Varieties of Representation in Evolved and Embodied Neural Networks

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

ABSTRACT
In this paper I discuss one of the key issues in the philosophy of neuroscience: neurosemantics. The project of neurosemantics involves explaining what it means for states of neurons and neural systems to have representational contents. Neurosemantics thus involves issues of common concern between the philosophy of neuroscience and philosophy of mind. I discuss a problem that arises for accounts of representational content that I call "the economy problem": the problem of showing that a candidate theory of mental representation can bear the work required within in the causal economy of a mind and an organism. My approach in the current paper is to explore this and other key themes in neurosemantics through the use of computer models of neural networks embodied and evolved in virtual organisms. The models allow for the laying bare of the causal economies of entire yet simple artificial organisms so that the relations between the neural bases of, for instance, representation in perception and memory can be regarded in the context of an entire organism. On the basis of these simulations, I argue for an account of neurosemantics adequate for the solution of the economy problem.

KEYWORDS
artificial life; evolution; mental representation; neural networks; philosophy of neuroscience

1. Introduction

Suppose you were given a handful of neurons and assigned the task of using them to build a mind. How would you proceed? Would you ask for more neurons, being convinced that minds arise from only the most complex brains? Would you ask for a body to put the neurons in, being convinced that the mind is essentially embodied and can only exist when there's a body to shove and be shoved by? Would you ask for an evolving population of bodies and brains, being convinced that the mind is essentially a product of evolution through natural selection?

The prospect of having a handful of neurons to play with along with the options of embodying and evolving any neural network is not as terribly far off nowadays as it may have once seemed. The virtual equivalent of the various activities described above are all available through the employment of current Artificial Life software. A few years ago, in describing the philosophical potential of Artificial Life (A Life), the philosopher Daniel Dennett wrote,

> 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 projects Dennett recommends is one that attempts to answer the following question "Can we build a gradualist bridge from simple amoeba-like automata to highly purposive intentional systems, with identifiable goals, beliefs, etc.?" (Dennett 1998: 262). In this paper, I describe artificial life experiments I have conducted in the pursuit of precisely this goal.

If given a handful of neurons and assigned the task of using them to build a mind, one of the first things I would do is figure out what it takes to endow neural states with representational content. This isn't a terribly original thought, but popularity oft enhances plausibility. The thought is this: minds are made of mental representations. Thus, if a mind is to be constructed of neurons, there better be a way to make representational states out of neural states. Further, if embodiment and evolution turn out to be requirements on mindedness, it is perhaps because they are requirements on mental representation. While the methods employed here are from Artificial Life, the targets are neurophilosophical: to understand, in abstract terms, the simplest sets of conditions for the implementation of mental representations in evolved and embodied neural networks.

The outline of things to come is as follows. First I spell out what I take key issues of neurosemantics to be. Next I discuss a problem that arises for attempts to naturalize representation that I call "the economy problem". The economy problem is the problem of demonstrating the consistency of a naturalization of the representation relation with an account of the roles representations play within the causal economy of an entire mind as well as an entire organism. Next I propose the utility of methodologies from artificial life for getting a grip on the economy problem for representation. I also sketch a pre-naturalized account of the minimal features of representations so that we know what we are looking for when we evaluate the control structures of artificial organisms. I turn then to describe artificial life simulations that I have conducted to show the varieties of representation at work in the neural controllers for evolving populations of relatively simple artificial organisms. Such simulations allow the causal economies of entire organisms to be laid bare for neurophilosophical scrutiny. Finally, I sketch a solution to the economy problem with reference to the neural networks of these simulated organisms.

2. Two questions of neurosemantics

This paper takes up two questions of neurosemantics. The first is the question of how neural states of organisms have representational contents. The second is the related question of how organisms *evolved* to have neural states with representational contents.

The answers offered here depend on certain assumptions, among which include the following. There really are representational contents: they exist (contra antirepresentationalists such as Beer (1990) and Brooks (1991)) and do so independently of our finding it convenient to say that they exist (contra Dennett (1987) and various other instrumentalists, interpretationists, and ascripitivists). Among the things that have representational contents are states of organisms. States of organisms with representational contents are one and the same as mental representations. For organisms with nervous systems, the right level to look for mental representations is the neural level, thus, for these creatures; psychosemantics is one and the same as neurosemantics (contra Fodor (1975) and other closet dualists who claim instead that representational states merely *supervene* on or are merely *realized by* neural states.)

The above list of assumptions treats as closed so many open questions in contemporary philosophy that one may wonder what is left to say of philosophical significance concerning the two above-mentioned questions of neurosemantics. As I see it, the two questions constitute philosophical "how possible" questions. Answering them will constitute, in part, an argument that the above-mentioned set of assumptions constitutes a consistent set of philosophical propositions. Demonstrating this consistency is itself a worthy philosophical project. It is not, however, the sole ambition of this paper, as we will see in a bit.

I intend the two questions of neurosemantics to differ in that the first addresses synchronic concerns while the second addresses diachronic concerns. I spell this out in further detail by beginning with the synchronic concerns. I am an organism, and as I view these words on the screen of my laptop various states of my nervous system carry various representational contents. Some states represent the luminance of the viewed objects. Some states represent the shapes of objects in my immediate environment. Some states represent the egocentric position and orientation of various objects. Some states represent events in the past and constitute my memory. I remember that I turned the coffee pot on 15 minutes ago. Other states represent events in the future and constitute my will. I currently intend to rise from my chair in a few minutes and make some eggs and toast to go with my coffee. If all goes according to plan, then in the immediate future my exertions will bring the world into conformity with my will. In spite of the temporally extended nature of memory and intention (and to some degree, perception), there is nonetheless a synchronic story to tell about how certain states of my nervous system, but not others, currently serve to constitute representations about objects and events in my

present, past, and future. What structural and functional features of my nervous system serve as the synchronic base of my representational capacities?

The second, diachronic, question of neurosemantics serves to constrain answers to the first. It is one thing to hypothesize how a system works, it is yet another to explain how it came to be. A hypothesis about the current workings of some target system lacks a certain amount of plausibility if it remains an utter mystery how such a thing could have come into existence. This is not to assert that having a history is essential to representational contents, as does, for instance Millikan (1984, 1993). It is instead to raise, as both a philosophical and scientific question, what those histories are. A widespread assumption is that both nervous systems and representations came relatively late in the evolution of organisms. Organisms without nervous systems had been around much longer than organisms with nervous systems. Likewise, organisms without representations predate organisms with representations[1]. Related to the question of how organisms evolved neural states with representational contents are questions of the temporal and causal priority of representations and nervous systems. Did nervous systems exist before or after organisms with representational states? Did nervous systems evolve in order to provide the means of representing or did nervous systems serve some nonrepresentational function first?[2]

Among the synchronic and diachronic questions are those questions that constitute the central problematic of the cottage industry in philosophy of mind that churns out theories of representational content. Must representations resemble or bear some non-trivial isomorphism to that which they represent? That they do is an idea historically attributed to Aristotle and espoused in contemporary debates by Cummins (1996). To what degree can informational or causal head-world relations underwrite a naturalistic representation relation? That they do is an idea historically attributed to Locke and espoused in contemporary debates by Dretske (1995). What roles must the forges of evolution or learning history play in welding representations to their contents? That they must is espoused in contemporary debates by Milllikan (1984, 1993) and Dretske (1988).

3. The economy problem.

One sort of problem plagues many if not all previous attempts to answer these philosophical questions. The various authors of competing theories of representational content have done relatively little to spell out how their naturalized representations would live and breathe in the economy of a single mind, yet alone in the network of forces contributing to the well functioning of an entire organism. Call this "the economy problem". Raising this as a problem is not to deny that philosophers have paid a certain amount of lip service to the importance of such causal relations. The complaint is that they have supplied insufficient information to show how mental representations, defined in their proprietary ways, could combine to constitute a mind and control an organism. This problem is especially acute in face of the fact that the

point of explaining what representations are is to understand what minds are. Contemporary interest in mental representation is largely driven by the view that explaining what representations are is, if not the lion's share of the journey to a completed theory of mind, at least a major step. This view is held aloft by the belief that the success, or at least popularity, of cognitive science vindicates the representational theory of mind.

To charge theories of representation with the challenge of solving the economy problem is not to question the truth of the theories, but instead to question their completeness. It is one thing to say that a dollar is worth a certain quantity of gold—this is true yet incomplete. It is yet another to explain how a dollar interacts with other items and processes in a way that genuinely explains the value and worth of a dollar. Analogously, it is one thing to say that a representation is, say, a state of an organism that is asymmetrically dependent on what that state is about (Fodor 1990). This may very well be true. But the questions remain of how and whether such asymmetrically dependent states can do what representations need to do, such as underwrite perception, memory, etc.

To see in further detail how the economy problem arises, consider the kind of stock example typical of this literature. Cognitive agent x has a mental representation heretofore referred to as "/cow/". As the story goes, /cow/ means cow, that is, agent x has /cow/ in its head and /cow/ represents a cow, or cow-ness, or cows in general. Representational content theorists like to focus on the case of perceptual belief (which is not, in itself, a bad thing), so the typically discussed example of when /cow/ gets tokened is when agent x is in perceptual causal contact with a cow and comes to believe that there is a cow, presumably by having, in its head /there/ + /is/ + /a/ + /cow/ or some other concatenation of /cow/ with various other mental representations. The main question addressed in this literature is how /cow/, a physical sort of thing in the agent's head, comes to represent a cow, a physical sort of thing outside of the agent's head. Focus on the perceptual case has made causal informational proposals seem rather attractive to quite a few people[3], so let us focus on the following sort of suggestion, namely, that /cow/ represents cow because in typical scenarios, or ideal scenarios, or in the relevant evolutionary scenarios, /cow/s are caused by cows, that is, /cow/s carry information about cows. Thus, tokenings of /cow/s in the heads of agent x and his relatives are part of the operation of a cow-detector. Now this sort of story has a certain amount of plausibility for the case of perceptual belief, or the representations involved in sensory perception more generally, but presumably, /cow/s, that is, mental representations of cows, have a lot more work to do than that. Consider that /cow/s are used to remember cows, to make plans concerning future encounters with cows, and to reason about counterfactual conditions concerning cows (like, what if a cow burst into this room right now?). The methodological concern I intend to raise here is the concern that what might seem like an good thing to say in connection with perception may not generalize to all the other sorts of things mental representations are supposed to do. A widespread

presumption, and a not necessarily bad one, is that the /cow/s you find in the perceptual case are the same things that will be deployed in the memory, planning, and counterfactual reasoning cases too. The presumption, inherited from a long empiricist tradition, is that what ever happens in perception to wed representations to their contents, can simply be passed along and retained for use in non perceptual mental tasks. A really literal implementation, then, would have whatever happens to items in the perception box be sufficient to mark those items (picture them as punch cards, if you like) as bearing representational contents. Those items can thus be passed to other boxes in the cognitive economy, and retain their marks of representational content even after they may go on to play quite different causal roles. This is an interesting suggestion, but certainly open for questioning. Perhaps, instead, the sorts of conditions that bestow representational contents onto perceptual states are very different than the conditions on representation in memory, which are yet different from the conditions for representation in planning, counterfactual reasoning, and so on.

A second concern, not unrelated to the first, is how you tell what and where the /cows/ are in the first place. Focusing on the case of perceptual belief brings with it certain natural suggestions: point agent x at some cows and look for the brain bits that seem to "light up" the most. Much talk of representation in neuroscience is accompanied by precisely this sort of methodology. But what lights up during the retrieval of memories of cows or counterfactual reasoning about cows? Do the same bits light up as in perception or not? And more to the point, how will various theories of representational content cope with the different possible answers to the previous two questions?

The economy problem is thus a cluster of many closely related questions and concerns. They may seem particularly daunting to answer. What I am asking about is how representations fit into the rest of a mind, that is, whether recent stories about representational content are consistent with the most plausible stories about what counts as fitting in. It would be nice, in this context, to have some examples of minds simple enough so that we can examine them in their entirety. So, instead of starting with a theory of representation, either explicit or tacit, and poking around in real, complex, human brains looking for the /cow/s, what I propose is a somewhat reversed strategy: start of with some simple minds of some simple organisms, describing how their survival promoting behaviors are accomplished, and then work backwards to a naturalization of representation.

4. Animat methodology

These questions concerning representation are pursued here by employing a cognitive scientific methodology come to be known recently as bottom-up AI or the animat approach (for a review see Guillot and Meyer 2001). An animat is an artificial animal, either computer simulated or robotic. Animat methodology involves three characteristic explanatory strategies: synthesis, holism, and incrementalism. The synthetic element involves explaining target phenomena

by attempting to synthesize artificial versions of them, a characteristic inherited in large part from earlier versions of Artificial Intelligence (Good Old Fashioned Artificial Intelligence (GOFAI) as well as connectionist approaches). The holism referred to here is not necessarily restricted to the semantic holism familiar in other areas of philosophy of mind or cognitive science[4] but is instead concerned with function more generally. The holistic take on function is that the function of an organ or a behavior is best understood in the context of the whole organism, or, more broadly still, in the context of the organism's physical and/or social environment. It is thus both embodied and embedded (Clark 1997). However, this holistic impulse might seem to conflict with attempts to synthesize phenomena. Synthesis must simplify to be tractable, yet whole organisms are more complex than their subsystems, and social systems and ecosystems are even more complex. An older strategy of simplification involves focusing on subsystems of human cognitive processes as in GOFAI and connectionist models of word recognition. The comparatively newer strategy of simplification embraced by the Animat approach involves focusing on the entirety of organisms much simpler than the human case, thus heeding Dennett's rallying cry/question, "Why not the whole iguana?" (1998: 309). In animat research projects of synthesis involve modeling the simplest intelligent behaviors such as obstacle avoidance and food finding by chemotaxis. The incrementalism of the animat approach involves building up from these simplest cases to the more complex via a gradual addition of complicating factors, as in, for instance, roboticist Rodney Brooks'(1999) ongoing project of building an incrementalist bridge from robotic insects like Attilla through to the humanoid robot, Cog.

Some of the earliest practitioners of animat methodology did not emanate from the engineering and computer sciences, but were instead neuroscientists. The neuroscientists Grey Walter (1963) and Valentino Braitenberg (1984) have had a deep impact on the practice of animat methodology. Walter built his robotic "turtles" Elmer and Elsie out of vacuum tubes and other electric components of the day. Elmer and Elsie were wheeled animats with perceptual sensitivity to light and sound and capable of a rudimentary form of associative learning. Unlike Walter, Braitenberg did not implement his ideas in hardware, but the thought experiments conducted in *Vehicles: Experiments in Synthetic Psychology* inspired the projects of many robotocists. Braitenberg's animats, the vehicles of his book's title, were envisioned as relatively simple collections of sensors and motors with excitatory and inhibitory connections between them. Figure 1 depicts two of Braitenberg's simplest vehicles in the proximity of a stimulus.

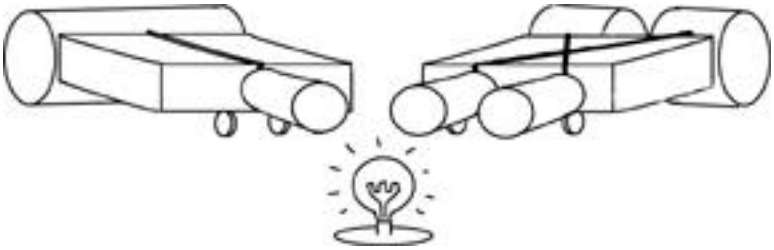

Figure 1. Two Braitenberg vehicles and a stimulus source. Figure drawn by Pete Mandik.

The vehicle on the left has of a single sensor on its front connected to a single motor in its rear. If the line connecting the sensor to the motor is excitatory, then increased sensor activity will result in increased motor activity. Stimulation of the sensor will result in the vehicle accelerating toward the stimulus. The vehicle on the right has two sensors with crossed connections to two motors. If the connections are excitatory, the vehicle will turn toward a stimulus. For example, if the stimulus is to the right of the vehicle this will result in higher activity in the right sensor than the left sensor, resulting in higher activity in the left motor than the right motor. If, in contrast, the excitatory connections are parallel and not crossed, then the creature will move away from the light.

Braitenberg discusses the degree to which it seems natural to attribute psychological states to the vehicles, for example, to describe these creatures as loving or fearing the stimulus source. Another point of interest is how relatively simple systems can give rise to models of coherent behaviors such as taxis (the movement toward or away from a stimulus source) and kinesis (movement triggered by a stimulus).

Two points of immediate concern fall out of considerations of Braitenberg's vehicles. The first, already touched on briefly, is how amenable they are to psychological description. The second is how amenable they are to neuroscientific and neuroetholgical description. The terms "sensor", "motor", "excitatory connection" and "inhibitory connection" have natural applications in the neurosciences. And the promise of seeing how they work together in the context of an entire organism to give rise to survival promoting behaviors like food finding by positive phototaxis or chemotaxis sparks the hope that along this path lies accounts of the evolutionary function of the earliest brains and nervous systems more generally.

Contemporary practitioners of animat methodology have at their hands techniques for addressing these diachronic evolutionary questions that arise. Of course Braitenberg speculated as to how, hypothetically, the design of his vehicles might be relegated to a process of "natural" selection. Braitenberg imagines several engineers well stocked with lots of kinds of sensors, motors, connecting wires other parts for making all sorts of vehicles. As each vehicle is

made, it is placed onto a table-top containing a variety of obstacles and beneficial and noxious stimulus sources. Any vehicle that falls off of the table is considered a failure and its parts used for scrap. As the engineers make additional vehicles to put on the table, they pretty much copy from the vehicles already on the table, not the ones that have fallen over the table ledge. Thus, relatively successful designs are perpetuated in future generations. Additionally, the fidelity of the copying process is not perfect. Novel "mutations" are thus introduced into the pool of designs. (1984: 26-27). Nowadays many computer programs exist that allow for the evolution of minimally cognitive behaviors in populations of relatively simple neural network controlled critters (for a review see Taylor and Massey 2001). Such programs allow for the simulation of evolution by natural selection by providing for the mechanisms of the variable inheritance of fitness. Such programs allow for simulations that capture the embodied, embedded, and evolutionary aspects of cognition.

The point of the simulations described below is to show that relatively simple autonomous agents—agents with neural controllers of only, for example, a dozen neurons and neural connections—are capable of acquiring and sustaining in an evolutionary context several varieties of mental representation. The successes of these simulations have implications for answering the synchronic and diachronic questions of neurosemantics as well as addressing the economy problem. The strategy of the remainder of the paper is as follows. First I present a pre-naturalized characterization of mental representation: a sketch of, in general, why we think there are any mental representations at all and what it means to say so. Second, after the pre-naturalized characterization is in hand, I present the simulations. Third, and finally, I discuss the implications of the simulations for fleshing out naturalizations of representation.

5. A pre-naturalized characterization of representation

The point of this section is to preempt the following common response to the simulations that follow: "Yeah, but why call *that stuff* 'representation'?" The brief story that follows is not designed to convince any philosophers reading it, but to remind them of longer stories of which they should be quite familiar.

Let us begin by considering why we think there are any representations at all, by looking at how the notion of mental representation applies to our own, human, case. As already discussed above, perception plays a central role in philosophical thought concerning representation. At least one explanation for this is the Cartesian preoccupation with perceptual error. The idea that perception is representational gets a powerful grip once we begin to puzzle over the distinction between the way things are and the way they seem. A coin may seem oval even though it is round and a half submerged stick may seem bent even though it is straight. In dreams, we may seem to be flying or chased by monsters even though we are tucked safely into bed. Instances in which our perceptions tell us something contrary to reality call attention to the fact that

perception is in the business of telling in the first place. Like sentences, which we use to tell each other information (like "the sky is blue"), the products of our senses have a content, (like "the sky is blue") and in both cases, this content can sometimes be accurate and at other times be inaccurate. Once the representational character of perception is called to our attention, it takes little additional effort to notice important analogies between perceptions and other mental states. Just as our perceptions represent the world as being certain ways, so do our beliefs and memories. I believe that the world is round and remember that the speed of light is faster than the speed of sound.

The analogy between mental states and language mentioned earlier helps us draw out another way in which mental states have representational contents. In spoken and written language there is a distinction between declarative sentences (e.g. "You have a clean room") and imperative sentences (e.g. "Clean your room"). Declaratives have truth conditions but imperatives do not. Imperatives have, instead, satisfaction conditions. A similar distinction applies to mental states. Perceptions, beliefs, and memories have declarative contents whereas intentions have contents more akin to imperatives. One way of understanding the distinction here is in terms of the direction of fit between mind and world (Searle 1983). For example, the point of intentions is to have the world brought into conformity with them whereas the point of perceptual beliefs is to have them be brought into conformity with the world. This distinction echoes the one drawn out by Anscombe (1957) in her discussion of two kinds of lists. A person in a grocery store goes up and down the aisle putting into his cart items on his grocery list. The shopper is tailed by a private investigator making a record of the shopper's selection. At the end of the day both the shopper and the private investigator have very similar pieces of paper with the same words written on them. But these two lists serve very different purposes as is evident by their differing correctness conditions. If the private investigator has "bananas" on the list, but no bananas were in the cart, he can rectify the situation by erasing "bananas". In contrast, if the shopper returns home without bananas and he realizes that "bananas" was on the list, unlike the investigator, the shopper cannot rectify the situation by erasing "bananas".

Based on the above remarks, we can come up with the following sketchy characterization of mental representations. First off, mental representations are states of organisms that are capable of being about non-actual as well as actual states of affairs, that is, they exhibit intentionality. Second, the various kinds of mental representations (memories, percepts, etc.) may be distinguished both in terms of what they represent and in the causal roles they play within the organism. For example, percepts represent states of affairs in the present and are supposed to be the causal consequences of these states of affairs. Memories represent states of affairs in the past and, through causal relations to percepts, are causal consequences of these past events. Intentions represent states of affairs in the future and are supposed to count among the causal antecedents of these states of affairs. There is, of course, much more to be said in order to transform these sketches into full-blown theories, but for now they will serve as

rough guides for the search for representations in the causal economies of the artificial creatures discussed below.

6. Overview of the simulations

The simulations described below were run using Framsticks 3-D Artificial Life Software developed by Maciej Komosinski and Szymon Ulatowski (Komosinski 2000, 2001). Framsticks allows for the simulation and evolution of artificial organisms. Organisms are modeled as collections of connected line segments ("sticks"), although visualizations usually depict these sticks as cylinders. A sample creature is depicted in figure 2. The simulated physics of the Framsticks virtual world allows for the specification of the properties of the sticks such as weight, friction, elasticity, and resilience. Additionally, the world may be modified to allow for the simulation of underwater or dry land environments. Creature construction allows for the use of neural network controllers for the determination of creature behavior. The neural networks of the creatures may be composed of sensory input neurons, motor output neurons (for muscles located at the joints between sticks) and interneurons. The state of each neuron is a sigmoidal function of the weighted sum of the neuron's inputs. Figure 3 depicts the neural network of the creature from figure 2.

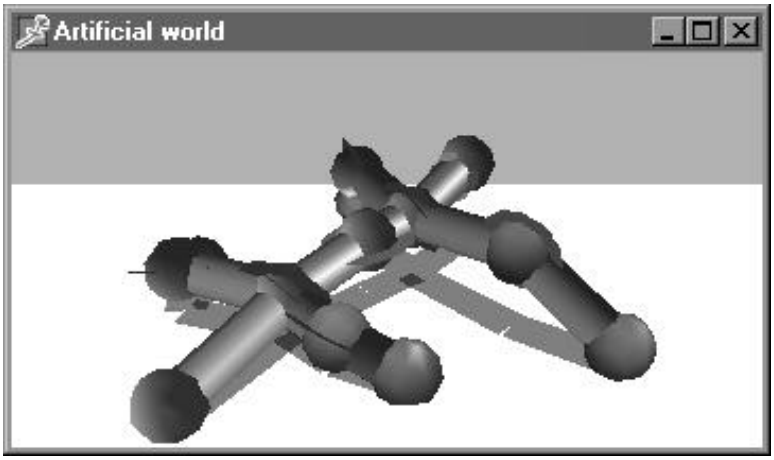

Figure 2. A sample Framsticks creature. This is a four legged land creature walking from the lower left of the figure to the upper right. The creature has a single sensor on its head.

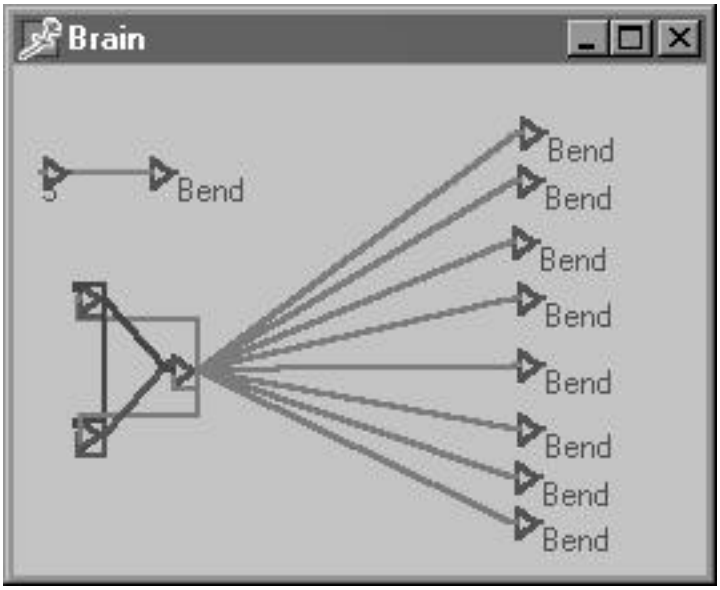

Figure 3. The nervous system of the creature depicted in figure 2. This nervous system is composed of two distinct parts. The smaller part is a stimulus orientation network that connects a single sensor to a steering muscle in the creature's torso. The larger portion is comprised of a three neuron central pattern generator that drives the bending muscles of the limbs, thus driving the creature's walking.

The arrangement and properties of sticks and neurons that comprise a single creature is determined entirely by the creature's genotype (no developmental factors are modeled). The genotype of the creature is a string of symbols. The symbol string may be hand coded by the user or modified by the evolutionary algorithm. A population of creatures is represented as a collection of symbol strings. Evaluation of creature fitness involves translating the genotype into a creature, and assessing the behavior of the creature in the world during the creature's lifetime. Fitness may be defined by the user in terms of weighted proportions of the following criteria: life span, horizontal velocity, horizontal distance, vertical position, vertical velocity, body size, and brain size.

The experiments described below involve four categories of mobile creatures. The first category—The Creatures of Pure Will—contains creatures that have no sensory inputs and the remaining categories all involve creatures with sensory inputs. The second category –The Creatures of Pure Vision—contains creatures that directly perceive certain environmental properties. The remaining categories, in contrast, have to compute or infer the presence of environmental properties based on comparatively degraded sensory input. The

third category—The Historians—contains creatures that employ a memory mechanism that allows the comparison between a current stimulus and a remembered stimulus. The fourth category—The Scanners—contains creatures that infer or compute the locations of environmental properties based on a comparison of sensory representations of the environment and representations of the states of their own bodies and actions. The scanners thus employ a form of action oriented representation as described in Mandik 1999.

7. Creatures of Pure Will: Procedural Representation

The projects of synthetic psychology (Braitenberg 1984) and synthetic neuroethology (Mandik 2002, Cliff 1998, Beer 1990), count among their goals to determine what the simplest possible systems are that exhibit phenomena interestingly considered as mental. A closely related question is the one of what the most primitive forms of life to exhibit mentality are. The assumption shared by all investigators is that the target systems will need to be capable of movement. Motile organisms as simple as euglena are thus more plausible candidates for mentality than sessile organisms as complex as oak trees.

The initial investigations of the neural bases of the sustenance and modulation of locomotion must come to terms with complexities completely ignored by the synthetic psychology informing Braitenberg's vehicle designs. In Braitenberg's vehicles, the locus of propulsion is conceived of simply as *motors*, black box devices on the posterior of the vehicle that might be implemented by powered wheels, propellers or turbines. The control networks for any Braitenberg vehicle need only send some level of activation or other to the motors. But when we turn to consider how biological locomotion is accomplished, we quickly realize the neural network controllers will have more to do than simply relay a signal to the motors with the content equivalent of "full steam ahead". Natural instances of motile organisms do not have motors that can be simply turned on or off. Instead, forward propulsion is maintained by some repetitive action: swimming animals must repetitively flagellate a fin or tail, walking and crawling animals must repetitively move their legs, and flying animals must repetitively beat their wings. Sending repetitive signals to the relevant muscles is thus one of the major tasks of neural control structures. One hypothesized class of neural mechanisms thought to generate such signals are known as Central Pattern Generators (CPGs) (Eliasmith and Anderson 2000).[5] CPGs are thought to be typically instantiated in neural networks as sets of reciprocally connected neurons. The simplest CPG would consist of a single, self-connected neuron: a neuron whose sole input is a self directed output as depicted in Figure 4a. More complex CPGs would include additional neurons and/or additional connections. Figure 4b shows a CPG with two neurons and two connections and figure 4c shows a CPG with two neurons and four connections.

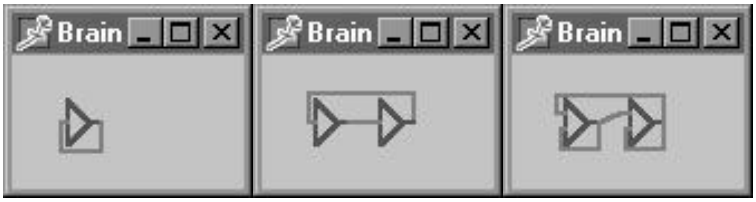

Figure 4. Figures 4a, 4b, and 4c depict three central pattern generator networks of increasing complexity. They are, respectively, a single neuron with a single connection, two neurons with two connections, and two neurons with four connections.

One hypothesized advantage of more complex CPGs is that they allow for the creation of a more complex command signal that is better suited to the dynamics of the creature's body in motion. That is, they allow for a more appropriate motor representation of the ideal configuration of bodily motions that will propel the creature forward. Note that the hypothesized representations output by the CPGs are conceived of here as representations with imperative contents, contents with success conditions instead of truth conditions.

Designing the topology of such networks by hand is a relatively simple task, but specifying the connections weights that will give rise to the oscillating mutual excitation required to generated a repetitive command signal is considerably more daunting. The challenge is not merely to create a repetitive oscillation, but further, one suited to the musculo-skeletal configuration of the motor organs and rest of the creature's body. Fortunately, the evolutionary algorithm in the Framsticks software allows for an automated solution to this problem. Creatures with the connection weights in their CPGs initially set to zero can be evolved to have weights optimized to generate a command signal suited to the repetitive motion of their limbs. The following experiment involves a comparison of the evolved performance of CPGs of varying complexity.

The body of the creature used in this experiment—a two-legged land creature—is depicted in figure 5. Three kinds of creatures were compared, each different kind had one of the three different kinds of central pattern generators depicted in figure 4. The creature's bodies and neural topologies were designed by hand, with the connection weights in the central pattern generators initially set to zero. The creatures were subjected to an evolutionary scenario in which fitness was defined as horizontal distance and mutations were allowed to only the neural network connection weights. Five populations of each kind of creature were evolved for 200 million steps of the simulation.

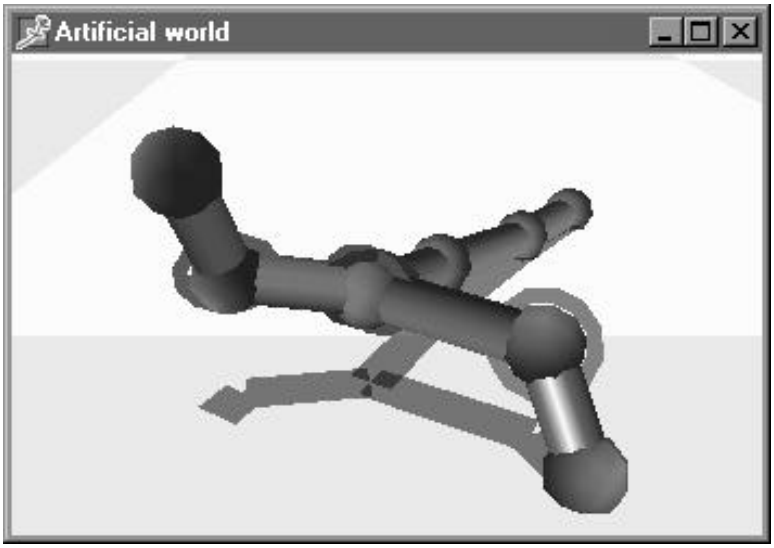

Figure 5. The two-legged land creature used to compare the performance of central pattern generators with the topologies depicted in figure 4.

The results of the experiment are depicted in the graph in Figure 6: creatures with more complex central pattern generators out performed creatures with less complex central pattern generators. The results support the representational hypothesis mentioned above: the ability to create more complex imperative representations enhanced the networks' ability to sustain the creature's motion.

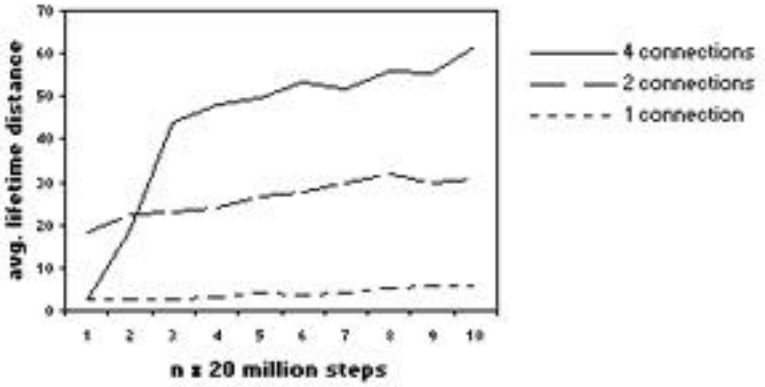

Figure 6. Results of the experiment comparing the performance of creatures with central pattern generators of varying complexities.

While the evolutionary scenario modeled here is highly constrained, this simulation illuminates the plausibility of evolving central pattern generators in less constrained, more realistic evolutionary scenarios. Additionally, it helps to see how our first variety of representation—motor imperative or procedural representations—might be the products of evolutionary processes. Note that the possibility sketched here is the possibility of evolving imperative representations in the absence of any sensory input, that is, in the absence of representations with any indicative declarative contents. The philosopher Ruth Millikan (1996) has argued that such a thing is impossible, claiming instead that representations with only imperative contents cannot exist without there first being representations that combine both imperative and declarative contents. I will return to this topic later. For now, let us move to consider what would be involved in introducing indicative representations into the Framsticks creatures. It is time now to turn attention to slightly more complex creatures and consider the addition of sensory inputs.

8. Creatures of pure vision: Sensory representation

The simulations discussed in this and the next couple of sections all involve creatures that have sensory inputs sensitive to the presence of food in the environment. This allows us to consider the next level of complexity in our exploration of the simplest neural control structures that will support minimally cognitive behaviors. The artificial creatures described below will utilize sensory inputs to exhibit both taxis and kinesis. Taxis and kinesis are commonly distinguished as follows. Taxis involves motion toward or away from some stimulus, as in, for instance, positive phototaxis, the motion toward a light source. Kinesis, in contrast, is not as sensitive to the location and heading of the stimulus, but is instead motion that is either triggered by or

suppressed by a stimulus. An example of kinesis would be if an animal ran around at random within a certain temperature range and stopped moving when outside of that range (Hale and Margham 1991).

For a simple example of how positive taxis can be modeled within Framsticks, consider the following creature, the 4 legged food finder designed and evolved by Miron Sadziak. Figure 7 shows the body of the creature, figure 8 shows the creature's neural network.

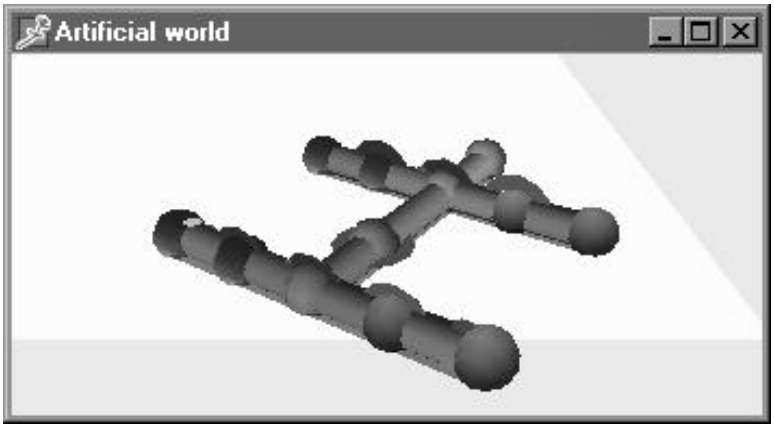

Figure 7. Miron Sadziak's four legged food finder.

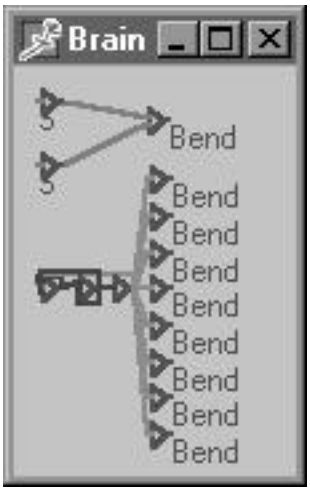

Figure 8. The nervous system of Miron Sadziak's four legged food finder.

Note that there are two distinct portions of the creature's nervous system. One part contains a central pattern generator that drives limb muscles for walking. The other part is a stimulus orientation network consisting of two sensors[6] feeding into a single bending muscle in the torso of the creature that controls the creature's steering. The stimulus orientation network functions so that if the activity in the right sensor is higher than the left, due to a food source being closer to the right sensor, the steering muscle will guide the creature toward the right, and conversely for a food source on the left. Driving taxis with a two sensor system has some relatively obvious biological validity, as in the photo taxis exhibited by caterpillars. (Rachlin 1976: 125-126).

Miron Sadziak's four legged food finder was largely hand designed. The question arises of how evolveable such a solution might be within the Framsticks software environment. Much of the trick involves the correct specification of fitness, since the software does not have a specific fitness criterion for amount of food found. One seemingly obvious approach would be to provide food-rich environments and define fitness in terms of life span. Creatures are born into the world with a finite store of energy and they die when their store reaches zero. Their life span may be extended indefinitely if they replenish their store with an indefinite quantity of food. However, selecting for life span turns out not to be an optimal way of evolving food finders because in many evolutionary runs sessile solutions like growing roots are often favored over motile solutions of going to the food. On casual

experimentation I have found that selecting for distance is a more reliable means of evolving food finders. Selecting for distance not only gets the creatures moving in the first place, but increasing the distances they traverse before they starve to death requires that they increase their likelihood of finding a meal along the way.

There is a relatively straightforward sense in which creatures that find food through the use of a pair of sensors have neural states that represent two dimensions of spatial location of the stimulus in an egocentric reference frame. The activation in a single sensor indicates one-dimension of spatial location: how near or far the stimulus is from the sensor and thus from the creature. The addition of the second sensor allows for the representation of a second dimension of spatial information: in addition to near or far, right and left are added to the mix. (I postpone momentarily discussing how a third dimension might be added). A walking creature or one swimming in relatively shallow water is essentially confined to a two dimensional world and a two sensor stimulus orientation system thus allows the creature to represent the (egocentric) location of food items in that world. A creature with only a single sensor is at a comparative disadvantage, since it will be incapable of telling whether a given stimulus is, say, five feet to the left or five feet to the right. However, there might still be some advantage to representing one dimension of stimulus location as opposed to none at all. Single sensor creatures may perhaps not have a genuine form of taxis (although this assumption will be subjected to further scrutiny in sections 9 and 10) but may nonetheless make use of it for a form of kinesis: the creature may scurry about randomly until it is close enough to the food to absorb it. Being able to detect a single dimension of proximity can thus allow the creature to stop long enough to enjoy the meal.

Elsewhere (Mandik 2002), I describe a Framstick experiment, the results of which confirm the above hypothesis that representing two dimensions is better than representing one which is itself better than none at all. I evolved creatures in conditions similar to those described in the CPG experiment described above. I hand designed bodies and neural topologies for legged land creatures. Fitness was defined as horizontal distance. Additionally, food items were randomly distributed throughout the environment. I compared three kinds of creatures: creatures with two sensors, creatures with one sensor, and creatures with no sensors. Five populations of each kind of creature were evolved for 200 million steps of the simulation. As expected, two sensor creatures performed better than one sensor creatures which in turn performed better than creatures with no sensors (Mandik 2002: 26-27).

There are interesting parallels between the performance and neural networks employed by these single sensor Framstick creatures and the real life example of the nematode worm *C. Elegans*. *C. Elegans* utilizes chemoreceptors to navigate up nutrient gradients. However, even though *C. Elegans* has more than a single chemoreceptor, these organs are thought to be too close together to give rise to meaningful differences of activation within the very diffuse gradients that the worms navigate (Pierce-Shimomura et al. 1999: 9557). They

are thus, for all sakes and purposes, single receptor creatures. Pierce-Shimomura et al. (1999) also note the pirouette motions that C. Elegans make in nutrient gradients and Morse et al. (1998) have modeled similar pirouette behaviors with a single sensor robot. I have observed that the single sensor Framsticks creatures exhibit similar motions.

The Framsticks software allows for the modeling of swimming creatures as well as walking creatures. Simulating food finders in a deep water environment allows one to utilize the evolution of taxis to address the issue of representing three dimensions of the spatial location of a stimulus: near-far, left-right, and up-down. A swimming creature able to represent two dimensions of stimulus location is depicted in figure 9 and its nervous systems is in figure 10.

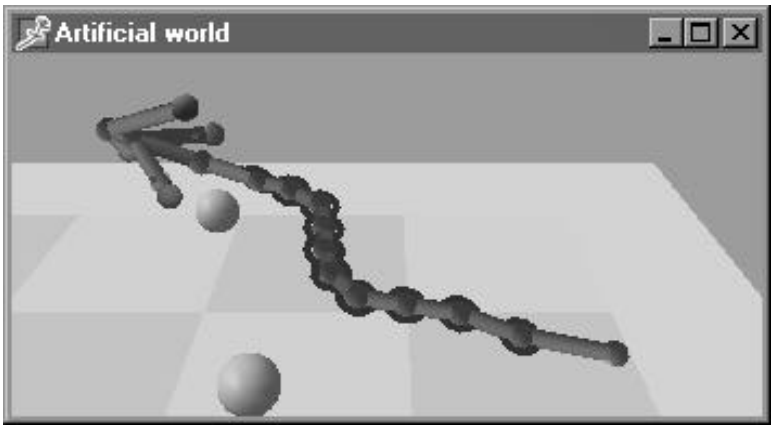

Figure 9. A 2-D food finder: a creature capable of detecting the location of a stimulus source within two spatial dimensions.

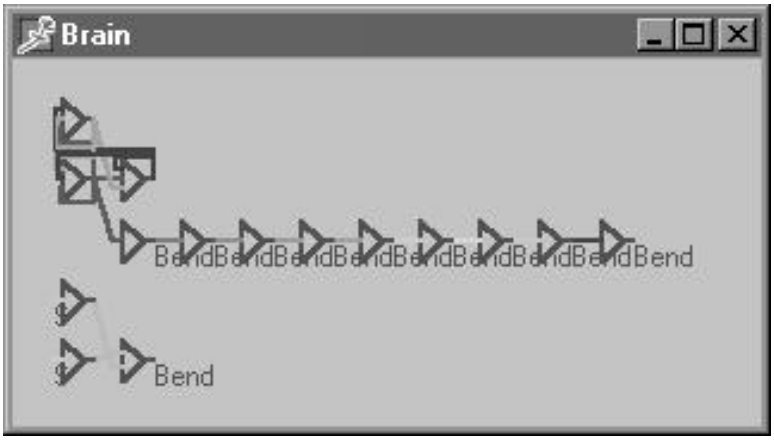

Figure 10. The nervous system of the 2-D food finder depicted in figure 9. A central pattern generator drives a chain of muscles in the creature's tail. A pair of sensors and a single left-right steering muscle constitute the creature's stimulus orientation network.

Swimming is achieved by flagellation of a tail which is in turn achieved by the sinusoidal activation of a chain of muscles driven by a central pattern generator. This means of sinusoidal swimming is oft hypothesized in models of, for example, lamprey locomotion (Ijspeert et al. 1999). This Framsticks creature has two smell sensors on its left and right, which feed into a steering muscle. Although this creature is virtually flawless in its ability to locate food in shallow water environments, in deep water it cannot tell whether a given food source is above it, below it, or in between. One possible way that one might attempt to endow a swimming creature with the ability to represent all three spatial dimensions is by giving it a second two-sensor orientation network mounted perpendicular to the first one. Such a four sensor creature would have top and bottom sensors to drive an up-down steering muscle in addition to left and right sensors corresponding to a left-right steering muscle. However, I have run Framsticks simulations to prove that a four sensor system is not the minimal way to achieve the perception of three dimensions of stimulus location: the feat may be accomplished with only 3 sensors. I hypothesized that a configuration of three sensors, one on the top and two on the bottom left and right would be sufficient for finding food in three dimensions. I created the creature "Trishark" to implement this idea.[7] Trishark's general body style is the same as depicted in figure 9. Trishark's nervous system is depicted in figure 11. Trishark's stimulus orientation system consists of three smell sensors, a hidden layer of 4 reciprocally connected interneurons, and two steering muscles (up-down and left-right). The connection weights in the stimulus orientation network were developed in an evolutionary scenario in which food was

present, fitness was defined in terms of lifetime distance and mutations were allowed to only the connection weights.

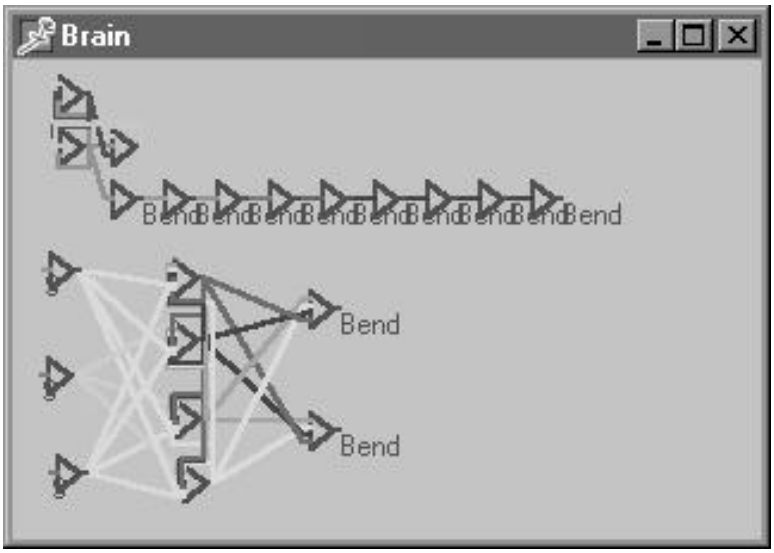

Figure 11. The nervous system of Trishark, a swimming creature with a body of the same style as depicted in figure 9. Trishark's stimulus orientation network, depicted in the bottom portion of the figure, consists of three sensors, a four neuron hidden layer with lateral and self-connections in addition to sensory inputs, and a two muscle steering system (containing a left-right muscle and an up-down muscle).

The graph in figure 12 shows the comparative performance of the two sensor and three sensor food finders in varying depths of water. The 2-D food finder excels in water depths between 1 and 7, but its performance dips far below the performance of the 3-D food finder in depths of 8 or greater.

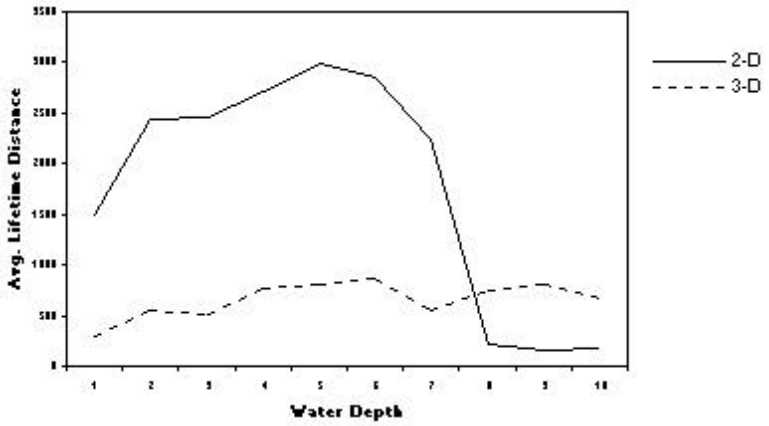

Figure 12. Results of the experiment comparing the performance of 2-D and 3-D food finders in water of varying depths.

It is worth noting the inference based on differential performance used here. It is not the case that the creature that represents more is, across the board, more fit or better adapted than the creature that represents less. This is especially evident in the comparison of performance between the 2-D and 3-D food finder in the shallow water environments. There are, nonetheless, demonstrable effects on the performance of the creature's different representational abilities, as is clear regarding their performance in the deep water environments.

So far we have seen three kinds of creatures with sensory inputs concerning the egocentric spatial location of food: single sensor creatures capable of representing one dimension of the spatial location of the distal stimulus, two sensor creatures capable of representing two dimensions of spatial location, and three sensor creatures capable of representing three dimensions of spatial location. Based on this observation, one might be tempted to accept the following hypothesis: the minimal number of sensors for the representation of N dimensions of the spatial location of the distal stimulus is N. However, the simulations described below show some interesting challenges to this hypothesis. In the simulations below, creatures utilize memory and active scanning of the environment to do with a single sensor what the above creatures required two sensors to accomplish. However, a modified version of the hypothesis equating number of dimensions represented with number of sensors will be true if restricted to creatures utilizing pure sensory representation and thus helps to serve to distinguish pure sensory representation—the creatures of pure vision—from the other kinds discussed below.

9. The historians: memorial representation

The creatures described in the previous section used relatively simple neural mechanisms to achieve positive taxis. I have been happy to attribute representational contents to certain neural states of these creatures. However, other philosophers might be more conservative with their attributions. For example, Dretske (1988) is skeptical of describing such mechanisms as genuinely representational, in large part because they do not exhibit learning. Whether or not Dretske's grounds for dismissing non-learning systems as non-representational are ultimately sound, refection on learning and memory does inspire the pursuit of implementing it in simple neural network controlled creatures. Another reason for seeking to implement memory in the simple creatures described here is that memory provides clearer instances of representation than the so-called pure sensory cases. As discussed in Mandik (2002) many would think it a requirement on representations that they represent the spatially and/or temporally remote. If this "remoteness constraint" is indeed a constraint on representation (and I am not saying that it is), then creatures utilizing memory offer instances that more obviously satisfy of the remoteness constraint than the kinds of creatures discussed in the previous section (Mandik 2002: 17).

Implementing memory in Framsticks requires overcoming several challenges. First, Framsticks allows no changes to a creature's topology or connection weights within a creature's lifetime, thus no kind of associative learning or Hebbian mechanism can be implemented. Further, the early version of Framsticks utilized here (version 1.78) allows very little control over the placement of creatures and environmental features. For example, creatures and food sources are placed in the world at random locations, so there is not enough environmental stability for creatures to learn something like where the food usually is.

In spite of the above mentioned challenges, there are certain aspects of memory that are relatively easy to model within Framsticks. If memory is conceived of encoding information about the past, then the construction of such networks should be relatively easy. As discussed in Mandik 2002, the simple recurrent networks that serve as central pattern generators might also be pressed into service as short term memory stores. Input to a recurrent network can trigger a series of oscillations that eventually decay, thus mirroring at lest one aspect of natural memory. However, this solves only a fraction of the memory problem. Memory involves three components: encoding, maintenance, and retrieval. The memory network envisioned so far only supplies a mechanism for encoding and maintenance. The challenge remains of supplying a mechanism of retrieval, that is, supplying a means whereby the creature is able to utilize the information that is stored. Additionally, there is the further challenge of finding a use for memory within the rather limited domain of food finding.

Below I describe a solution to these challenges that I have arrived upon. Before describing the solution, it is worth briefly describing the

solution's inspiration: some fascinating studies of memory in bacterial chemotaxis performed by Daniel Koshland (1977, 1980). The bacterium *E. Coli* is able navigate extraordinarily diffuse nutrient gradients. Due to the small size of the bacterium, it is incapable of making use of anything like the multiple sensor solutions described above. The difference between the concentrations of nutrient impinging the different sides of the creature are too small to give the creature any means of determining the direction of greatest concentration. The problem faced by the bacterium might be thought of as analogous to determining, by looking out a airplane window while flying through a dense cloudbank, what the direction of greatest cloud density is. Koshland hypothesized that the bacterium was making use of some kind of memory to solve the problem. To see how memory might be employed, think of the airplane example. Suppose that you took a Polaroid photograph of the fog outside of the window, and then waited a while. After waiting, you compare the photograph to the current perception of the fog. If the current perception is that the fog is lighter than the photographed scene, then you may infer that the plane is heading away from the center of concentration. However, if the perception is darker than the photographed scene, then the plane is heading deeper into the clouds. By comparing percept to memory (instantiated here as the external memory of the photograph) a moving creature can infer whether it is heading up or down a gradient. Koshland tested the hypothesis by placing the bacteria in different uniform concentrations and noting their change in direction. A bacterium placed in a higher uniform concentration than it was in previously will continue its heading, but if placed in a lower concentration will change its heading. In both cases it is evident that the bacterium is storing some record of past events, since how it acts in some particular environment is not determined solely by the current environment, but depends on what the difference between the present and the past is. Perhaps the mechanism employed by the bacterium is analogous to the one illustrated in the airplane example insofar as it involves a comparison of the current stimulus to a memory of a past stimulus.

The network in Figure 13 depicts an attempt to implement in a Framsticks creature an analogous mechanism for comparing present and past stimuli. The stimulus orientation network, like the network in a 2-Dimensional food finder, involves a steering muscle that receives a pair of inputs. However, the input to the network itself is only a single sensor. One of the inputs to the steering muscle is a direct connection to the single sensor. The second input to the steering muscle also comes from the single sensor, but the signal is passed through a memory buffer consisting of a chain of neurons. There is a slight delay between the receipt of a neuron's input and the discharge of its output. Thus, by increasing the number of neurons connected in serial, one introduces an increased delay of the signal transmitted across the channel.

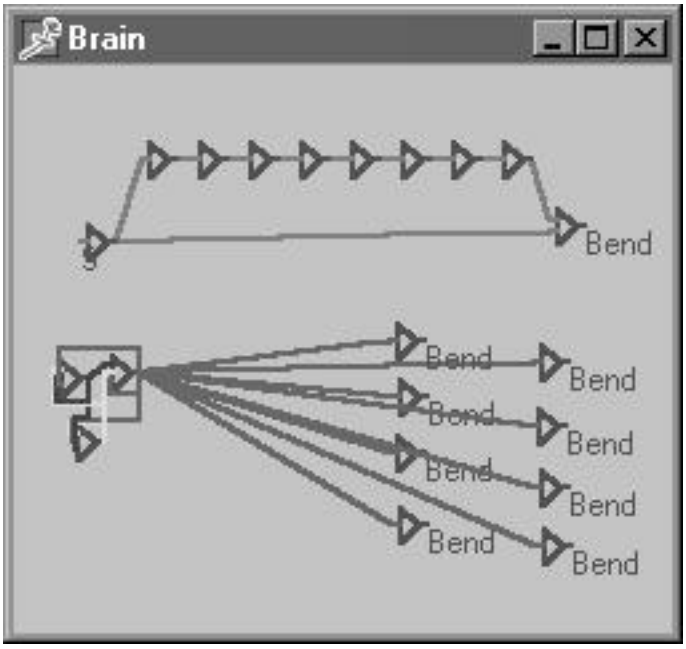

Figure 13. The nervous system of a creature that uses memory. The stimulus orientation network is depicted in the top portion of the figure. Information from a single sensor passes to a single steering muscle via two routes. The top route is a memory buffer containing eight neurons wired in serial. The second route is a direct and thus faster connection from the sensor to the steering muscle.

The stimulus orientation network depicted in figure 13 thus offers a means of implementing a memorial solution to taxis with a single sensor via a means similar to that hypothesized for *E. Coli*. An intriguing hypothesis is that through the use of a memory circuit, a creature with only a single sensor can use a comparison between the present and past to build up a representation of more than just a single dimension of spatial distance from the stimulus, and do with one sensor what the 2-D food finders were doing with two sensors: represent the egocentric location of the stimulus in two dimensions.

The Framsticks software allows two benefits in the pursuit of the truth of this hypothesis. First, it allows for an experimental test of whether such a scheme is feasible, and second, the evolutionary algorithm allows for the possibility of tuning the connection weights in such a way to allow for the information encoded to also be utilized by the behaving organisms. I describe several experiments conducted along these lines.

In the first experiment, four legged single sensor creatures (similar in body style to the creature depicted in figure 2) were divided into two groups, those with memory buffers and those without. The creatures with memory buffers had neural topologies as depicted in figure 13 whereas the creatures without memory buffers and nervous systems similar to that depicted in figure 3. Creatures with memory buffers had the weights of all the buffer neural weights set to an initial value of one, to guarantee that signals would be propagated through the buffer at the earliest stages to the simulation. The weights of the inputs to the steering muscle were 10.0 and –10.0 respectively. Creatures had pre-evolved central pattern generators, so at the beginning of the simulation they were already quite capable of forward locomotion. Five populations of each of the two groups were evolved for 200 million steps. Food was present, mutations were allowed to only neural weights, and fitness was defined as lifetime distance. Results are shown in the chart in figure 14. The results show a clear superiority of the creatures with memory over the creatures without. Whether the creatures are constructing a representation of the two dimensional location of the food is not entirely clear, but it is more clear that the evolved creatures are utilizing a representation of the past. They not only encode and maintain the memory record, they retrieve it as well.

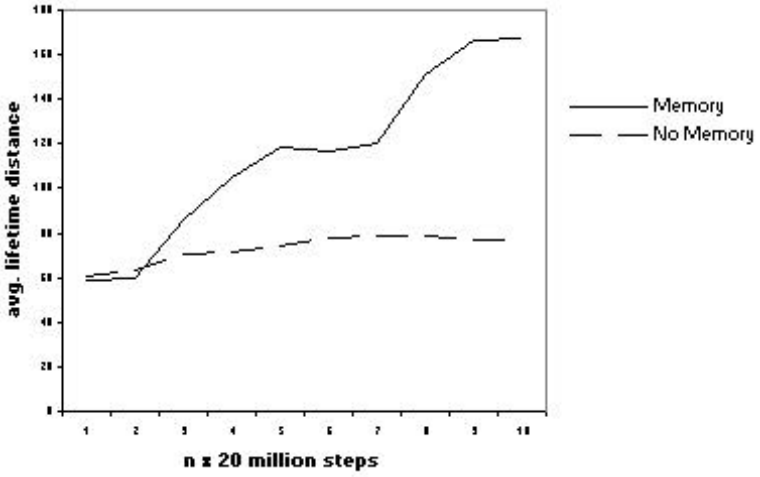

Figure 14. Results of comparison between creatures with memory and creatures without.

Despite the success of the first experiment, many questions remain unanswered and here I focus on two. The first question concerns whether the memory creatures are really utilizing a comparison between perception and memory or whether the introduction of the delayed signal is alone bearing the burden of their superior food finding.[8] To test this, I selected the best individual

from the last generation of the best population of creatures with memory buffers. By looking at the portion of the genotype of the evolved creature that coded for the connections and neural weights in the orientation network, I verified that all of the weights were non-zero, thus showing that the steering muscle would be receiving sensor information through both the direct route and the memory-delay route. As a further test, I subjected the creature to various lesions and compared the creature's performance in intact and lesioned conditions. The categories of lesioned creature were creatures with memory only, creatures with the direct (non delayed) sensory information only, and creatures with absolutely no sensory information arriving at the steering muscle. Intact and lesioned creatures were run for 4 million steps in a version of the simulation that disallowed mutations. As shown in the chart in figure 15, intact creatures out performed the various lesioned creatures. The results of the lesion study provide yet further evidence that the food finding ability was not achieved by mere reliance on the memory information, but involved a comparison between the remembered and current percept.

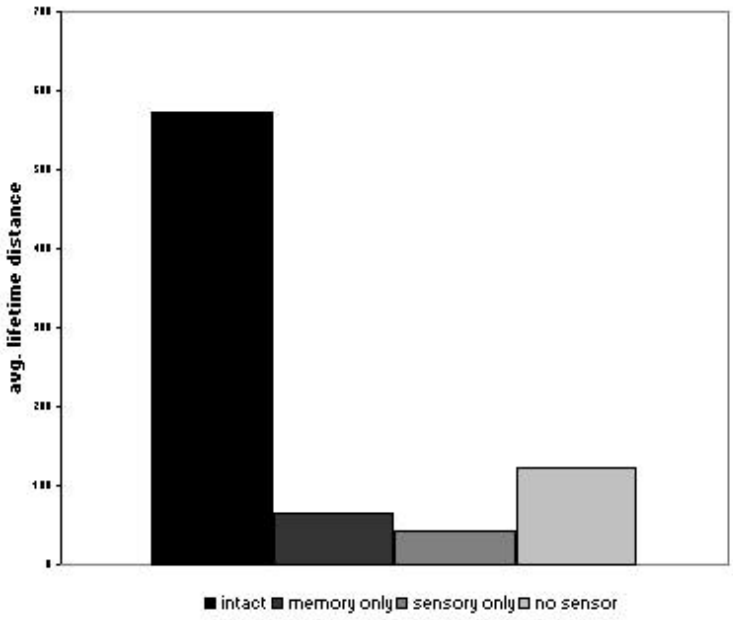

Figure 15. Results of the comparison of intact and lesioned memory utilizing creatures.

A second question raised by consideration of these experiments is whether creatures utilizing memory can be evolved without having the hand coding of their connection weights set them so close to the solution at the start.

That is, can memory evolve in conditions where buffer neuron weights are not set to one at the beginning? The next experiment addresses this. Creatures had all of the initial weights in their stimulus orientation networks—including both memory buffer neurons and connections to the steering muscle—set to zero. Results are shown in figure 16. As anticipated, creatures without any memory were inferior to those with memory. A lesion study analogous to the one described above was conducted yielding analogous results. These are presented in figure 17.

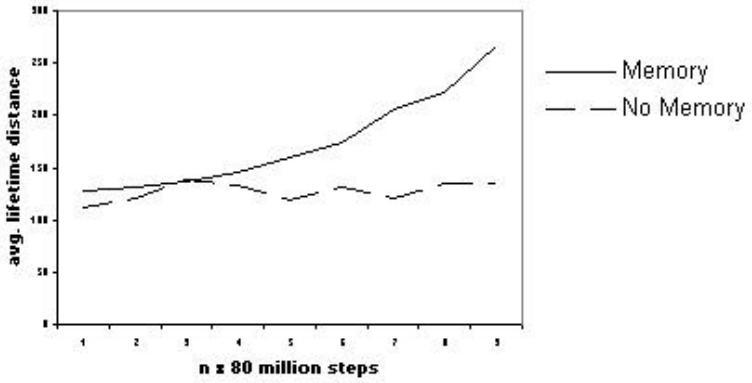

Figure 16. Results of memory experiment in which creature's initial memory buffer connections weights were set to zero.

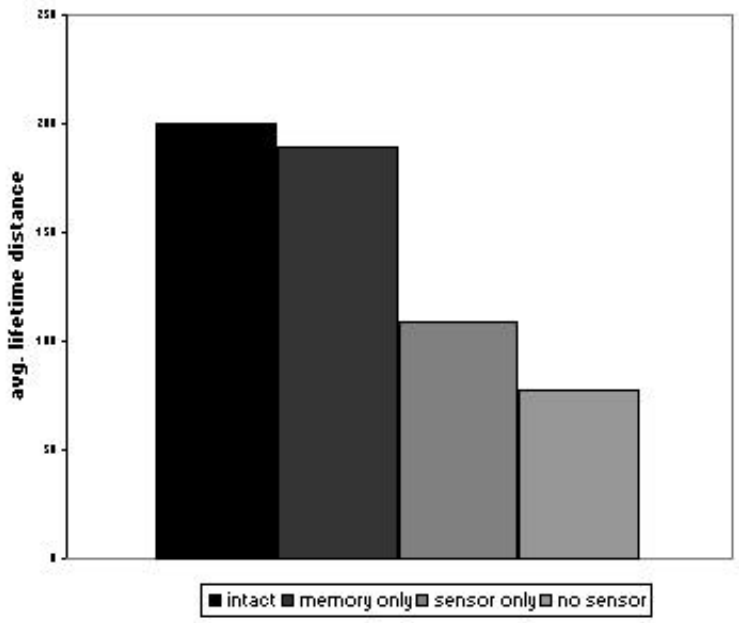

Figure 17. Results from the lesion study on the second batch of memory creatures.

Note the somewhat closer performance between the intact condition and the memory only condition. This result suggests that a somewhat heavier load is born by the memory portion of the orientation network. Nonetheless, two points remain. The first is that the use of memory seems to be superior to the cases lacking memory. The second is that the case in which both the memory and the percept are compared are superior to the memory only case.

It is worth mentioning that casual observation of the trained creatures reveals pirouette motions similar to those discussed above in connection with the gradient navigation of *C. Elegans* worms. This raises the possibility that the kinds of strategies employed by the nematodes involves a kind of memory similar to these Framstick creatures. Whether this suggestion bears fruit remains to be seen. But currently, this much remains clear: having information about the past can provide a demonstrable benefit in evolved creatures.

10. The scanners: action-oriented representation

In the previous experiments the limits of representing only a single dimension of spatial information in sensory input were overcome through the capacity to represent past as well as present events. In the current section I explore a different way in which these limits may be overcome. The creatures employed

in this next simulation had a single sensor mounted on a long limb that was used as an oscillating scanner. The creature "Radar" is depicted in figure 18. Radar's neural network is depicted in figure 19.

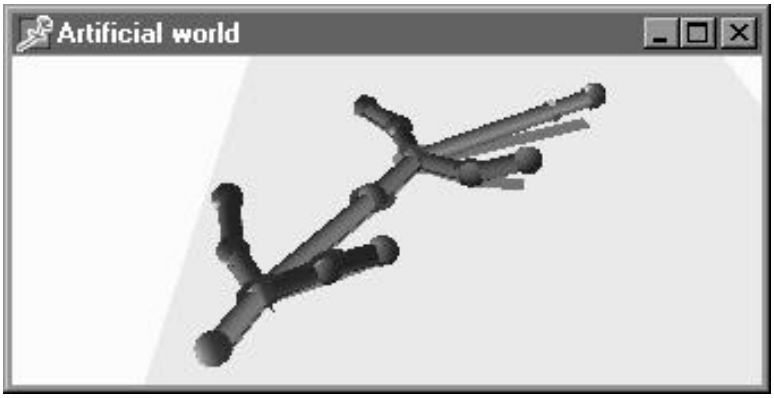

Figure 18. The creature Radar has a long scanning organ for a head that the creature utilizes to sweep a single sensor from side to side.

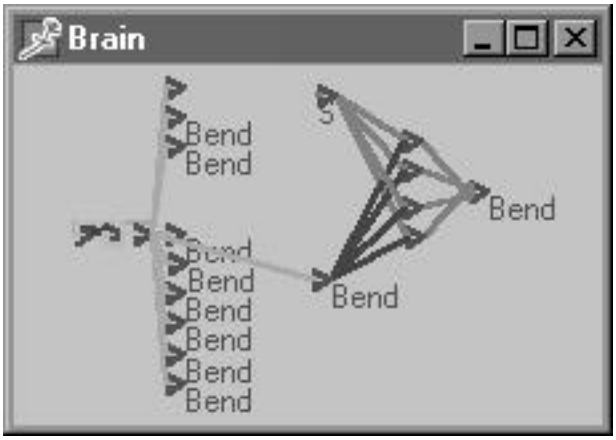

Figure 19. Radar's nervous system. The stimulus orientation system, depicted on the right portion of the figure, involves input from a single sensor and muscular information concerning the state of the scanning muscle. These two sources of information are connected to a four neuron hidden layer which is connection to the single steering muscle. The scanning muscle is driven by the same central pattern generator (on the far left) that drives all of the walking muscles.

Radar's stimulus orientation network receives as inputs the activity from a single smell sensor and feedback from the muscle that controls the scanning motion. This pair of inputs may, in theory, encode information about the two dimensional stimulus location in the following way. Sensor activity encodes proximity information, thus providing the first dimension of location. The second dimension—left-right—is achieved by a comparison between sensor state and muscle feedback. If sensor activity is high and the muscle is bending to the right, then the food is to the right. If sensor activity is high while the muscle is bending to the left, then the food is to the left. If sensor activity is low while bending to the right, the food is to the left, and if sensor activity is low while bending to the left, then food is to the right.

Another way in which a single sensor can be used to build up a two dimensional representation through scanning is by comparing sensor activity to an efference copy of the command signal sent to the scanning muscle, instead of comparing the sensor activity to muscular feedback. Even though the efference copy is an imperative (efferent) representation and the muscular feedback is a declarative (afferent) representation, they have overlapping representational content: both concern the bending of the scanning muscle. (This efference copy strategy of building up a representation of the two dimensional location of the stimulus helps give further credence to describing the outputs of the central pattern generators as representational in the first place, as discussed above in section 7.) Such a representational scheme thus implements the action oriented representations I discuss in Mandik 1999. In that paper, describe a thought experiment concerning a creature named "Tanky" that locomotes through the use of tank treads (1999: 53-55). I discussed two different ways in which Tanky could keep track of his location. The first was by sensory feedback counting the rotations of his tank treads. The second, and potentially equally reliable, method would be to keep track of the commands sent to the tank treads. This latter kind of solution constitutes the use of action oriented representations of Tanky's egocentric space. See Mandik 1999 for further discussion of the psychological and physiological evidence for the prevalence such action oriented representations in natural systems. My main concern here is to see how such representations might fit within the evolutionary context of my synthetic creatures.

To see if Framstick creatures could utilize action oriented representations, I conducted an experiment to compare the evolved performance of three different kinds of neural controllers for Radar's body. The first kind of controller—the feedback condition—had as inputs to the stimulus orientation network both sensory and muscular information as depicted in figure 19. The second kind of controller—the efference copy condition—had an efference copy instead of muscular feedback sent to the stimulus orientation network. The third—sensory only—condition had neither muscular feedback nor an efference copy but only information from the sensor sent to the stimulus orientation network. All three kinds of creatures were hand coded and pre-evolved to have fully functioning central pattern generators and

active scanning muscles. The initial connection weights of the stimulus orientation networks were set to zero. Five populations of each of the three kinds of creatures were evolved for 200 million steps in an evolutionary scenario containing food in which fitness was defined as horizontal distance and mutations were allowed to only the neural connection weights. Results are depicted in figure 20.

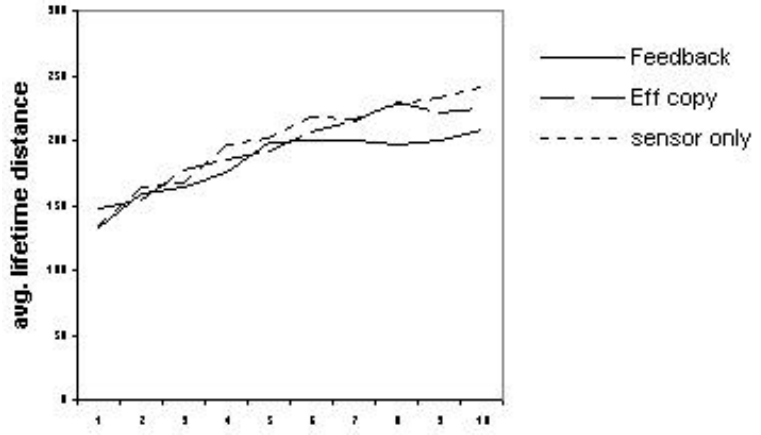

Figure 20. Results of experiment comparing different neural controllers for the scanners.

While the differences in performance between the different kinds of creatures were not as dramatic as the previous experiments discussed in this paper, there were still mild differences with the sensory only condition doing the best and the feedback condition doing the worst. While the performances of the creatures remained approximately the same, the question arises of whether the neural control strategies they evolved were approximately the same. One possibility is that each of the three conditions were relying on only the sensory information, like a one-dimensional food finder, and thus neither the so called feedback and efference copy conditions were using either muscular feedback or efference copies. A different possibility, and one that seems to be the correct one, is that the creatures actually were utilizing the efference copies or muscular feedback information that they had access to. One means that confirmed that this was indeed the case was by an inspection of the genotype similar to the inspection described for the memory creatures in the previous section. The two best creatures in both the efference copy and feedback condition (thus, four creatures altogether) were inspected and confirmed to have non-zero weights in connections leading to the steering muscle from both the sensor and the muscle feedback/efference copy connections. A second means of confirmation was by a lesion study similar to those conducted for the memory creatures. There were two lesioned conditions in this case. The first

left the sensory connections intact while depriving muscle feedback or efference copy. The second deprived the stimulus orientation networks of sensory information altogether. The results are shown in figure 21.

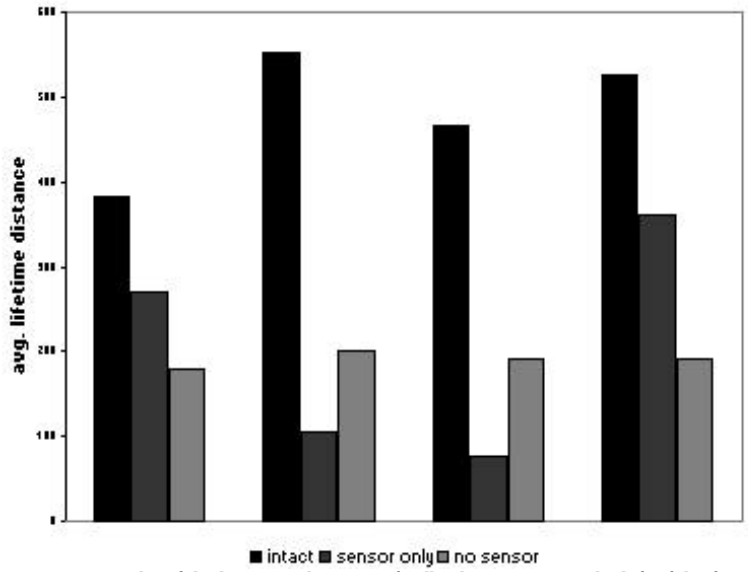

Figure 21. Results of the lesion study on two feedback scanners (on the left of the figure) and two efference copy scanners (on the right).

In all four of the lesion studies of the scanners, the intact creatures outperformed the lesioned creatures, thus showing the these scanners were not relying on a strategy identical to that employed by the sensory only condition. That is, these scanners did indeed evolve to make use of the muscular feedback or efference copies in their stimulus orientation networks. Further, the performance of the creatures utilizing the efference copies seems not significantly worse, and may perhaps even be better than the performance of the creatures utilizing direct muscular feedback. This result shows the evoveability in synthetic creatures of the kinds of action oriented representation discussed above and in Mandik 1999.

11. Discussion

It is time now to take stock of the varieties of representation in the evolved and embodied neural networks. In so doing, it is crucial to make note of both the vehicles and the contents of the representations, that is, to say both what the representations *are*, and what the representations are representations *of*. What

are the representations in the Framstick creatures? There are several varieties, although in each case they are the states of activation in neurons and the corresponding signals sent from one neuron to the next. What are the representations representations of? Again, there are several varieties for the several varieties of creatures. In the Creatures of Pure Will, the representations are the motor commands issued from the central pattern generators and what they are representations *of* are patterns of muscular movement. In the Creatures of Pure Vision, the representations are states of sets of sensory transducer neurons and the signals those neurons passed to orientation muscles. What those representations are representations *of* are the current egocentric locations of food sources in one, two, and three dimensions. The Historians had some of the same kinds of representations as the Creatures of Pure Vision and additionally had representations in their memory buffers that were memories of *past* egocentric locations of food sources. The Scanners similarly had some of the same kinds of representations as the Creatures of Pure Vision and, in some cases, combined these with the sorts of representations highlighted in the discussion of the Creatures of Pure Will. The Scanners utilizing efference copies of command signals combined those signals with sensory signals to arrive at action oriented representations of the egocentric location of food.

With these varieties of representation in view, and further, in view within the context of the causal economies of entire organisms and their evolutionary histories, we can begin to see how these simulations shed light on the synchronic and diachronic concerns of neurosemantics with special attention paid to the economy problem. I begin here by considering a sketch of a naturalization of representation that emerges from consideration of these Framsticks creatures.

First, let us make note of how far a teleological informational (teleo-informational) account goes with regard to these creatures. On a teleo-informational account, a representation is a state of an organism that has the function of carrying information about some thing, that is, being caused by something (Dretske 1995). On a teleo-informational account then, what a representation is a representation *of*, is certain causal antecedents of the representational state. Thus the paradigm instance of a representational state on this account would be a percept. This account has obvious application to at least some of the Framstick creatures. For example, in the Creatures of Pure Vision, the states of activation in ensembles of sensory transducers had the function of carrying information about the egocentric spatial location of the food source. Carrying this information is the teleological *function* of these states because carrying this information has been survival conducive to this creature and the creature's ancestors.

Let us turn to see how well the teleo-informational account applies to the memories of the Historians. Here I think the account fares quite well: the signals that constitute the outputs of the memory buffer have the function of carrying information about past events, events that happened some significant increment of time earlier than the what the sensor currently represents. Note

how this account is *not* an instance of the overly literal empiricism lampooned earlier in the allegory of the punch cards and the boxes. What is *not* happening in the memory creatures is that items in the sensor receive the stamp of approval that bestows sensory content and then gets sent off to a memory module while retaining that mark. Instead, the causal relations that give sensory states their contents are different from the causal relations that give memory states their contents.

Informational accounts of representation are often viewed as opposed to isomorphism based accounts, accounts in which a representation does its work in virtue of resembling that which it represents (Cummins 1996). There may seem to be a certain applicability of the isomorphism accounts to the Framstick creatures. For example, the relations between greater and lesser degrees of activity in the sensor neurons are isomorphic to the relations of closer or more distant locations of the food source. However, the isomorphisms here are entirely consistent with the application of the informational account. Indeed, not only are they consistent, but the relevant isomorphisms are integral to the way that information is carried and processed in the networks. This is in keeping with the view of information that I have elsewhere described as

> 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 litterally inFORMed by the boot (Mandik 2002: 4-25. Compare Cummins 1989: 2-4.)

While the teleo-information account is pretty successful in capturing the notions of representation applied in descriptions of the Creatures of Pure Vision and the Historians, the account is inadequate to account for the cases of the Creatures of Pure Will and the Scanners. For these latter creatures, crucial representational states are those that represent their causal consequences, not their causal antecedents. What a command is a representation of is that which would occur if the command were obeyed, and thus, this is something that happens after and because of the issuing of the command. The teleo-informational account thus may supply sufficient conditions for representation, but not necessary conditions. However, the account may be appended to account for procedural (effector) representations as well as informational (affector) representations by specifying an additional set of sufficient conditions for being a representation: a state of an organism is a representation of some thing if it has the function of causing that thing (Mandik 1999, p. 52). Lumping together effective and affective causal relations under the heading of causal covariation yields a formulation of representation that handles both informational and procedural representations: "it is sufficient for X to represent Y that X has the function of being causally related to Y (alternately: causally covarying with Y)" (Mandik 2001: 190).

I close this section with some remarks concerning some of the issues raised earlier under the heading of diachronic questions of neurosemantics. These questions all concern where representations came from, how they evolved, what came first, and what came last. There are two points worth noting along these lines. The first concerns the temporal priority of procedural over indicative representations. The second concerns the temporal priority of egocentric over allocentric representations.

As touched on briefly above, Millikan has argued that representations that combine imperative and indicative contents, so called "Pushmi Pulyu Representations" or PPRs, are more primitive than—that is, must be evolved before—representations that have only imperative content or only indicative content (Millikan 1996). Her arguments for these priority claims are quite brief. I find Millikan's argument that indicative representations cannot precede PPRs more persuasive than her argument against the priority of imperative representations. Millikan's argument against the priority of indicative representations seems to be that an indicative representation cannot come into being (because of the theoretical weight she places on evolution by natural selection) unless it has some effect on behavior, and it can have no effect on behavior unless it can combine with some representation that tells you what to do. The closest Millikan comes to arguing against the priority of imperative representations is simply to state "of course, representations that tell what to do have no utility unless they can combine with representations of facts" (Millikan 1996). The Framtick creatures described above offer no counter examples to Millikan's claim about indicative representation being unable to evolve unless there are first representation with imperative content. However, I offer the Creatures of Pure Will as counterexamples to Millikan's claim about imperative representations. Contra Millikan, representations with only imperative content may be evolved in the total absence of representations with indicative content. The Creatures of Pure Will do not have states with indicative content of the location of food, since they have no sensors. Neither are the Creatures of Pure Will detecting the states of their own muscles, since the connections to muscles are strictly feed-forward. The only representations then are the outputs of the central pattern generators, messages that do not indicate, but only command.

A second issue concerning the temporal priority in the evolution of the varieties of representation concerns a contrast between egocentric ("self-centered") representation and allocentric ("other-centered") representation. All of the varieties of representation instantiated in the Framstick creatures discussed above are egocentric representations: what they are representations of concern in each case either something *in* the creature (as in muscular feedback) or something *in relation to* the creature (as in representing a food source to the right or left). Egocentric representations are thus perspectival for, unlike allocentric representations, they do not abstract away from the perspective of the representing subject. In the causal covariational account of representation, egocentric (perspectival) representations are defined as follows:

> A subject S has a perspectival representation R of X if (but maybe not only if) R has the function of causally covarying with X and relations Z_1-Z_n S bears to X (Mandik 2001: 191).

For example, then, the sensory representational states in the Creatures of Pure Vision do not simply represent food the way that the proverbial /cow/ simply represents cows. Instead, the creature's sensory states represent the location of the food source in relation to the creature itself in virtue of causally covarying with both the food and the spatial relation the food bears to the creature. Insofar as these Framsticks creatures count among the most primitive evolveable instances of representing subjects, they lend credence to the view that egocentric representations are more primitive than allocentric representations. An allocentric representation would be one has no contents about how things stand in relation to the representing subject. So, for instance, you may acquire the belief that Neptune has uranium in its core without thereby representing anything about the various relations you may bear to Neptune or uranium. I currently have very little clue as to what it would take to evolve such allocentric representations in Framstick creatures, but it certainly appears more difficult, and thus, less primitive than egocentric representations.[9]

12. Objections and replies

I want to briefly consider two objections to the utility of the above computer simulations in addressing the central neurophilosophical concerns of representation.

Are the current simulations too constrained? One concern with the current simulations that I take very seriously is that the conditions were so highly constrained. The representational systems that evolved were not evolved from scratch but from pre-designed, and in some cases, pre-evolved creatures. Further mutations were allowed to only neural weights. More realistic would be less constrained scenarios, scenarios in which, for instance, mutations were allowed to neural topologies and/or body structures. While I admit that such simulations would be good, I will not admit that this thereby renders the current simulations useless. This is especially clear if the goal of the imagined less constrained simulations would be to look for the varieties of representation sketched here. The current simulations can act as guides, making it easier to recognize the varieties of representation once we turn to look for them "in the wild". Further, the current simulations have helped to suggest what sorts of parameters might be useful in less constrained scenarios. For example, the evolution of food finders in the current simulators seemed best achieved by defining fitness in terms of life time distance. Such a fitness parameter may be similarly useful in less constrained future simulations.

The current simulations are mere *simulations.* This kind of objection comes in two flavors, one that I take seriously and another that I do not. The flavor that I take seriously is that simulations abstract away from the real phenomena in ways that may leave out crucial features. This is a real danger,

but it must be noted that it is not a danger peculiar to computer models. All theorizing and all modeling must necessarily abstract away from the thing in its self. The object of study is presumably indefinitely complex but our descriptions and models cannot do justice to this indefinite complexity. We hope instead that our simplifications leave out only what is inessential, but there is always a risk of getting it wrong. I have done the best to focus on what is essential to representational phenomena in neural networks. Whether I have failed will not be settled by merely pointing out that I am dealing with *mere* computer simulations. This leads to the flavor of the objection to simulations that I have considerably less respect for. On this version of the objection, nothing that goes on in a computer simulation is really *real*, but instead some mere *virtual* and thereby *fictional* process. My sympathies on this issue are so well summarized by someone else, that I will simply quote them at length. The artificial life researcher Bruce MacClennan, in commenting on this sort of objection to his artificial life simulations of the emergence of communication writes:

> The objection may still be made that any communication that might take place is at best simulated. After all, nothing that takes place in the computer is real, the argument goes; no one gets wet from a hurricane in a computer. To counter this objection I would like to suggest a different way of looking at computers. We are accustomed to thinking of computers as abstract symbol-manipulating machines, realizations of universal Turing machines. I want to suggest that we think of computers as programmable mass-energy manipulators. The point is that the state of the computer is embodied in the distribution of real matter and energy, and that this matter and energy is redistributed under the control of the program. In effect, the program defines the laws of nature that hold within the computer. Suppose a program defines laws that permit (real!) mass-energy structures to form, stabilize, reproduce, and evolve in the computer. If these structures satisfy the formal conditions of life, then they are real life, not simulated life, since they are composed of real matter and energy. Thus the computer may be a real niche for real artificial life-not carbon-based, but electron-based. Similarly, if through signaling processes these structures promote their own and their group's persistence, then it is real, not simulated, communication that is occurring." (1991: 638).

I see the same being applicable to the Framstick creatures. The Framsticks creatures are patterns of energy that really exist in the computer. They really have evolved, they really do survive, they really have environments and they really have the capacity to represent features of their environments.

13. Conclusion

I have attempted to shed light on the issues of representation by constructing and evolving simple neural networks in simple creatures. Many have used similar artificial life work to argue *against* attributing representations to such simulated organisms. For example, Randall Beer writes of his experiments on the nervous systems of synthetic insects

> 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. (1990: 162-163).

Elsewhere (Mandik 2002) I consider arguments of antirepresentationalists such as Beer and find them wanting. I do not wish to recount my negative arguments from the previous work here. May aim here has been instead one of continuing the positive line of thought in favor of attributing representational states to the neural controllers of the evolved Framsticks organisms. I have tried to show how the use of representation talk is in keeping with a relatively widespread pre-naturalized conception of mental representation. Further, I have tried to show how the attributions constitute items in empirically predictive discourse: attributing representational states to the varieties of Framstick creatures served to both predict and explain their behavior.[10]

In this paper I have set out sketch the minimal requirements for building micro-minds out of a small number of components. Even if my little critters do not help to accomplish the philosophical work advertised—like, for instance, shedding light on the economy problem for theories of mental representation—I think they should nonetheless be of interest to philosophers of neuroscience, for they present novel opportunities to see the functioning of an entire neural network in an entire organism. We have seen the whole iguana. We have also seen its brain, if not its mind.

Acknowledgements

This work was supported in part by grants from the McDonnell Project in Philosophy and the Neurosciences as well as William Paterson University. Early versions of the current work were presented as talks. For valuable feedback I thank the audiences of the City University of New York Graduate Center Cognitive Science Symposium and Discussion Group, The Franklin and Marshall College Scientific and Philosophical Studies of Mind Program, and

the William Paterson University College of Humanities and Scoical Sciences Faculty Research Seminar. For especially helpful comments and discussion I give special thanks to Tony Chemero, Patrick Grim, Carrie Figdor, Shawn Gaston, Emily Mahon, Doug Meehan, Roblin Meeks, Jim Moor, David Rosenthal, Eric Steinhart, Jonathan Waskan, Josh Weisberg, and Alison Wylie.

Notes

[1] Although in section 9 we will see challenges to this later assumption.

[2] Akins (1996) advances a non-representational view of neural function. In contrast, Mandik (2002) suggests that creatures evolved nervous systems in order to represent and process information.

[3] Such as Fodor (1998), Dretske (1995), Lycan (1996), and Tye (1995).

[4] As discussed, for instance, in Fodor and Lepore (1992).

[5] Note that I am in no means claiming that central pattern generation, that is, purely endogenously initiated command signals, is the only means of creating and sustaining creature locomotion. Of course there is ample evidence of varying degrees of sensory input involved in the maintenance of locomotion. My intention on focusing on pure central pattern generated motion is to get a chemically pure sample of one of several varieties of representation: procedural representations—and show how they might be arrived at in evolutionary scenarios.

[6] Food sources located in the environment emit a gradient that the sensors are responsive to. Activity in the sensor corresponds to the sensor's position in the gradient. The creators of the Framsticks software label the sensors "smell sensors" thus making any taxis modeled a form of chemotaxis. However, the gradients and sensors can just as easily be interpreted as optical, thus making any taxis modeled a form of phototaxis. The preference for calling the creatures described in this section "Creatures of Pure Vision" as opposed to "Creatures of Pure Smell" or "Little Sniffers" is purely poetic.

[7] Trishark is available for free download from the Framsticks web site (Komosinski 2001) as part of the package of the latest version of the Framsticks software.

[8] I thank to Emily Mahon for bringing this concern to my attention.

[9] I am especially grateful to Tony Chemero for a stimulating discussion of the relative priority of egocentric (subjective) and allocentric (objective) representations. Much of what the both of us think in this regard, and what it may or may not imply for metaphysics and epistemology more generally, is aired in public in Mandik and Clark (in press).

[10] For an excellent discussion of the differenential empirical strengths of representationalism and antirepresentationalism, see Chemero (2000).

References

Akins, K.: 1996, 'Of sensory systems and the 'aboutness' of mental states', *The Journal Of Philosophy* 93, 337-372.

Anscombe, G.E.M.: 1957, *Intention*. Cornell University Press, Ithaca.

Beer, R.: 1990, *Intelligence as Adaptive Behavior*, Academic Press, San Diego, CA.

Braitenberg, V.: 1984, *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA.

Brooks, R.: 1999, *Cambrian Intelligence*, MIT Press, Cambridge, MA.

Brooks, R.: 1991, 'Intelligence Without Representation', *Artificial Intelligence* 47, 139-159.

Chemero, A.: 2000, 'Anti-Representationalism and the Dynamical Stance', *Philosophy of Science* 67, 625-647.

Clark, A.: 1997, *Being There: Putting Brain, Body and World Together Again*, MIT Press, Cambridge, MA.

Cliff, D.: 1998, 'Computational Neuroethology, in M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge, MA, pp. 626-630.

Cummins, R.: 1996, *Representations, Targets, and Attitudes*. MIT Press, Cambridge, MA.

Dennett, D.: 1987, *The Intentional Stance*. MIT Press, Cambridge, MA.

Dennett, D.: 1998, *Brainchildren*. MIT Press, Cambridge, MA.

Dretske, F.**:** 1988, *Explaining Behavior*. MIT Press, Cambridge, MA.

Dretske, F.: 1995, *Naturalizing the Mind.* MIT Press, Cambridge, MA.

Eliasmith, C. and Anderson, C: 2000, 'Rethinking central pattern generators: A general framework', *Neurocomputing*. 32-33(1-4): 735-740 .

Fodor, J.: 1975, *The Language of Thought.* Cambridge, MA: Harvard University Press.

Fodor, J.: 1990, *A Theory of Content and Other Essays*. MIT Press, Cambridge, MA.

Fodor, J. 1998, *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press,

Fodor, J. and LePore, E.: 1992, *Holism: A Shopper's Guide.* MIT Press, Cambridge, MA.

Guillot, A. and Meyer, J.A.: 2001, 'The Animat Contribution to Cognitive Systems Research', In *Journal of Cognitive Systems Research.* 2(2), 157-165.

Hale, W. and Margham, J.: 1991, *The HarperCollins Dictionary of Biology*. New York: HarperCollins.

Ijspeert A.J., Hallam J. and Willshaw D.: 1999, 'Evolving swimming controllers for a simulated lamprey with inspiration from neurobiology', *Adaptive Behavior* 7:2, , pp 151-172.

Komosinski, M.: 2000, 'The World of Framsticks: Simulation, Evolution, Interaction'. In: *Proceedings of 2nd International Conference on Virtual Worlds*, Paris, France . Springer-Verlag (LNAI 1834), 214-224.

Komosinski, M.: 2001, Framsticks Website. http://www.frams.poznan.pl.

Koshland, D.: 1977, 'A response regulator model in a simple sensory system. *Science* 196: 1055-1063.

Koshland, D.: 1980, 'Bacterial chemotaxis in relation to neurobiology, in *Annual Review of Neurosciences 3*, ed. by Cowan, W. C. et al, Annual Reviews, Inc., Palo Alto, 1980 pp. 43-75.

Lycan, W.: 1996, *Consciousness and Experience*. MIT Press, Cambridge, MA.

MacLennan, B.: 1991, 'Synthetic Ethology: An Approach to the Study of Communication" *Artificial Life II: Studies in the Sciences of Complexity*, vol X, edited by C.G. Langton, C. Taylor, J.D. Farmer, & S. Rasmussen, Addison-Wesley.

Mandik, P.: 1999, 'Qualia, Space, and Control'. *Philosophical Psychology* 12 (1): 47-60.

Mandik, P.: 2001, 'Mental Representation and the Subjectivity of Consciousness'. *Philosophical Psychology* 14 (2): 179-202.

Mandik, P.: 2002, 'Synthetic Neuroethology', *Metaphilosophy* 33, 11-29.

Mandik, P. and Clark, A. (in press). Selective representing and world-making. *Minds and Machines.*

Millikan, R.: 1996, 'Pushmi-pullyu Representations', in May, L., Friedman, M. and Clark, A. (eds.), *Minds and Morals,* MIT Press, Cambridge, MA., pp. 145-161.

Millikan, R.: 1984, *Language, Thought, and Other Biological Categories.* MIT Press, Cambridge, MA.

Millikan, R.: 1993, *White Queen Psychology and Other Essays for Alice.* MIT Press, Cambridge, MA.

Morse, T.M., Ferree, T.C., and Lockery, S.R.: 1998, 'Robust spatial navigation in a robot inspired by chemotaxis in *C. elegans*.' *Adaptive Behavior.* 6:393-410.

Pierce-Shimomura, J.T., Morse, T.M., and Lockery, S.R.: 1999, 'The fundamental role of pirouettes in C. elegans chemotaxis', *Journal of Neuroscience* 19:9557-9569.

Rachlin, H.: 1976, *Behavior and Learning*. San Francisco: Freeman.

Searle, J.: 1983, *Intentionality: An Essay in the Philosophy of Mind*. New York, Cambridge University Press.

Taylor, T. and Massey, C.: 2001, 'Recent Developments in the Evolution of Morphologies and Controllers for Physically Simulated Creatures'. *Artificial Life* 7 (1), 77-87.

Tye, M.: 1995, *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press, Cambridge, MA.

Walter, G.: 1963, *The Living Brain*, W. W. Norton, New York.