tail growth and one for right-angled turning. The gene signal states circulate around the core of the loop and down the length of the tail. When a gene signal arrives at the end of the tail, the gene signal triggers an increment of tail growth along either the straight path or at a right angle. What is of special note concerning the gene signals is how they constitute a certain kind of signaling or information flow within the organism. However this flow of information is only from the inner to the outer and from parent to offspring. Individuals do not have the capacity to pick up information from the environment. Nor does the population as a whole have the capacity to pick up information, since while Langton's loops implement reproduction, they do not constitute an implementation of evolution. However, the researcher Sayama modified Langton's loops to create patterns capable of evolution by natural selection.

The first innovation Sayama introduced was to create a state designated as a "dissolving state" that could travel along the structure of a pattern and dissolve any adjacent structures. The dissolving state allowed for a kind of death. The second innovation Sayama introduced was to modify the reproductive mechanism so that not only one kind of pattern was a viable self-reproducer. Thus different sized patterns could compete for space. These two innovations allowed for a kind of natural selection, but this did not constitute evolution because the variation did not arise spontaneously. Sayama's third innovation was to modify the cellular automaton further so as to allow for the appearance of variation over the natural course of the running of the simulation. With these three modifications to Langton's loop implemented, evolution can be demonstrated in a cellular automaton space that starts off with a single loop. As Sayama describes such a simulation:

> At first an ancestral loop is set alone in the center of the space. When simulation begins, the ancestral loop soon proliferates to all the space. Then, self-reproduction and structural dissolution of loops begin to happen frequently in the space, which produce various kinds of variants such as sterile loops, loops with two arms, loops not self-reproducing but reproducing smaller offsprings than themselves, and so forth. A self-reproducing loop of smaller species also emerges by accident from this melee, and once it appears, it is naturally selected to proliferate in the space, due to its quicker self-reproductive ability. Such an evolutionary process develops in the space as time proceeds, and eventually, the whole space becomes filled with loops of species 4, which is the strongest species in this world.

I turn now to explore the question of how far such a cellular automaton takes us in the quest for a cognitive cellular automaton.

5. Evolving mind from life

The evolving artificial creatures in simulations such as Sayama's exhibit no adaptability in a single life time. Contrast these against simple motile organisms that must change something about themselves, namely their position, in order to go to where there is food and/or to where there are no predators. One of the most simple of such adaptive behaviors is positive chemotaxis. This sort of adaptability demands the pick-up of information by an organism in its lifetime.

Langton's and Sayama's loops both involve a kind of representation or information processing but not the kind that constitutes the pick-up of information by an individual in its lifetime. In both Langton's and Sayama's loops, genes code for aspects of loop growth and reproduction but the flow of information is from the inside out, whereas pick-up of environmental information is from the outside in. Sayama's evoloops potentially involve pickup of environmental information, but only at the species level over evolutionary time, not at the level of an individual organism.

The question naturally arises as to what modifications to Sayama's loops would suffice to implement sensory and/or mnemonic aspects of mentality. In order for Sayama's loops to exhibit the above-mentioned kind of adaptability, there would have to be some sort of point to it for the organism. Perhaps there would be items in the environment that were required for reproduction or needed avoiding on pain of death. Sayama himself provides some possible mechanisms for encouraging adaptation, although he is not explicitly concerned with the questions of cognition. In subsequent work with the evoloops, Sayama (2004) explores modifications of the evoloops that would constitute mechanisms of self-protection of a loop from other loops. Sayama explores three potential strategies:

> **Shielded:** the attacked loop generates a dissolving state …at the tip of the attacker's arm. Since the dissolving state usually becomes canceled by the gene flow in the arm, most likely the attacker suffers from only a partial dissolution of the tip of its arm, and the same attack occurs repeatedly.
> **Deflecting:** The attacked loop generates an umbilical cord dissolver … at the tip of the attacker's arm. The umbilical cord dissolver goes back to the attacker, removing the entire arm, and then deceives the attacker into believing that self-replication has been completed. The attacker then starts another attempt of self-replication in a different direction rotated by 90 degrees counterclockwise.

> **Poisoning:** The attacked loop generates a poison … at the tip of the attacker's arm. The poison works as a kind of dissolving state with extra strength that will never be canceled until it deletes the whole contiguous structure. This may be one of the most radical and merciless mechanisms of self-protection (p.85).

Two remarks are especially in order. The first concerns Sayama's attribution of beliefs to the deflecting loops. The second concerns how all three strategies employ an attack detector state. Regarding the belief attribution it is especially pertinent to the current paper whether it is in fact true since if it is, then Sayama has thereby produced a cognitive cellular automaton. The belief in question is the belief that "self-replication has been completed". This is allegedly a false belief had by an attacker as the result of being tricked by a loop employing the deflecting strategy of self-protection. If an organism is capable of having a belief that "self-replication has been completed" then it makes sense to ask what kind of belief it is. Is it a perceptual belief? An introspective belief? A volitional belief? I take it that the most plausible candidate is perceptual belief. If the loop has a belief at all, then it has a perceptual belief. However, the having of a perceptual belief has certain necessary conditions that the loop fails to satisfy. In particular, a necessary condition on having my the perceptual belief that P--that is, a perceptual belief concerning some state of affairs, P--is that I have a state S that is at one end of an information channel which has at the other end P. Further, S must carry the information that P and be caused by P. Thus if I am to have the perception that there is a fly on the far side of the room, then I must have a state that carries the information that there is a fly. Lacking the ability to have such a state I might come to believe that there's a fly, but that belief certainly cannot be a perceptual belief. In other words, perceivers of flies must be capable of detecting flies. Failing an ability to detect flies, one fails to perceive them and likewise fails to have perceptual beliefs about them. Do Sayama's loops have any capacity to detect the termination of their self-reproductive procedures? It seems not, since they have no detector states that carry the information that self-replication has terminated. They thus fail to satisfy a crucial condition for the having of perceptual belief. And on the assumption that perceptual beliefs were the only plausible candidates, then we can conclude that insofar as Sayama's attribution was literal, it is literally false. However, just because Sayama's loops do not have detector states for replication termination, they are not devoid of detector states altogether. As previously mentioned, they have attack detecting states. The question arises as to how far the attack detection schemes in Sayama's loops go toward the evolution of cognition. One thing to note is that the self-defensive strategies triggered by the attack detection state, as well as the attack detection state itself, were designed by Sayama and are not products of evolution in the loops.

It would be interesting if modifications to Sayama's rules could result in the evolutionary emergence of self-protection. One potentially crucial difference between the detection mechanisms in Sayama's loops and those of the Framsticks organisms described earlier is that the latter have states that vary along an ordering relation (lower to higher neural activation) which is causally correlated with environmental magnitudes that themselves vary along an ordering relation (e.g. higher-lower chemical concentration). Because of the preexisting causally correlated isomorphisms between neural activations and chemical concentration, when an organism acquires the capacity to represent any concentration it simultaneously acquires the capacity to acquire representations of many concentrations. Thus the existence of causally correlated isomorphisms between various environmental states and possible physiological states allows for the evolutionary acquisition of many representations via only small genotype changes.

If an evolutionary cellular automaton is to implement the emergence of cognition along the sorts of lines suggested above, then future explorations would do well to include implementations of sufficient environmental complexity to allow for the possibility of causal correlations between ordered states of evolving organisms and ordered states of their environments. Such inclusions are likely to provide a base from which minimally cognitive creatures can emerge and constitute artificial individuals capable of the acquisition of novel information in their individual lifetimes.

Acknowledgements

For extremely helpful comments on an early version of this paper, I thank Pierre Poirier and Benoit Hardy-Vallée.

References.

Bedau, M., McCaskill, J., Packard, N., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., Ray, T. (2000). Open problems in artificial life. *Artificial Life*, 6: 363-376.

Berlekamp, E., Conway, J., Guy, R. 1982. *Winning Ways for Your Mathematical Plays, Volume 1: Games in General*. New York: Academic Press, , N.Y., 1982.

Braddon-Mitchell, D., Jackson, F. 2000. *Philosophy of Mind and Cognition.* Oxford: Blackwell.

Churchland, P. M. 1989. A Neurocomputational Perspective: The Nature of Mind and the Structure of Science. Cambridge, MA: MIT Press.

Dunn, N., Lockery, S., Pierce-Shimomura, J., Conery, J. (2004) A neural network model of chemotaxis predicts functions of synaptic connections in the nematode Caenorhabditis elegans. *Journal of Computational Neuroscience*, 17(2):137-147.

Dyson, F. (1999). *Origins of Life*. Cambridge: Cambridge University Press.

Feree, T., Lockery, S. (1999). Computational rules for chemotaxis in the nematode C. Elegans. *Journal of Computational Neuroscience*, 6, 263-277

Feynman, R. 1996. Feynman Lectures on Computation. Reading, MA: Addison-Wesley.

Koshland, D. (1977). A response regulator model in a simple sensory system, *Science*, 196, 1055–1063.

Laakso, A., Cottrell, G. (1998). How can I know what You think?: Assessing representational similarity in neural systems. In *Proceedings of the Twentieth Annual Cognitive Science Conference*. Mahwah, NJ: Lawrence Erlbaum.

Laakso, A., Cottrell, G. (2000) Content and cluster analysis: Assessing representational similarity in neural systems. *Philosophical Psychology*, 13(1):47-76.

Langton, C. (1984). Self-reproduction in cellular automata. *Physica D*, 10, 135–144.

Langton, C. (1990). Computation at the edge of chaos: phase transitions and emergent computation. *Physica D,* 42, 12-37.

Mandik, P. (2003). Varieties of representation in evolved and embodied neural networks. *Biology and Philosophy*, 18 (1): 95-130.

Mandik, P. (2002). Synthetic neuroethology. *Metaphilosophy*, 33 (1-2): 11- 29.

Mandik, P., Collins, M., Vereschagin, A. (in press). Evolving artificial minds and brains. In Andrea Schalley and Drew Khlentzos (eds.) *Mental States: Nature, Function, Evolution*. Amsterdam: John Benjamins Publishers.

Prinz, J. (2006). Empiricism and state-space semantics. In B. Keeley (ed.) *Paul M. Churchland: Contemporary Philosophy in Focus*. Cambridge: Cambridge University Press.

Putnam (1967). The nature of mental states. In D.Rosenthal (ed.) *The Nature of Mind*, pp. 197-203.

Ramsey, W., Stich S., Garon, J. (1991). Connectionism, eliminativism, and the future of folk psychology, in: W. Ramsey, S. Stich & D. Rumelhart (eds.) *Philosophy and Connectionist Theory*. Hillsdale NJ: Lawrence Erlbaum.

Rendell, P. (2002). A Turing machine in Conway's game Life. www.cs.ualberta.ca/~bulitko/F02/papers/tm_words.pdf

Rosenthal, D. 2006. *Consciousness and Mind*. Oxford: Oxford University Press.

Sayama, H. 1999. A new structurally dissolvable self-reproducing loop evolving in a simple cellular automata space. *Artificial Life*, 5 (4): 343-365.

Sayama, H. 2004. Self-protection and diversity in self-replicating cellular automata. *Artificial Life,* 10: 83-98.

Turing, A. (1950). Computing machinery and intelligence. *Mind* 49, 433-460.

von Neumann, J. 1966. *Theory of Self-Reproducing Automata*. Urbana, IL: University of Illinois Press.

Wolfram, S. 1984. Universality and complexity in cellular automata, *Physica D*, 10:1-35

Wolfram, S. 2002. *A New Kind of Science*. Champaign, IL: Wolfram Media.