

## SYNTHETIC NEUROETHOLOGY

PETE MANDIK

---

**ABSTRACT:** Computation and philosophy intersect three times in this essay. Computation is considered as an object, as a method, and as a model used in a certain line of philosophical inquiry concerning the relation of mind to matter. As object, the question considered is whether computation and related notions of mental representation constitute the best ways to conceive of how physical systems give rise to mental properties. As method and model, the computational techniques of artificial life and embodied evolutionary connectionism are used to conduct prosthetically enhanced thought experiments concerning the evolvability of mental representations. Central to this essay is a discussion of the computer simulation and evolution of three-dimensional synthetic animals with neural network controllers. The minimally cognitive behavior of finding food by exhibiting positive chemotaxis is simulated with swimming and walking creatures. These simulations form the basis of a discussion of the evolutionary and neurocomputational bases of the incremental emergence of more complex forms of cognition. Other related work has been used to attack computational and representational theories of cognition. In contrast, I argue that the proper understanding of the evolutionary emergence of minimally cognitive behaviors is computational and representational through and through.

Keywords: philosophy, artificial life, mental representation, cognitive science, neural networks, evolution, computer simulation.

---

### **Introduction: Diachronic Metaphysics and Synthetic Methodologies**

One of the core questions of philosophy, especially philosophy of mind, is the question of representation. What is the relation of the mind to the world such that the mind comes to have representations of the world? Asked in a materialistic vein the question is one of how brains, or physical systems more generally, have representations of the world. The neurophilosophical presumption of this essay is that brains are the relevant physical systems in question.

The question may be asked in both of two versions, one synchronic and the other diachronic. Synchronically speaking, we have reason to believe that brains are sufficiently structured and situated in environments, right now, to traffic in representations of aspects of their environments and beyond. What patterns of structure and activity in the physical universe

support, synchronically, the representation of objects, properties, and states of affairs?

Diachronically speaking, we have reason to believe that the universe did not always contain representations. What had to happen, over time, for physical structures to come to bear representational properties? Presumably representations postdate the emergence of organisms but predate the emergence of humans. Thus, the temporal course of the existence of representations in the universe is nested in the temporal course of the existence of biological organisms in the universe. Given this biological contextualization of the problem, a natural supposition is that an adequate account of this temporal course will be heavily imbued with remarks on the contributions of evolution by natural selection – the variable inheritance of fitness – to the emergence of representational structures.

The question of representation is foundational to the cognitive scientific enterprise. Further, certain cognitive scientific methodologies supply tools for answering both the synchronic and diachronic aspects of the question. The methodologies I have in mind are heavily constructivist in the sense of Daniel Dennett's view of cognitive science as reverse engineering (Dennett 1998, 249). The advantage accrued by such an approach helps one cope with what neuroscientist Valentino Braitenberg calls the law of uphill analysis and downhill synthesis: it is more difficult to figure out how Mother Nature contrived to design an organism to accomplish some task than it is to come up with one's own artificial solution (Braitenberg 1984). Once, however, a well-functioning artifact has been created to produce the target phenomena – once the synthesis has been effected – the researcher is often in a better position to tackle the analysis.

The cognitive scientific methodologies of artificial intelligence and computational neuroscience supply the means for a synthetic approach to the synchronic aspects of the problem of representation. The techniques of connectionist modeling and the construction of neural network controllers for autonomous robots allow us to test synthetic hypotheses about the possible neural architectures that will support intelligent behavior. The diachronic questions, however, remain relatively untouched by such techniques. How might the various proposed neural architectures have evolved from other systems? One supposition often made in evolutionary contexts is that each relatively adaptive solution must have as a precursor an incrementally distinct but nonetheless relatively adaptive solution. Thus, for instance, the evolution of eyes with lenses had as incrementally adaptive precursors "pinhole" eyes without lenses, light-sensitive invaginations, and light-sensitive skin patches, respectively (Llinas 2001, 101). For the synthetic approach to the diachronic aspects of cognitive questions we may turn to the techniques of artificial life.

Artificial life involves the application of biological solutions to computational problems and the application of computational solutions to biolog-

ical problems (Liekens 2001). Typical artificial-life projects involve the computer-simulated evolution of populations of synthetic creatures. Artificial-life approaches to cognitive scientific problems have several features that distinguish them from other synthetic techniques, such as artificial intelligence (including both GOFAI – Good Old Fashioned AI – and connectionist modeling). First, such artificial-life projects involve the modeling of entire organisms. Natural examples of cognitive creatures are extremely complex, and modeling must necessarily simplify. Whereas GOFAI and connectionist approaches typically simplify by focusing on subsystems of agents (by, for instance, creating a program that can convert text to speech), artificial-life approaches focus on relatively simple creatures. Thus, such approaches echo Dennett's "why not the whole iguana?" approach (Dennett 1998, 309).

A second mark of contrast between artificial-intelligence and artificial-life approaches to cognition is the contrast between the relative reliance on designed versus evolved solutions. Artificial life employs the creation of synthetic cognitive systems via evolutionary algorithms. The variable inheritance of fitness may be defined over a finite set of combinatorial elements. Fitness functions may be specified for the evaluation of combinations. Combinations are copied with varying degrees of fidelity, allowing for both the inheritance of fitness and the introduction of mutations into the gene pool. While the specification of the elements and the fitness functions are up to the designer, the evolutionary products are not. In fact, they are often quite surprising to the designer.

Artificial-life approaches to cognition to date have focused on the design and evolution of "minimally cognitive behavior," like obstacle avoidance and food finding (Beer 1990). Such approaches allow for both the evolution of cognitive systems from noncognitive systems and the evolution of complex cognitive systems from comparatively simple cognitive systems. Such projects have involved the evolution of controllers for robots, the evolution of controllers for simulated morphologies, and the co-evolution of both controllers and morphologies.

### **When Animats Attack: The Revolt against Representation**

Animats are synthetic animals, either computer simulated or robotic. Many prominent animat researchers describe their results as posing challenges to the cognitive scientific assumptions that intelligent behavior requires mental representation and computation. Two prominent representatives of this line of attack are Randall Beer and Rodney Brooks. Beer's work concerns computer simulations of insects with neural network controllers capable of guiding them through environments and finding food (Beer 1990). Brooks's early work concerns six-legged mobile robots with control structures that implement a subsumption architecture: a collection of systems each capable of guiding behavior, some of which are able to

modulate (subsume) the activity of others (Brooks 1991). Brooks sees the subsumption architecture as avoiding a bottleneck that is introduced by more hierarchical control systems that employ a central control unit that uses representations. According to Brooks, “Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems” and “explicit representations and models of the world simply get in the way” (Brooks 1991, 140). Similarly, Beer states of his computer-simulated insect,

There is no standard sense of the notion of representation by which the artificial insect’s nervous system can be said to represent many of the regularities that an external observer’s intentional characterization attributes to it. Even the notion of distributed representation which is currently popular in connectionist networks does not really apply here, because it still suggests the existence of an internal representation. . . . The design of the artificial insect’s nervous system is simply such that it generally synthesizes behavior that is appropriate to the insect’s circumstances. (Beer 1990, 162–63).

Much animat research, thus, construes the animats as *reactive agents*. Reactive agents are able to exhibit a surprising variety of behaviors in spite of their alleged lack of internal representations of their environments. For simple schematic illustrations of the basic principles of reactive agents, consider some of Valentino Braitenberg’s thought experiments from his influential book *Vehicles: Experiments in Synthetic Psychology* (Braitenberg 1984). Figure 1 depicts a bird’s-eye view of three of Braitenberg’s simplest animats (vehicles). A stimulus source in the form of a light is in the upper left-hand corner of the figure. The vehicle on the left has a single sensor with a single excitatory connection to a single motor. Increased sensor activity results in increased motor activity; thus,

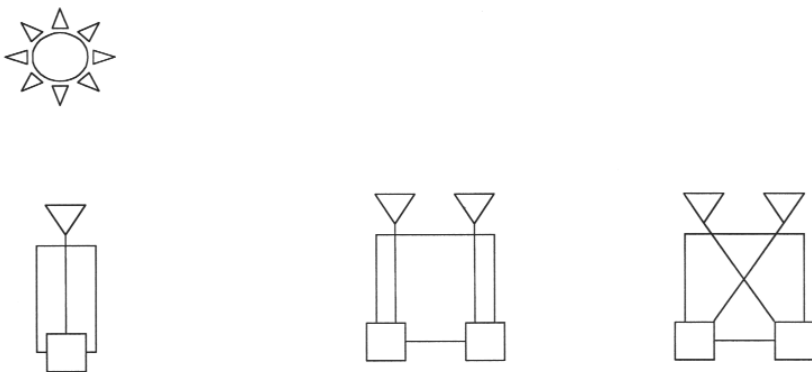


Figure 1. Three Braitenberg vehicles and a stimulus source.

increased proximity to the stimulus results in higher velocities of the vehicle. The vehicles in the middle and on the right have slightly more complex architectures.

The middle vehicle, with excitatory connections wired in parallel, will turn away from a stimulus because the motor closer to the stimulus will turn faster. If instead this vehicle had inhibitory connections it would move toward the stimulus. The vehicle on the right, with crossed excitatory connections will move toward a stimulus, and with crossed inhibitory connections will move away. Braitenberg's vehicles illustrate how relatively simple architectures can form the basis of coherent, survival-enhancing behavior. By multiplying the number of sensors (for example, light, chemical, temperature, obstacle proximity, and so on) and kinds of connections (excitatory, inhibitory, parallel, and crossed), a single vehicle could be capable of finding nutrients while avoiding toxins and obstacles. These vehicles are reactive agents insofar as their behaviors are driven by reactions to environmental stimuli. In spite of their simplicity, reactive agents are capable of exhibiting minimally cognitive behavior. Creating such agents is the focus of much animat research. This involves situating an animat in an environment in which it must perform tasks that seem (to the researcher) to be conducive to survival. Thus, typical behaviors to model include avoiding obstacles and finding "food."

One of the rallying cries of antirepresentational approaches to cognition is that in many, if not all, instances of intelligent behavior, the environment is sufficiently rich that the agent need not represent it, just react to it. Some, such as Brooks, even go so far as to describe the world as its own model (Brooks 1991). Such a view is not limited to those working on animats and includes those working on humans and other natural organisms. Thus, for instance, O'Regan and Noë, in describing the function of human visual perception, describe the environment as being its own representation (O'Regan and Noë 2001). This echoes one of the main themes of Ecological Psychology developed by J. J. Gibson and his followers: the contention that sufficient amounts of information are simply in the environment, just waiting to be picked up by the behaving organism, and thus need not be represented or computed (Gibson 1966).

If there is an argument against positing representations implicit in these sorts of remarks, it seems to be the following two-stage argument. The first stage involves the thesis that representation is required only for the guidance of behaviors concerning things that are somewhat spatially and temporally remote. Remembering what happened last night and a thousand miles away seems to involve some kind of representation or record of these past events. Similarly, plotting a course for the Eiffel tower while one is in Spain cannot be stimulus-driven by the tower itself. Call this the *remoteness requirement* of representation. Haugeland illustrates a similar point by comparing the sun-tracking abilities of a sunflower to an imaginary "supersunflower" (Haugeland 1991). A

sunflower is able to track the position of the sun in the sky, but only if the sun is not blocked by a cloud or a building. An imaginary *super* sunflower, in contrast, is able to track the position of the sky by representing its trajectory and can do so even when the sun ducks behind an obstruction.

The second stage of the antirepresentationalist argument involves construing perceptually guided (or stimulus-driven) action as not satisfying the remoteness requirement. Consider, for instance, the task of reaching for and grasping a coffee cup that you see before you: why bother to represent the cup when it is just there for the taking? Like the movement of the sunflower, the motion of the hand may be driven by what is present in the environment. Similarly, obstacle avoidance and food finding by positive chemotaxis and phototaxis exhibited by simple animals and animats is stimulus-driven and thus fails the remoteness requirement for representation.

Where there is no representation, there is no computation either. Computation involves transformations that map semantically evaluable inputs onto semantically evaluable outputs. For example, a device that computes a vehicle's velocity from the rate of revolution of its wheels can do so only if it has some states that are semantically evaluable as representing velocity and other states that are semantically evaluable as representing rate of revolution. Thus, showing that an organism or animat lacks states that satisfy the remoteness requirement suffices to show both that the organism lacks representations and that it lacks computational processes.

### **The Metaphysics of the Neuron**

Views that go hand in hand with the above thinking, therefore, depict representation as an evolutionary Johnny-come-lately. Such views deny that the appearance of organisms with nervous systems that allow them to move toward nutrients and away from toxins coincides with the appearance of representation and computation. I think that this view is mistaken. Even so-called stimulus-driven behavior satisfies the remoteness constraint. Failure to see this is due to a failure to appreciate the relatively enormous distances (both literal and metaphorical) between sensory inputs and motor outputs. It may be appropriate to describe single-celled organisms as simply reacting without representing, but well before the emergence of vertebrates a significant distance opened between stimulus and response. In animals – complex motile organisms – parts on opposite sides of a multicellular organism need to know what the others are doing. The distances that have to be traversed may be measured in the vast number of cell membranes that must either be crossed or circumnavigated to match appropriate responses to stimuli. Nervous systems are the evolutionary solution to this complex coordination problem. Further, the solution is representational and computational through and through. Transducer neurons pick up environmental information and encode it in representa-

tional formats that may be processed by central systems. Among the products of central processing are motor representations that are decoded by motor systems and eventuate in appropriate muscular activity. These representation-manipulating processes are, by definition, computations.

Returning to the questions raised under the heading of diachronic metaphysics, questions arise of how evolution and the incremental emergence of function give rise to the appearance of the first instances of biological representation and computation. Llinas describes the early evolution of nervous systems as follows. In relatively primitive multicellular organisms like sponges, contractile cells respond to direct stimulation. In more evolved organisms like sea anemones, sensory and contractile functions are handled by separate cells, and on some occasions sensory neurons are connected directly to motor cells without any intervening interneurons (Llinas 2001, 11). The evolution of more complex animals is accompanied by the evolution of nervous systems with more interneurons. Thus, the evolution of nervous systems involves an increase in nervous tissue intervening between stimulus and response. This involves an increase in the number of membranes that must be traversed or circumnavigated to relay a signal from one end of the organism to the other, increasing the satisfaction of the remoteness criterion for representation. The evolutionary growth of nervous systems, however, does not involve lengthening straight-line paths from stimulus to response but instead involves complex branching structures. According to Llinas,

The great advantage provided by such often widely branching interneurons is the ability to “steer with multiple reins.” The sensory stimuli activating a few sensory cells may activate a small set of interneurons, which may in turn and, through many spinal segments of connectivity, evoke a complex motor response involving a large number of contractile elements. Through this profusely branching forward connectivity, the animal becomes capable of performing well-defined gross movements that involve many muscles along its body. (Llinas 2001, 81)

I would add that another advantage of the emergence of branching networks of interneurons is that it marks the emergence of processes that not only *relay* information but also *process* information. Thus, the appearance of these networks constitutes the appearance of computation. I explore this theme in further detail in later sections.

The emergence of memory introduces further remoteness (temporal remoteness) and thus clearer satisfaction of the remoteness constraint. Let us define as memory any persistence of encoded information in nervous systems. One plausible neural mechanism for at least short-term memory is recurrence. Figure 2 depicts a simple network with a recurrent connection from the third neuron back to the second neuron. (Longer-term memory may be accomplished by mechanisms that alter the connectivity and/or the connection weights of the network as a function of experience,

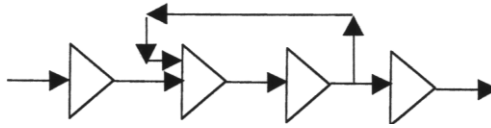


Figure 2. A simple recurrent neural network.

as in the proposed mechanisms of Hebbian learning or Long Term Potentiation.)

Recurrence, which may serve to implement certain forms of memory, also potentially serves another purpose: the generation of repetitive motion. Most natural forms of locomotion involve the repetitive motion of some limb. Wings are flapped in flying, legs step in walking, and a tail is flagellated in swimming. Further, the muscles driving these repetitive motions need themselves to be driven by repetitive neural signals. One possible source of such signals is a central pattern generator. In the artificial creatures described in this essay, central pattern generators involve recursively connected collections of neurons that sustain oscillatory activity. It is possible that the recurrent connections evolved initially to serve as pattern generators for repetitive motions and can be adapted to serve as systems for short-term memory.

### The Simulated Evolution of Creatures and Their Neural Networks

The simulations described in the rest of this essay employ the Framsticks artificial-life simulator created by Maciej Komosinski and Szymon Ulatowski (Komosinski 2000; Komosinski 2001). This piece of software allows for the simulation of three-dimensional creatures able to move on land and in water. A sample creature is depicted in figure 3. Creatures are modeled as composed of “sticks”: connected finite-length line segments (though typical visualizations depict them as cylinders). Sticks are subjected to simulated physical forces, such as friction, gravity, and buoyancy (in a water environment). Stick creatures may have networks composed of three kinds of neurons: sensors, muscles, and interneurons. Connections between segments may contain either of two kinds of muscles: bending muscles and rotating muscles. The natural pairing of tensors and flexors is simplified into a single muscle that is able to move in either of two opposed directions. Creatures may also be equipped with sensors, of which there are three: smell sensors, touch sensors, and equilibrium sensors. Smell sensors allow the detection of energy sources – food balls – and other creatures, which may be cannibalized under certain conditions. Smell sensors give an increased signal both as a function of proximity of the stimulus and concentration of energy in the stimulus. Touch sensors give an increased signal as a function of proximity to envi-



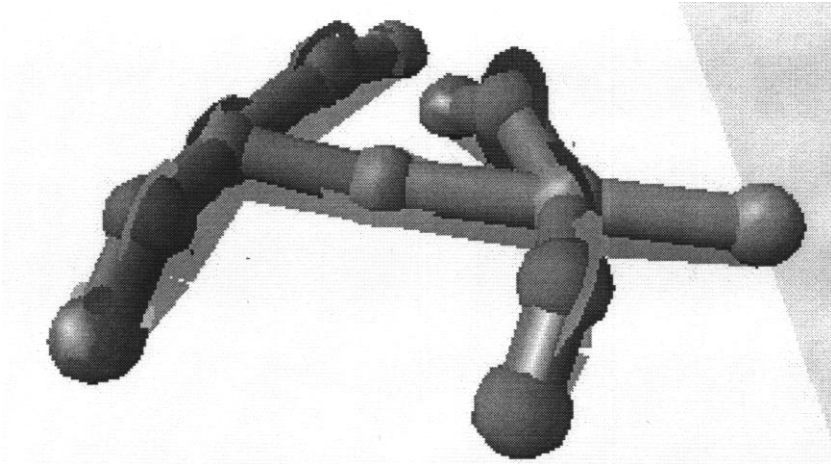


Figure 3. A sample Framsticks creature.

ronmental surfaces, such as the ground, obstacles, and barriers. An equilibrium sensor on a stick detects orientation relative to the gravitational field. A horizontal orientation yields a signal of 0, and the signal moves toward -1 or 1 as it is rotated in one of two directions to a vertical position.

Sensors and muscles may be connected by any number of interneurons. Framsticks allows for neural networks of any topology. Neuron states are a sigmoidal function of a weighted sum of inputs. The Framsticks simulator allows for the evolution of creatures via optimization in accordance with weighted combinations of the following fitness criteria: lifespan, horizontal velocity, horizontal distance, vertical position, vertical velocity, body size, and brain size (the number of inputs to neurons in a creature's neural net). Fitness is evaluated by simulating creatures in a virtual world. The structure of a simulated creature is determined by one of the genotypes in the gene pool. When a creature is "born" its genotype is either a clone of a preexisting genotype, a mutation of a preexisting genotype, or a crossover combination of two preexisting genotypes. (There is no sexual differentiation between creatures, but speciation may be introduced by restricting crossovers to relatively similar creatures.) The creature is born into the world with a finite store of consumable and replenishable energy, and the creature dies when either its energy runs out or it suffers a destructive collision with another creature. Fitness of a creature is evaluated by its performance during its lifetime. The calculation of the creature's fitness influences the number of instances of that creature's genotype represented in the genepool. Fitter individuals are more likely to have their genotypes reproduced.

User control of the fitness function is not the only way that evolution may be guided. Framsticks also allows for the user to select separately

which aspects of the genotype may undergo mutation and at what intensities and probabilities. One kind of option would be to allow mutations only to the creature's neural networks (for example, adding or deleting a neuron, adding or deleting connection), not to their bodies. Many of the simulations described below allowed mutations only of neural network properties and further restricted the mutations to only introducing changes of the weights of neural inputs and connections.

### **Modular and Nonmodular Control of Chemotaxis**

In this section I describe several creature designs that exhibit positive chemotaxis (food finding). Evolving minimally cognitive creatures from scratch is a daunting task, and so is building them outright. The creatures described in this section were "free formed," by which I mean that some of their architecture is due to user design, and some is due to evolution. A typical feature of free forming is that when creature designs are subjected to evolutionary pressures, the user frequently adjusts the fitness function along the way to "get what he wants." Free forming is a kind of engineering where evolution is used as a forge to temper creature designs. For example, a creature will be designed with a neural architecture such that it seems to the designer that it will map inputs to outputs appropriately to exhibit some target behavior like obstacle avoidance. While specifying the neural connections by hand is relatively trivial, hand coding optimal connection weights is daunting. The evolutionary algorithm, however, can pick up the slack: the designed creatures can be subjected to an evolutionary run whereby only mutations to connection weights are allowed and fitness is defined as amount of horizontal distance achieved in the creatures' lifetime. The creatures are allowed to evolve overnight, and when the experimenter returns to the lab the next day, the connection weights may have been optimized.

One of my goals in this section is to introduce the taxonomies of modular and nonmodular architectures in describing design solutions to the problem of positive chemotaxis. What I mean by modular control of chemotaxis is that the motor system responsible for the maintenance of locomotion is separate from sensory motor systems sensitive to the spatial location of the stimulus.

Creatures that exemplify the modular control of chemotactic behavior have separate systems for the continuation of forward locomotion and stimulus orientation. A particularly successful food finder is the creature Modular-B depicted in figure 4. Modular-B is a modification of the four-legged food finder created manually and then evolved by Miron Sadziak (see Komsinski 2001).

The neural network controller for Modular-B is composed of two distinct systems, depicted in figure 4. One system is the stimulus-orientation system, which consists of two smell sensors connected to a bending muscle in the creature's torso. When there is a higher degree of activity in

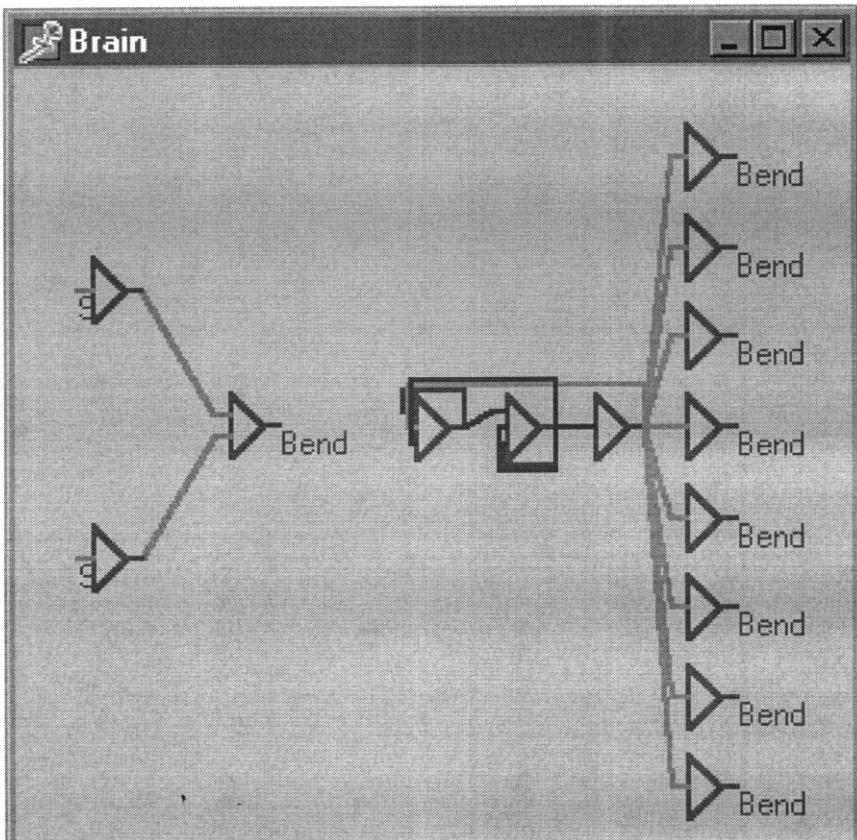


Figure 4. Neural network controller for Modular-B.

the right sensor than in the left, the torso muscle bends the creature toward the right. Likewise, *mutatis mutandis*, for greater activity in the left sensor. (Crossing the connections from the sensors to the torso muscle would result in an energy avoider. This can be a relatively useful behavior in environments where creatures must avoid each other to avoid destructive collisions.) The other neural system in Modular-B is the locomotion system, which consists of a central pattern generator that drives the leg muscles in a synchronous gait.

Creatures that exhibit nonmodular neural solutions to the problem of chemotaxis do not have distinct circuits for forward locomotion and stimulus detection. For example, consider the water creature *Eel2* depicted in figure 5. *Eel2*'s neural network, depicted in figure 6, involves a central pattern generator that drives sinusoidal swimming and is modulated by a

single-smell-sensor input. Sufficiently high input signals seize the activity of the central pattern generator. When the input decreases, oscillations in the CPG resume. Eel2 exhibits a pattern of food-finding behavior that is relatively typical of nonmodular food finders. Eel2 swims around in wide curved arcs. Increased activity in the smell sensor results in the arcs becoming tighter, causing the creature to swim in a small circle close enough to the stimulus source to absorb it. In cases where the stimulus signal is sufficiently high, Eel2 goes into a “seizure” and stops swimming altogether. Over evolutionary time Eel2’s network became tuned in such a way that the seizures would occur only if the creature was close enough to the stimulus source to absorb the food. After a certain amount of food is absorbed, the seizure ends and Eel2 swims in ever-widening circles, though still making contact with the food source. Only when the food is gone will Eel2’s body straighten sufficiently to swim away from that location to find the next meal.

The contrast between modular and nonmodular solutions is especially clear in a comparison between Modular-B and Eel2. Proximity to food does nothing to modulate Modular-B’s central pattern generator and thus does nothing to modulate Modular-B’s means of forward locomotion. In contrast, stimulus signals modulate Eel2’s locomotion system by changing the waveform of the pattern-generator output. The contrast between food-finding networks of Modular-B and Eel2 also involves important differences in the ways these networks represent and compute information about

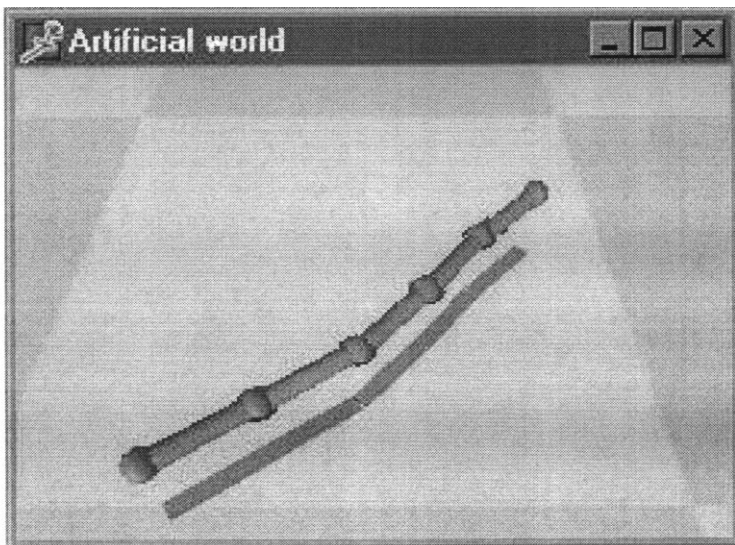


Figure 5. Eel2, a water creature that employs sinusoidal swimming.

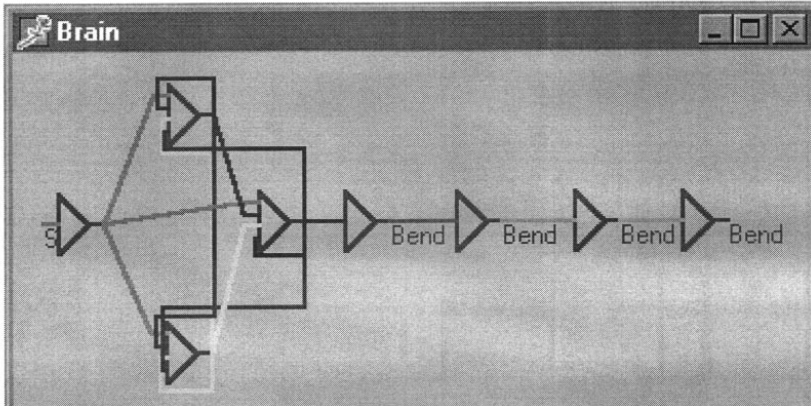


Figure 6. Eel2's neural network.

the creatures' environment. I postpone directly addressing these issues until after laying down further background on the notion of representation.

### Mental Representation in a Biological Context

Studying the minimally cognitive behavior of chemotaxis in artificial life provides an excellent opportunity to get at issues concerning the most basic cases of mental representation in an evolutionary context. The basic approach to representation assumed here is both teleological and information theoretic: a representation is a state of an organism (typically, a state of the organism's nervous system) that has the function of carrying information about environmental and bodily states (Dretske, 1988; Milikan 1984; 1993). It is presumed here as a starting point for discussion, but as the discussion progresses, we shall find it necessary to alter the initial characterization of representation.

To see what is useful about this characterization of representation, or at least how it is used, consider the following example. Von Uexkull describes the mental life of a tick as being composed of the following behavioral components (Von Uexkull 1934). A tick clings to leaves on trees waiting for a mammal to pass underneath. The mammal gives off butyric acid, which the tick has detectors for. The detection of butyric acid triggers the tick to release its grasp of the leaf, and the tick falls onto the mammal. Touch sensors detect the appropriate proximity between the tick and the mammal to trigger the tick to run around. The tick continues this running until thermoreceptors detect a high enough temperature to trigger the tick's burrowing response. The tick will then burrow into the skin of the mammal, where it will find the blood it feeds on.

Key adaptive behaviors of the tick are thus driven by mechanisms that

function as detectors: they have the functions of carrying information about environmental features. Like ticks, the creatures modeled in Framsticks have chemical receptors and touch sensors, but instead of thermoreceptors, they have equilibrium sensors. Thus, basic abilities to represent environmental features may be built into Framstick creatures at the level of individual transducers. For example, heightened activity in a smell sensor serves to carry information about, and thus represent, concentrations of chemical energy.

Informational approaches to understanding representation are typically cast as opposed to isomorphism-based approaches: approaches that see the relation between representation and represented as fundamentally one of resemblance (see, for example, Cummins 1996). Prototypical instances of nonmental representations can help draw out the opposition. On the informational side of the contrast, a ringing doorbell can have the function of carrying the information that someone is at the door and thus represent the fact that someone is at the door without in any significant way resembling the state of affairs of someone being at the door. In contrast, a photograph of someone standing at a door represents that state of affairs in virtue of there being much meaningful resemblance between the photograph and the photographed scene. The geometric arrangement of color and shade in the photograph is isomorphic to the arrangement in the actual scene. For instance, the door is represented as being taller than the person in virtue of the door image being taller than the person image, and so on.

There are, however, other examples that may make the contrast between informational representations and isomorphic representations seem not as sharp. Consider thermometers as devices that carry information about temperature. A mercury thermometer has a column of mercury that changes in height such that the higher the temperature gets, the higher the mercury gets. The column height of, say, three centimeters represents 50° C in the environment in part because that column height is caused by, and thus carries information about, that temperature. But we may also see that isomorphism is at play in the way column heights represent temperatures. The thermometer instantiates an ordered series of physical magnitudes (mercury-column heights) that is isomorphic to an ordered series of physical magnitudes in the environment (temperature). So thermometers constitute a mixed case of informational and isomorphic representations.

Indeed, returning to the case of photographs, they do not represent solely in virtue of isomorphism; causal-cum-informational relations play constitutive roles as well. A photograph may be a picture of Joe and not his identical twin brother Moe in virtue of the causal relations to Joe and in spite of equally resembling both Joe and Moe. The view of information put forward here is thus in keeping with an etymological understanding of information as *inFORMation*: something carries information about something else in part because of a sharing of form. The boot print carries information about the boot in part because the mud becomes rather literally

inFORMed by the boot. The mixture of information and isomorphism illustrated in terms of the thermometer is also exemplified by Framsticks sensor neurons. For example, the range of activity of a sensory neuron is isomorphic to the distance away from the stimulus source. And part of what makes activity in a sensor represent food, and not something else, is that the activity is caused by food, and thus carries information about food, and not something else.

The mere fact that a state of an organism carries information about some environmental feature is insufficient to make that state have the function of carrying that information. Consider people who have fair hair that gets bleached in sunlight. The bleached color of the hair carries information about the light it was exposed to, but the human organism does not use that information, at least not in the significant way in which information about light picked up through the eyes is used. The state of the organism may *carry* the information, but there is no state of the organism that constitutes the *use* of that information. In other words, the information may be *encoded*, but to count as a representation there must be a process by which the organism *decodes* that information. Millikan makes the point in terms of mechanisms of representational production and mechanisms of representational consumption (Millikan 1984; 1993). According to Millikan, in order for a state of an organism to come to have the function (in the teleological sense of the term) of carrying information, that state has to be naturally selected to carry information. In order to be naturally selected, the carrying of information must have some effect on the organism's behavior that contributes to the organism's fitness. And in order for that to happen, there must be some system of the organism that can channel that information into the modulation of behavior: a system that decodes the encoded information in a potentially adaptive way.

### **Representation in Synthetic Neuroethology**

With these minimal remarks about key notions of representation in place – the notions of information, isomorphism, encoding, and decoding – we are now in a position to see how Framsticks offers a platform for prosthetically controlled thought experiments concerning mental representation.

I focus here on the way food-finding Framstick creatures utilize networks with smell-sensor inputs to represent the spatial relations between the creature and the stimulus source. A return to the discussion of the contrast between modular and nonmodular solutions to food finding will be especially useful in this regard. Recall the way that Eel2, a creature with a single smell sensor, was capable of finding food. Activity in a single smell sensor represents proximity to the stimulus source: the higher the activity, the closer the stimulus source. Only one dimension of spatial information can be encoded in this one-sensor system. Eel2's network is incapable of representing or computing anything more specific about the

stimulus source, like the direction of the source with respect to the creature. In contrast, the two-sensor system of Modular-B is capable of encoding information about the two-dimensional location with respect to the stimulus source, information that is decoded by the single turning muscle in Modular-B's torso. A greater amount of activity in the right sensor than in the left sensor indicates the stimulus being farther to the right. Conversely, a greater amount of activity in the left sensor than in the right sensor indicates the stimulus being farther to the left. The nervous system of Modular-B is thus capable of representing (in an egocentric reference frame) the two-dimensional location of a stimulus source. Activity in each of the individual sensors represents one dimension of spatial information: near versus far. The network involving two sensors and one turning muscle is able to compute two dimensions of spatial information: right versus left as well as near versus far.

While the one-sensor system seems to be at a clear disadvantage to the two-sensor system, the one-sensor system is not entirely useless to Eel2. Information about proximity encoded in the sensor is decoded by the muscular system by making Eel2 swim in circles such that higher amounts of sensor activity result in smaller-diameter circles. This behavior increases the likelihood of hitting the food source.

The evolutionary advantages of being able to represent more features of the environment can be demonstrated experimentally with the Framsticks platform. In the experiment described below, creatures with identical morphologies (bodies) but different neural architectures were evolved in conditions in which food was present, fitness was defined only in terms of horizontal distance, and mutations were allowed only in neural weights (defined as input weights to sensors, motors, and interneurons). The morphologies employed were the same as for Modular-B described above. The neural network controllers were significantly more complex than Modular-B's. The creatures below differed from each other in having either two, one, or no smell sensors. Otherwise, the topology of their networks all conformed to the following scheme. There were feed-forward connections from each sensor to each of four interneurons in a hidden layer. An additional four interneurons (a second hidden layer) each had feed-forward and feedback connections to each of the neurons in the first hidden layer. Each of the nine outputs (one torso and eight leg muscles) received feed-forward connections from each of the neurons in the first hidden layer. Creatures began with all neural weights set to zero and were evolved for 200 million steps of the simulation. Each of the three creature architectures (with initial weights of 0) were subjected to five evolutionary runs. Population statistics were sampled after every 20 million steps of the simulation. The results are graphed in figure 7. Having two smell sensors bestowed a clear advantage over having only one smell sensor, though having one sensor was better than having none at all.



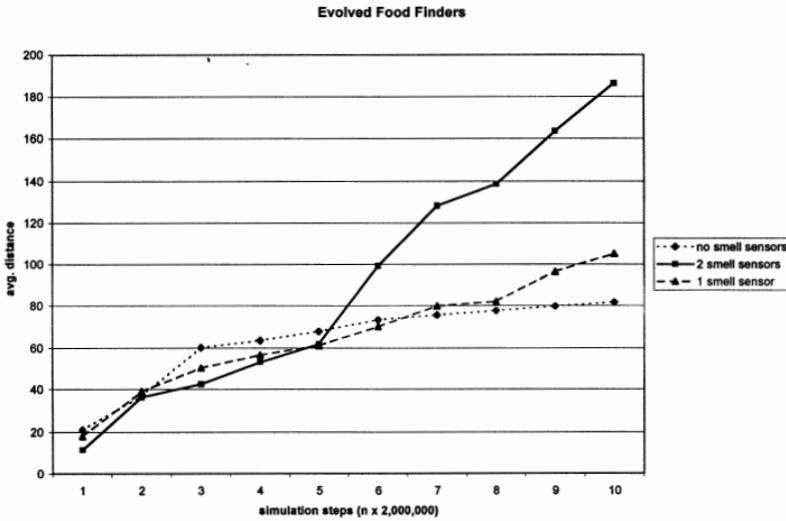


Figure 7. Results of the experiment comparing the evolution of neural weights for land food finders with two, one, and no smell sensors.

Examining creatures that walk on land to find food allows us to see how simple artificial neural networks are capable of supporting representations of spatial locations of stimuli in one and two dimensions. Switching from land to water renders it possible to make Framsticks creatures that represent the spatial location of stimuli sources in three dimensions. One way to accomplish this would be with a four-sensor swimming creature. Two of the sensors would be aligned along the creature's horizontal plane and drive a turning muscle that bends the creature along the horizontal plane. The other two sensors would be aligned along the creature's vertical plane and drive a second turning muscle that bends the creature along the vertical plane. Such a system would allow a swimming food finder in relatively deep water to find food sources placed at varying depths. A similar feat could be accomplished with a swimming creature that had only three sensors arranged in a triangle, with, say, one sensor on the top of the creature and the other two on the bottom. The difference between the right and left bottom sensors could be used to drive the horizontal turning muscle. The difference between the top sensor and the sum of the bottom sensors could be used to drive the vertical turning muscle.

## Conclusion

The sorts of questions asked and methods pursued in this essay follow

those advocated by Daniel Dennett in his description of artificial life as a mode of doing philosophy. Dennett writes: “In short Alife is the creation of prosthetically controlled thought experiments of indefinite complexity. . . . Philosophers who see this opportunity will want to leap into the field, at whatever level of abstraction suits their interests, and gird their conceptual loins with the simulational virtuosity of computers” (Dennett 1998, 262).

Among the various questions Dennett sees this approach as equipped to address is this: “Can we build a gradualist bridge from simple amoeba-like automata to highly purposive intentional systems, with identifiable goals, beliefs, etc.?” (Dennett 1998, 262). The sorts of projects I have sketched in this essay do not yet definitively answer this question, but they do provide a useful start. Further, I have argued what the gradualist bridge will look like: representational and computational systems will figure very early in the evolutionary trajectory from mindless automata to minded machines.

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

### **Acknowledgments**

This work was supported in part by a grant from the McDonnell Project in Philosophy and the Neurosciences. I thank Eric Steinhart and Shawn Gaston for stimulating discussions of this material.

### **References**

- Beer, Randall. 1990. *Intelligence as Adaptive Behavior*. San Diego: Academic Press.
- Braitenberg, Valentino. 1984. *Vehicles: Experiments in Synthetic Psychology*. Cambridge, Mass.: MIT Press.
- Brooks, Rodney. 1991. “Intelligence Without Representation.” *Artificial Intelligence* 47: 139–59.
- Cummins, Robert. 1996. *Representations, Targets, and Attitudes*. Cambridge, Mass.: MIT Press.
- Dennett, Daniel. 1998. *Brainchildren*. Cambridge, Mass.: MIT Press.
- Dretske, Fred. 1988. *Explaining Behavior*. Cambridge, Mass.: MIT Press.
- Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

- Haugeland, John. 1991. "Representational Genera." In *Philosophy and Connectionist Theory*, edited by W. Ramsey and S. Stich. Hillsdale, N.J.: Erlbaum.
- Komosinski, Maciej. 2000. "The World of Framsticks: Simulation, Evolution, Interaction." In *Proceedings of 2nd International Conference on Virtual Worlds* (Paris, France). Berlin: Springer-Verlag.
- . 2001. Framsticks website <http://www.frams.poznan.pl>.
- Liekens, Anthony. 2001. Artificial life website <http://www.alife.org>.
- Llinas, Rodolfo. 2001. *I of the Vortex: From Neurons to Self*. Cambridge, Mass.: MIT Press.
- Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, Mass.: MIT Press.
- . 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass.: MIT Press.
- O'Regan, J. Kevin, and Alva Noë. 2001. "A Sensorimotor Account of Vision and Visual Consciousness." *Behavioral and Brain Sciences* 24, no. 5.
- Von Uexkull, Jakob. 1934. *Streifzüge durch die Umwelten von Tieren und Menschen* (a stroll through the worlds of animals and men). Berlin: Springer.