Chapter 7: Uphill Analysis, Downhill Synthesis?

7.1 INTRODUCTION

The previous chapter used the construction, observation, and analysis of toy robots to provide a concrete example of the three basic steps that are required in synthetic psychology. The first step is the synthesis of a working system from a set of architectural components. The second step is the study of this system at work, looking in particular for emergent properties. The third step is the analysis of these properties, with the goal of explaining their origin. This general approach was given the acronym SEA, for synthesis, emergence, and analysis.

The demonstration project that was presented in Chapter 6 provides a concrete example of these three basic steps, but is not by its very nature a particularly good example of synthetic psychology. In terms of advancing our introduction of synthetic psychology, the "thoughtless walkers" that we discussed at best raise some important issues that need to be addressed in more detail. These issues were mentioned near the end of Chapter 6.

The purpose of the current chapter is to go beyond our toy robots to consider two related issues in more detail. First, we are going to be concerned with the attraction of the synthetic approach. Why might a researcher choose it instead of adopting the more common analytic approach? Second, we are going to consider claims about the kind of theory that the synthetic approach will produce. Specifically, one putative attraction of the synthetic approach is that theories that emerge from synthetic research are considerably less complex than those that are generated from analytic research. The theme of this chapter will be that the synthetic approach does offer an attractive perspective for explaining complex behaviors. However, it is not an approach that necessarily produces theories that are simpler than those that come from analytic research. Indeed, synthetic research depends heavily upon analysis if its goal is to explain, and not merely produce, emergent phenomena.

This chapter adopts an historical context to explore these issues concerning the relationship between synthetic and analytic traditions. Starting from an example from early research in cybernetics, the chapter will introduce some of the pioneering work on autonomous robots from the early 1950s. Then, the chapter will briefly describe a rebirth – of sorts – of this work in the early 1980s. In reviewing this research, we will see several examples of simple devices that produce behavior that is both intricate and interesting. But we will also become aware that even the researchers who constructed these devices did not have an easy task in explaining their performance.

7.2 FROM HOMEOSTATS TO TORTOISES

In the early stages of the Second World War, it was realized that advances in aviation technology needed to be met in kind by advances in anti-aircraft artillery. Specifically, the speed and maneuverability of German aircraft were such that classical methods of aiming this artillery were obsolete. New techniques for aiming – techniques that were capable of predicting the future position of a targeted plane, and sending a projectile to this predicted position – had to be developed, and had to be built right into artillery controlling mechanisms (Wiener, 1948).

One of the scientists who worked on this applied problem was Norbert Wiener (b. 1894, d. 1964), who had received his PhD in mathematical philosophy from Harvard when he was only 18, studied at Cambridge under Russell, and eventually became a professor in the mathematics department at MIT. Wiener realized that feedback was a key factor in designing a mechanism for aiming anti-aircraft artillery. For example, “when we desire a motion to follow a given pattern the difference between this pattern and the actually performed motion is used as a new input to cause the part regulated to move in such a way as to bring its motion closer to that given by the pattern” (Wiener, 1948, p. 6). Wiener also realized that processes like feedback were central to a core of problems involving communication, control, and statistical mechanics. He provided a unifying mathematical framework for studying these problems, and this framework defined a new discipline that Wiener called cybernetics, which was derived from the Greek word for “steersman” or “governor”. “In choosing this term, we wish to recognize that the first significant paper on feedback mechanisms is an article on governors, which was published by Clerk Maxwell in 1868” (p. 11).

7.2.1 Feedback And Machines

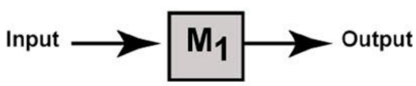

Figure 7-1. Simple definition of a machine.

A more definite understanding of feedback, and its relationship to synthetic psychology, begins with a very general definition of a machine (Ashby, 1956). William Ross Ashby (b. 1903, d. 1972) was one of the pioneering figures for the field of cybernetics, and was director of research at Barnwood House Hospital in Gloucester, and later was the director of the Burden Neurological Institute in the Department of Electrical Engineering at the University of Illinois, Urbana. For Ashby, a machine is simply a device which, when given a particular input, generates a corresponding output. In other words, a machine is a device that performs a transformation of an input signal to an output response. Figure 7-1 illustrates this simple and general definition of a machine.

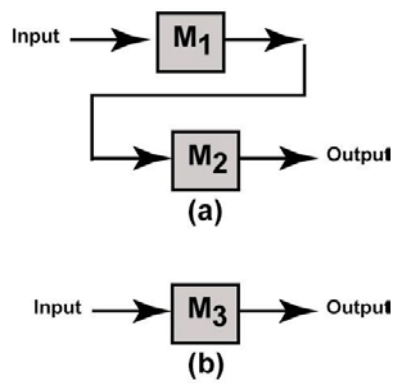

Figure 7-2. Simple coupling of two machines.

When a machine is defined in this way, then one can easily imagine a situation in which two machines are coupled together. In the simplest case, this is accomplished by having the output of one machine serve as the input to a second machine. With this kind of coupling, the behavior of the second machine is completely determined by the behavior of the first machine. For example, in Figure 7-2a the behavior of machine $\mathbf{M_2}$ is completely determined by the behavior of machine $\mathbf{M_1}$. This means that considering the machines separately does not really provide any additional insight into the function that transforms the input into the output. We could replace the two machines with a single machine ($\mathbf{M_3}$) that maintained the same input/output relationship, as is shown in Figure 7-2b. We saw this kind of relationship earlier in the book when we discussed the linear nature of regression equations, and noted that the behavior of the entire regression equation was exactly equal to the sum of its parts.

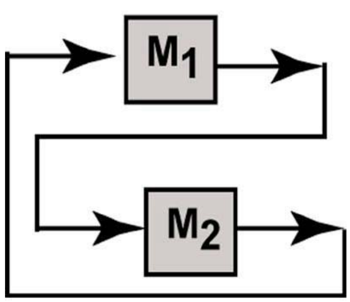

Figure 7-3. Feedback between two machines

A more complicated relationship between machines occurs with a different kind of coupling. The straightforward behavior of the two machines in Figure 7-2 occurred because the inputs of machine $\mathbf{M_1}$ were independent of the outputs of machine $\mathbf{M_2}$. If the output of $\mathbf{M_2}$ is fed backwards to serve as the new input to $\mathbf{M_1}$, then much more complicated behavior will result. The relationship between machines that is illustrated in Figure 7-3 is the basic sort of feedback that is of central interest to cybernetics. At one level of description, the “mechanical feedback” that was described in our analyses of the thoughtless walkers in Chapter

6 is of this type: the forces generated by the robot (machine $\mathbf{M_1}$) are transmitted to the surface ($\mathbf{M_2}$), which in turn transmits forces back to the robot.

Descriptions of feedback need not be limited to pairs of machines. Many more machines may be coupled together to create a more complicated system. Of particular interest to Ashby was a system of four different machines coupled together with feedback, as is shown in Figure 7-4. To foreshadow observations that we will be making later in this chapter about whether synthetic theories are simple or not, Ashby (1956, p.54) makes the following observation about a system of this complexity: "When there are only two parts joined so that each affects the other, the properties of the feedback give important and useful information about the properties of the whole. But when the parts rise to even as few as four, if every one affects the other three, then twenty circuits can be traced through them; and knowing the properties of all the twenty circuits does *not* give complete information about the system."

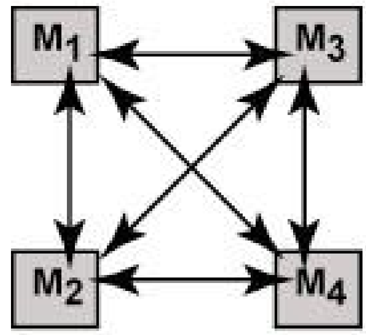

Figure 7-4. The double-headed arrows indicate mutual feedback relationships in a system of four different machines.

7.3.1 Ashby's Homeostat

Imagine if a researcher was interested in studying a system like the one illustrated in Figure 7-4. If understanding its twenty component circuits cannot provide complete information about the system, then how should the research proceed? Ashby (1960) provided a decidedly synthetic answer to this question by constructing an interesting system that he called the homeostat to study the properties of feedback amongst four mutually coupled machines.

7.3.1.1 Basic Design

The homeostat was a system of four identical components, and is illustrated in Figure 7-5. The input to each component was an electrical current, and the output of each component was also an electrical current. The purpose of each component was to transform the input current into the output current. This was accomplished by using the input current to change the position of a pivoted magnet mounted on the top of the component. In essence, each magnet could rotate a needle back and forth. The needle was connected to a wire that was dipped into a trough of water through which another constant electric current was passed. With this physical arrangement, it was possible for the component to output an electrical current that was approximately proportional to the needle's deviation from its central position. All things being equal, a large current that was input to the component would cause a large deflection of the magnet (and needle), which in turn would result in a proportionately large current being output from the component.

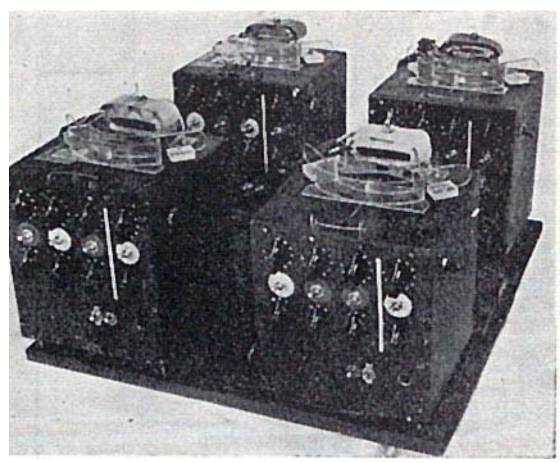

Figure 7-5. Ashby's (1960) photograph of the homeostat. Each of the four units is in the picture; the water trough and pivoting magnet is on top of each unit. (I will need to get the rights to use this picture!)

The four units were coupled together to create a system of the type that was drawn in Figure 7-4. Specifically, the electrical current that was input to one unit was the sum of the electrical currents that was output by each of the other three units, after each of these three currents was passed through a potentiometer. The purpose of the potentiometer was to determine what fraction of an input current would be passed on to deflect the magnet, and thus each potentiometer was analogous to a connection weight in a PDP network. The result of this interconnectedness was a dynamic system that was subject to a great deal of feedback. "As soon as the system is switched on, the magnets are moved by the currents from the other units, but these movements change the currents, which modify the movements, and so on" (Ashby, 1960, p. 102).

In order to dictate the influence of one unit upon another in the homeostat, one could set the resistance value of each potentiometer by hand. However, Ashby (1960) used a different approach to allow the homeostat to automatically manipulate its potentiometers. Each unit was equipped with 25-valued uniselector or stepping switch. Each value that was entered in the uniselector was a potentiometer setting that was assigned randomly. A unit's uniselector was driven by the unit's output via the deflected needle. If the output current was below a predetermined threshold level, the uniselector did not activate, and the potentiometer value was unchanged. However, if the output current exceeded the threshold, the uniselector activated, and advanced to change the potentiometer's setting to the next stored random resistance. With four units, and a 25-valued uniselector in each, there were 390,625 different combinations of potentiometer settings that could be explored by the device.

In general, then, the homeostat was a device that monitored its own internal stability (i.e., the amount of current being generated by each of its four component devices). If subjected to external forces, such as an experimenter moving one of its four needles by hand, then this internal stability was disrupted and the homeostat was moved into a higher energy, less stable state. When this happened, the homeostat would modify the internal connections between its component units by advancing one or more of its uniselectors to modify its potentiometer settings. The modified potentiometer settings enabled the homeostat to return to a low energy, stable state. The homeostat was "like a fireside cat or dog which only stirs when disturbed, and then methodically finds a comfortable position and goes to sleep again" (Grey Walter, 1963, p. 123).

7.3.1.2 Behavior Of The Homeostat

The homeostat was tested by placing some of its components under the direct control of the experimenter, manipulating these components, and observing the changes in the system as a whole. For example, in a simple situation only two of the four components might be tested (Ashby, 1960, Figure 8/4/1). In this kind of study, the feedback being studied was of the type $\mathbf{M_1} \leftrightarrow \mathbf{M_2}$. The relation $\mathbf{M_1} \rightarrow \mathbf{M_2}$ could be placed under the control of the experimenter by manipulating the potentiometer of $\mathbf{M_1}$ by hand instead of using its uniselector. The reverse relationship $\mathbf{M_2} \rightarrow \mathbf{M_1}$ was placed under machine control by allowing the uniselector of $\mathbf{M_2}$ to control its potentiometer. After starting up the homeostat and allowing it to stabilize, Ashby manipulated $\mathbf{M_1}$ to produce instability. The result was one or more advances by the uniselector of $\mathbf{M_2}$, which resulted in stability being re-attained.

Even with this fairly simple pattern of feedback amongst four component devices, many surprising emergent behaviors were observed. For example, in one interesting study Ashby (1960) demonstrated that the system was capable of a simple kind of learning. In this experiment, it was decided that one machine ($\mathbf{M_3}$) was to be controlled by the experimenter as a method of "punishing" the homeostat for an incorrect response. In particular, if $\mathbf{M_1}$'s needle was forced by hand to move in one direction, and the homeostat did not respond by moving the needle of $\mathbf{M_2}$ to move in the opposite direction, then the experimenter would force the needle of $\mathbf{M_3}$ into an extreme position to introduce instability. On the first trial of this study, when the needle of $\mathbf{M_1}$ was moved, the needle of $\mathbf{M_2}$ moved in the same direction. The homeostat was then punished, and uniselector-driven changes ensued. On the next trial, the same behavior was observed and punished; several more uniselector-driven changes ensued. After these changes had occurred, movement of $\mathbf{M_1}$'s needle resulted in the needle of $\mathbf{M_2}$ moving in the desired direction – the homeostat had learned the correct response. "In general, then, we may identify the behavior of the animal in 'training' with that of the ultrastable system adapting to another system of fixed characteristics. Ashby went on to demonstrate that the homeostat was also capable of adapting to two different environments that were alternated.

7.3.1.3 Implications

The homeostat counts, perhaps, as one of the earliest examples of the synthetic approach in action. It was a fairly simple analog device, constructed from well-understood component machines. It was wired up in such a way that complex feedback could be established among these components, and was used to study the dynamic processes that resulted. It had the advantage of permitting these processes to be studied at a time when a mathematical account of the device was not well established, and also at a time when computer simulations of this kind of feedback were not really possible. It demonstrated emergent behaviors, including interesting kinds of learning. Ashby (1960) was quite interested in drawing parallels between the behaviors of the homeostat and behaviors of the nervous system and entire organisms, although he was also aware of many limitations in his machine.

The interesting behavior of the homeostat arises from two general sources. The first is the rich possibilities of interactions between machines, as defined by the feedback relationships that were wired into the device. The second comes from the relatively large number of internal states that could be adopted by the machine when its uniselectors were used to modify potentiometer settings.

As a prelude to one theme that will be developed in more detail later in this chapter, the large number of different internal states that are available to a working homeostat provides the machine with many degrees of freedom with which to produce a low energy state. However, these same degrees of freedom make it difficult for the experimenter to explain the specific mechanisms that the homeostat uses to achieve this state. “A very curious and impressive fact about it, however, is that, although the machine is man-made, the experimenter cannot tell at any moment exactly what the machine’s circuit is without ‘killing’ it and dissecting out the ‘nervous system’ – that is, switching off the current and tracing out the wires to the relays” (Grey Walter, 1963, p. 124). In other words, it is much easier to produce interesting behavior in the homeostat than it is to explain this behavior.

7.3.2 Grey Walter’s Tortoises

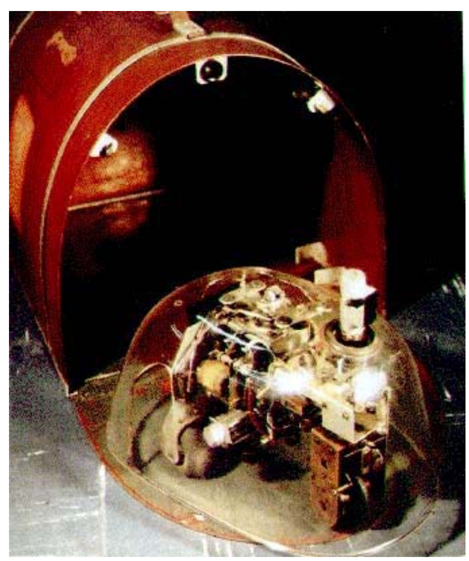

Figure 7-6. The tortoise Elsie emerges from her hutch. (I need permission for this image)

Ashby’s (1960) homeostat could be interpreted as supporting the claim that the complexity of the behavior of whole organisms largely emerges from a) a large number of internal components and from b) the interactions between these components. In the late 1940s, William Grey Walter (b. 1910, d. 1977) built some of the first autonomous robots to investigate a counter-claim (Grey Walter, 1950, 1951, 1963). His research program “held promise of demonstrating, or at least testing the validity of, the theory that multiplicity of units is not so much responsible for the elaboration of cerebral functions, as the richness of their interconnection” (Grey Walter, 1963, p. 125). His goal was to use a very small number of components to create robots that generated much more life-like behavior than that exhibited by Ashby’s homeostat. Grey Walter was a neurophysiologist who conducted pure and applied research at a variety of London hospitals from 1935 to 1939, and at the Burden Neurological Institute in Bristol from 1939 to 1970. While our interest in his research is with his robotics work, he was also a pioneer in the use of the electroencephalogram, and was the discoverer of theta and delta waves. His EEG research and his robotics work are both described in his 1963 text *The Living Brain*.

7.3.2.1 Basic Design

Grey Walter (1963) whimsically gave his autonomous robots the biological classification *Machina speculatrix* because of their propensity to explore the environment. (He gave Ashby's (1960) homeostat the classification *Machina sopora*, pointing out that if it were to be judged "entirely by its behavior, the naturalist would classify it as a plant" (p. 124).) Because of their appearance, which is illustrated in Figure 7- 6, his robots were more generally called tortoises. A very small number of components (two miniature tubes, two relays, two condensers, two motors, and two batteries) were used to create two sense reflexes. One reflex altered the behavior of the tortoise in response to light. The other reflex altered the behavior of the tortoise in response to touch.

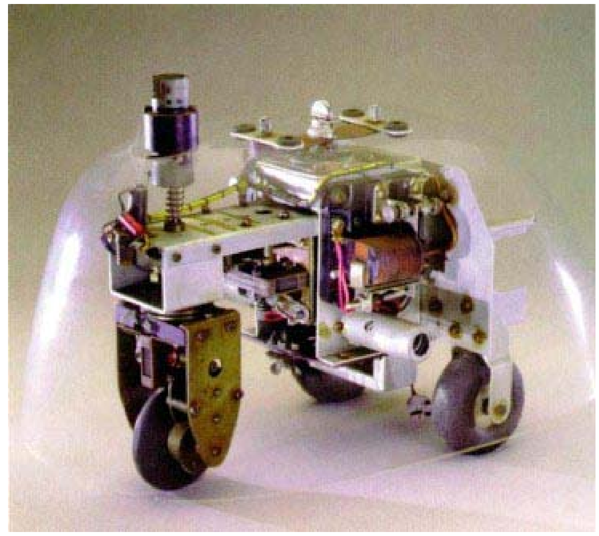

Figure 7-7. A picture of a copy of Elsie that is currently at the Smithsonian.

At a general level, a tortoise was an autonomous motorized tricycle (see Figure 7-7). One motor was used to rotate the two rear wheels forward. The other motor was used to steer the front wheel. The behavior of these two motors was under the control of two different sensing devices. The first was a photoelectric cell that was mounted on the front of the steering column, and which always pointed in the direction that the front wheel pointed. The other was an electrical contact that served as a touch sensor. This contact was closed whenever the transparent shell that surrounded the rest of the robot encountered an obstacle.

Of a tortoise's two reflexes, the light-sensitive one was the more complex. In conditions of low light or darkness, the machine was wired in such a way that its rear motor would propel the robot forward while the steering motor slowly turned the front wheel. As a result, the machine could be described as exploring its environment. The purpose of this exploration was to detect light -- when moderate light was detected by the photoelectric cell, the steering motor stopped. As a result, the robot moved forward, approaching the source of the light. However, if the light source were too bright, then the steering motor would be turned on again at twice the speed that was used during the robot's exploration of the environment. As a result, "the creature abruptly sheers away and seeks a more gentle climate. If there is a single light source, the machine circles around it in a complex path of advance and withdrawal" (Grey Walter, 1950, p. 44).

The touch reflex that was built into a tortoise was wired up in such a way that when it was activated, any signal from the photoelectric cell was ignored. When the tortoise's shell encountered an obstacle, an oscillating signal was generated that rhythmically caused both motors to run at full power, turn off, and to run at full power again. As a result, "all stimuli are ignored and its gait is transformed into a succession of butts, withdrawals and sidesteps until the interference is either pushed aside or circumvented. The oscillations persist for abut a second after the obstacle has been left behind; during this short memory of frustration Elmer darts off and gives the danger area a wide berth" (Grey Walter, 1950, p. 45).

7.3.2.2 Behavior

(Grey Walter, 1950, 1963) built two tortoises, and named them Elsie and Elmer using the initials of the terms that described them – "Electro Mechanical Robots, Light-Sensitive, with Internal and External stability." The question of interest to him was whether the intricate relationships between the small number of robot components, and the interactions between the robots and their environment, would be sufficient to generate complicated and interesting behaviors. He attempted to answer this question by observing the actions of the robots, together and separately, in a number of different environments. He mounted a light source on the robots, and recorded their behavior using time-lapse photography. As a result, the trajectory of a tortoise was traced out on the photograph by the light. The behavior that he observed was "remarkably unpredictable" (Grey Walter, 1950, p. 44).

For example, consider the behavior recorded in Figure 7-8. When this experiment was started, the light was hidden from view by an obstacle. As a result, Elsie began with its exploratory motion. As a result of this exploration, Elsie collided with the obstacle, which produced the avoidance behavior. Because of the movements taken to avoid the obstacle, the robot was able to detect the light. It approached the light, but circled it, because when it came too close to the light it was too bright, and caused the robot to veer away. "Thus the machine can avoid the fate of the moth in the candle" (Grey Walter, 1963, p. 128).

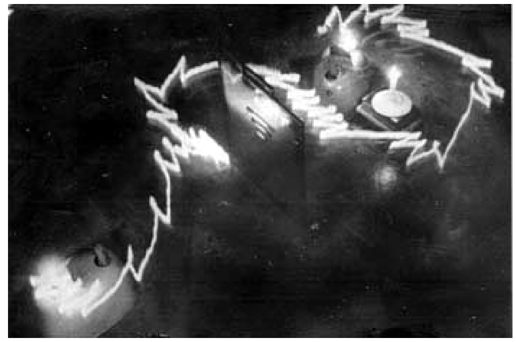

Figure 7-8. Elsie's behaviour in an experiment with a light and an obstacle.

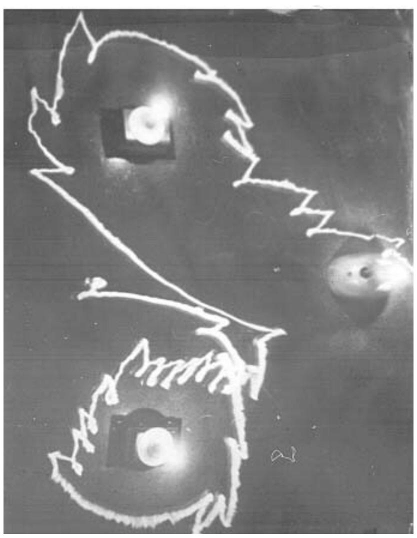

Figure 7-9. Choice behaviour as Elsie first visits one light source, and then visits the other. (rights

In a second experiment, Elsie was placed in an environment in which there were two lights, and (as can be seen in Figure 7-9) exhibited choice behavior. The robot started by being attracted to one of the two lights, and approached it. However, when it moved too close to that light, it veered away. As a result of veering away, it detected the second, "pleasantly" dimmer light, which it approached. Thus, the robot avoided the problem "of Buridan's ass, which starved to death, as some animals acting trophically in fact do, because two exactly equal piles of hay were precisely the same distance away" (Grey Walter, 1963, p. 128).

In a third experiment, the robot encountered a mirror, and its behavior was driven by the combined effects of its ability to detect its own reflected (and relatively dim) light source, and of its physical contact with the mirror. The result was the so-called "mirror dance", which is illustrated in Figure 7-10. The robot "lingers before a mirror, flickering, twittering and jigging like a clumsy Narcissus. The behavior of a creature thus engaged with its own reflection is quite specific, and on a purely empirical basis, if it were observed in an animal, might be accepted as evidence of some degree of self-awareness" (Grey Walter, 1963, pp. 128-129).

The electric components that were used to create the tortoises themselves led to an interesting emergent behavior. In particular, the sensitivity to light was dependent upon the degree to which the battery of a tortoise was charged. When fully charged, a bright light would repel the robot. However, when its battery was much weaker, the same bright light would attract the robot, because it would be recorded as being of moderate intensity. This enabled Grey Walter to use lights to control the ability of a tortoise to recharge itself. A hutch was built (see Figure 7-6); if the robot entered the hutch its battery would be recharged. Inside the hutch was a light. When a tortoise's battery began to fail, the robot was attracted by the bright hutch light, entered the hutch, and recharged. However, when the battery was fully recharged, the bright hutch light repelled the robot, so that it left the hutch and began to explore its environment once again.

(Grey Walter, 1950, 1951, 1963) reported the results of many different kinds of experiments, including some that involved a particularly complicated environment because it included two tortoises. He also designed a later version of the machine, *Machina docilis*, which was capable of being classically conditioned. It learned to be attracted to a high-pitched whistle. In general, the results of all of his experiments demonstrated quite clearly that the complexity of the behavior of his robots far exceeded the complexity of the components from which they were constructed.

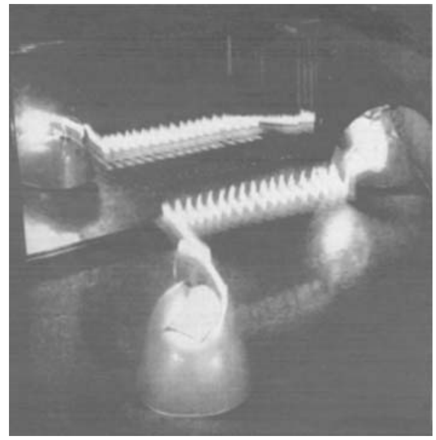

Figure 7-10. The mirror dance.

Recently, with the availability of robot construction toys such as Lego Mindstorms, new generations of Grey Walter's tortoises have appeared. For example, Figure 7-11 illustrates one such robot that my son Christopher and I constructed from Lego Dacta components. While not identical to a tortoise of the late 1940s, we attempted to build into it much of the functionality of Grey Walter's machines. For readers of the electronic version of this book, direct links to the instructions for building this robot, a listing of the program used to control it, and a movie illustrating its behavior, are included at the end of this chapter. For other readers, this material has been included on the CD that accompanies the text.

7.3.2.2 Implications

From where does the complexity of behavior arise? Simon (1996) explored this question with his famous parable of the ant. He imagined an ant walking along a beach, and that its trajectory along the beach was traced. This trajectory might be thought of as being a very complicated function; explaining the behavior of the ant was equivalent to explaining how the many twists and turns of this function arose. One might be tempted to attribute the properties of this function to fairly complicated internal navigational processes. Indeed, if one were to adopt an analytic approach, this kind of attribution would be expected. The trajectory would be taken as raw data, analyzed into key components, and the mechanisms that generate these key components would be attributed to the ant. However, Simon pointed out that this would likely lead to an incorrect theory. "Viewed as a geometric figure, the ant's path is irregular, complex, hard to describe. But its complexity is really a complexity in the surface of the beach, not a complexity in the ant" (p. 51). In other words, fairly simple dispositions of the ant – following the scent of a pheromone trail, turning in a particular direction when an obstacle is encountered – could lead to a very complicated trajectory, if the environment being navigated through was complicated enough.

Grey Walter's tortoises provide a robotic analog to the parable of the ant. As we have seen in the preceding section, the trajectories traced in the photographs of robot behavior are very complicated. However, this complexity is not reflected in the internal complexity of the tortoise. The inner workings of Grey Walter's robots were very simple and straightforward by design. The complexity in the observed behavior must be rooted in the complexity of the interaction between a simple robot and its environment. "So a two-element synthetic animal is enough to start with. The strange richness provided by this particular sort of permutation introduces right away one of the aspects of animal behavior – and human psychology – which *M. speculatrix* is designed to illustrate: the uncertainty, randomness, free will or independence so strikingly absent in most well-designed machines" (Grey Walter, 1950, p. 44). Again, feedback is a key – in this case feedback between the world and the machine.

Consider this issue from a different perspective, the analytic one that had to be taken by the "kids in the hallway" who were discussed in Chapter 1. When I introduce synthetic psychology in lectures, I often use Grey Walter's tortoises as an introduction. However, when I do this, I describe the performance of the robots first, presenting images like Figure 7-8, 7-9, 710 as behavioral data. Students are asked to infer the internal mechanisms of the machines on the basis of these images. Invariably, after analyzing the data that I have presented to them, they propose a far more complicated theory – one that involves many more internal properties – than is actually required. This is exactly the same situation that was observed in our Chapter 1 examples. It would appear that psychology students – and psychologists – have a strong tendency to ignore the parable of the ant, and prefer to locate the source of complicated behavior within the organism, and not within its environment.

Pfeifer and Scheier (1999) call this the frame-of-reference problem. "We have to distinguish between the perspective of an observer looking at an agent and the perspective of the agent itself. In particular, descriptions of behavior from an observer's perspective must not be taken as the internal mechanisms underlying the described behavior" (p. 112). This is because Pfeifer and Scheier believe that the behavior of a system cannot be explained by only appealing to internal mechanisms; an agent's behavior is always presumed by them to be the result of a system-environment interaction. "The complexity we observe in a particular behavior does not always indicate accurately the complexity of the underlying mechanisms."

Here we see one of the strong appeals of adopting the synthetic approach. By building a system and taking advantage of nonlinear interactions (such as feedback between components, and between a system and its environment), relatively simple systems can surprise us, and generate far more complicated behavior than we might expect. By itself, this demonstrates the reality of the frame-of-reference problem. However, the further appeal of the synthetic approach comes from the belief that if we have constructed the simple system, then we should be in a very good position to propose a simpler explanation of the complicated behavior. In particular, we should be in a better position than would be the case if we started with the behavior, and attempted to analyze it in order to understand the workings an agent's internal mechanisms. Later in this chapter I will argue that while this perspective is appealing, it is also very deceptive and dangerous.

7.4 VEHICLES

Surprisingly and disappointingly, Grey Walter's tortoises appear to have had a very short-lived academic impact, and had essentially disappeared from the scene by the end of the 1950s. The original machines, Elmer and Elsie, were constructed in 1948 and 1949. They were fairly unreliable machines when publicly demonstrated in 1949 and 1950, and were probably scrapped in 1951. That year, Grey Walter's technician Bunny Warren constructed six new, more reliable, tortoises that were demonstrated in public throughout the 1950s. Three of these had been on display at the 1951 Festival of Britain exhibition; two were auctioned off and later destroyed in an Australian house fire. The third is currently at the Smithsonian Institute. Of the other three, one went missing after being shipped to the United States. A second was dismantled, and never re-assembled, after an ill-fated attempt by an American company to use it to design a new toy. The third remained with Grey Walter until his death in 1977, and then was in the possession of his son Nicolas. This final machine was displayed in 1985, and was saved from disposal by Nicolas some years later. Stored in his basement for several years, it was rediscovered after a long search in 1995. It was refurbished, and used as a model for less fragile modern replicas. On display now at Bristol University, the Intelligent Autonomous Systems Engineering Laboratory there maintains a great deal of historical information about the tortoises at the Grey Walter Online Archive (http://idle.uwe.ac.uk/IAS/gwonline.html).

To my mind, one of the most striking examples of the disappearance of the tortoises is that Grey Walter's research was not cited in the book that provides the renaissance of his theoretical perspective. In *Vehicles*, Valentino Braitenberg proposed a series of 14 different thought experiments (Braitenberg, 1984). Each of these experiments involved conceptualizing a fairly simple machine, and considering how that machine might behave in different environments. Some of these machines are reminiscent of Elmer and Elsie. As Braitenberg's book progresses, the hypothetical machines become more sophisticated, as does their consequent behavior. One of the main themes of the book is one that is familiar from the current chapter: simple machines can generate far more complicated behavior than one might expect. A second theme pursued by Braitenberg is that his synthetic approach will lead to simpler explanations than those that would be attained if vehicle behaviors were approached analytically.

7.4.1 Braitenberg's General Approach

Valentino Braitenberg (b. 1926) is the emeritus director of the Max Planck Institute of Biological Cybernetics, an emeritus professor at the Institute of Medical Psychology and Behavioral Neurobiology of the Eberhard-Karls-University in Tübingen, Germany, and is the director of the cognitive science laboratory at the University of Trento in Italy. Braitenberg is a leading researcher in cybernetics and neuroscience, and the thought experiments that he presents in *Vehicles* are an attempt to understand some of the characteristics of the brain by adopting the synthetic approach. The final chapter of the book "sketch a few facts about animal brains that have inspired some of the properties of our vehicles, and their behavior will then seem less gratuitous than it may have seemed up to this point" (Braitenberg, 1984, p. 95). In general, Braitenberg takes an anatomical property of interest, reduces it to a very simple form, and considers the behavior of a simple machine that incorporates it.

In this section, we will briefly explore Braitenberg's (1984) approach by considering a couple of the devices that he proposed. After we have introduced some of these machines, we will be in a better position to seriously consider some of the pros and cons of adopting a synthetic research strategy.

7.4.2 Some Example Vehicles

7.4.2.1 Vehicle 1: Getting Around

Braitenberg (1984) constructed a deliberate evolutionary sequence that is traced from his early vehicles to the later ones. His early machines are very simple, and are easily thought of as organisms that swim around in water. The later, more sophisticated devices are better thought of as "little carts moving on hard surfaces" (p. 2).

His simplest vehicle is a swimming device that is best thought of as a cylinder or torpedo, with a sensor at one end (the front) and a motor at the other. The foundational design principle for this vehicle is the proportional relationship between the response of the sensor and the speed of the motor. As the sensor detects more of whatever quality it is designed to detect, the motor increases its speed. As the sensor detects less of this quality, the motor slows down. Under the assumption that this vehicle is moving in the real world, it will become under the influence of asymmetrical frictional influences. As a result, it will not travel in a perfectly straight line, but will instead follow a complicated trajectory that is both difficult to predict and to explain.

From the synthetic perspective used to create this vehicle, its overall behavior is very understandable. However, if faced with analyzing the behavior of the vehicle in the absence of any knowledge about its internal structure, it is likely to be very complicated. On observing this machine, "it is restless, you would say, and does not like warm water. But it is quite stupid, since it is not able to turn back to the nice cold spot it overshot in its restlessness. Anyway, you would say, it is ALIVE, since you have never seen a particle of dead matter move around quite like that" (Braitenberg, 1984, p. 5).

7.4.2.2 More Advanced Vehicles

The next set of vehicles proposed by Braitenberg (1984) are similar in spirit to Vehicle 1 in that they can be viewed as swimming devices propelled by motors whose speed is determined by the output of sensors. However, for these devices, there are two motors, one on each side at the back of the vehicle. Each sensor drives its own motor. The two sensors are mounted on each side at the front of the vehicle. Of interest is the anatomy of the connections between motors.

For instance, one vehicle might have excitatory connections (i.e., the same kind of sensor-motor relationship described for Vehicle 1) between the sensor and the motor on the same side of the vehicle. If the signal source being detected by the sensors is straight ahead of this vehicle, both motors will run at equal speeds, and the vehicle will run into the source. However, if the source is to one side, then the sensor nearer to the source will detect a stronger signal than will the sensor further from the source. As a result, the two motors will run at different speeds, causing the vehicle to turn away from the source. Braitenberg (1984) describes this vehicle as DISLIKING sources, becoming "restless in their vicinity and tends to avoid them, escaping until it safely reaches a place where the influence of the source is scarcely felt" (p. 9).

One could take the vehicle just described and cross its connections, so that the sensor on the right drives the motor on the left, and the sensor on the left drives the motor on the right. With these crossed connections, the sensor nearest the source drives the motor on the other side faster than the other sensor will drive the motor nearer the source. If the source is directly in front of the source, the vehicle will drive through it, as was the case for the previous vehicle. However, if the source is to one side of it, then the vehicle will turn towards the source instead of away from it. Braitenberg (1984) designates this vehicle as being AGGRESSIVE: "it, too, is excited by the presence of sources, but resolutely turns toward them and hits them with high velocity, as if it wanted to destroy them."

One common approach to studying the two types of vehicles that have just been described is to actually construct them, for instance using Lego Mindstorms or Lego Dacta components. The advantage of doing this is that their behavior is removed from the idealized domain of the thought experiment, and becomes subject to real-world influences. These influences include differential forces of friction on different robot parts, and the fact that no two presumably identical robot components will work in exactly the same way. "This means that it is usually more difficult than it seems to get a consistent and reliable automatic response to a stimulus" (Webb, 1996, p. 94). If the goal is to design a robot that will move in a straight line, then this is a serious problem. However, if the goal is to produce complex behavior from a simple system, then these vagaries of the environment become advantages. By adopting the synthetic approach "what seems like complex behavior in a robot can come from a surprisingly uncomplicated control algorithm" (p. 95).

The robots that were briefly described as being observed by the "kids in the hallway" in Chapter 1, and illustrated in Figure 1-1, were versions of the Braitenberg vehicles described in this subsection. These Lego Dacta machines were constructed and programmed by my daughter Michele and myself. Each robot used one motor to drive one rear wheel, and the speed of rotation of the wheel depended upon the output of a light sensor. In one robot, the connections between sensors and motors were crossed, in the other they were not. The only addition to these particular robots was that we used a "shell" to serve as an obstacle detector, as was the case with the Lego version of the tortoise that was described earlier. For readers of the electronic version of this book, direct links to the instructions for building these robots, a listing of the program used to control them, and a movie illustrating their behaviors, are included at the end of this chapter. For other readers, this material has been included on the CD that accompanies the text.

Braitenberg (1984) goes on to consider minor advances in the design of this kind of vehicle. For instance, the sensors might be tuned to be maximally sensitive to a particular range of signal from a source. When this is done, with crossed connections, the behavior of the vehicle mimics the phototropism exhibited by Grey Walter's tortoises. The connections between sensors and motors can be made inhibitory, so that a motor slows down when the sensor detects more of the signal. Motors can be driven by more than one sensor, each sensitive to a different kind of signal. In theory, one such vehicle would be straightforward to build, but would exhibit extremely complex behavior: “It dislikes high temperature, turns away from hot places, and at the same time seems to dislike light bulbs with even greater passion, since it turns toward them and destroys them” (p. 12). Again we see that producing emergent properties – where the whole of a system’s performance far exceeds the sum of its simple parts – are one of the key goals of the synthetic approach.

7.4.2.3 Vehicle 6: Selection, The Impersonal Engineer

When Braitenberg (1984) vehicles emerge from the sea to occupy the land, evolutionary ideas take a decidedly different role in his book. Braitenberg imagines a collection of vehicles, all operating on a table, a table that is surrounded by spare parts. A team of researchers also surrounds the table, and the goal of this team is to build new vehicles. The way that this process works is that a researcher takes one of the vehicles from the table, and uses it as a model for the creation of a copy from spare parts. Then both the original and the copy are placed back on the table.

A further twist to this thought experiment is the notion that the copies are being made in a hurry, and therefore the builders don’t have much time to check their work, or to test the adequacy of each copy. As a result, some of the copies that are placed back on the table will not be identical to the original that was used as a model. Many of these copies will be defective, and will therefore fall off the table to be used as parts for later generations of copies. “But it is also possible that we will unwittingly introduce a particularly shrewd variation into the pattern of connections, so that our copy will survive forever while the original may turn out to be unfit for survival after all” (Braitenberg, 1984, p. 27). Braitenberg argues that this is particularly likely if one vehicle is picked up and used as a model for one vehicle component, and a different vehicle is picked up and used as a model for a different vehicle component when the copy is being constructed. Of course, if the lucky mutation results in a longer life span for the copy, then this vehicle will be more likely to be picked up and used as the model for later generation systems.

7.4.2.4 Further Sophistications

Braitenberg (1984) proposes several additional modifications, and describes how they can be used to develop more advanced vehicles. Some of these vehicles have spatially organized sensors that permit them to detect the shapes of objects. Others have simple connectionist networks that enable them to learn from experience. Still others have feedback loops that enable them to predict the future.

All of these sophistications have two things in common. First, they are all made possible through the use of fairly straightforward materials and engineering. Second, when they are components of vehicles that are placed in interesting environments, extremely complicated behaviors can emerge. “It is pleasurable and easy to create little machines that do certain tricks. It is also quite easy to observe the full repertoire of behavior of these machines -- even if it goes beyond what we had originally planned, as it often does“ (Braitenberg, 1984).

7.5 SYNTHESIS AND EMERGENCE: SOME MODERN EXAMPLES

The historical examples that have been considered thus far in the chapter all point to two underlying themes. First, it is definitely possible to construct informative models by building complete systems from some set of assumed components, without the need of basing the model on extensive analyses of existing data. In other words, if one looks back at the previous examples, then one striking feature that should be noted is that neither the homeostat, the tortoises, nor the vehicles were models that were intended to fit extant behavior. Second, when this synthetic approach is taken, it is almost always the case that interactions between system components, and between these components and a complex environment, can produce surprising and interesting emergent behavior that usually exceeds the expectations of the system designer.

The research that has been reviewed above has inspired a great many modern research programs. In order to reinforce these two themes, let us take a moment to briefly review three more modern examples of complex behavior emerging (often unintentionally) from relatively simple systems that have been created via the synthetic approach.

7.5.1 NETtalk

DECtalk is a program for converting text into audible speech (Hallahan, 1996), and is widely viewed as the best commercially available product for this task. DECtalk consists of eight different processing "threads", each of which is concerned with a major stage of processing, ranging from buffering text in an input memory to generating audio via a computer's sound hardware. It does this by following a two-stage process. Of particular interest in the context of the current chapter is the letter-to-sound (LTS) thread that converts sequences of ASCII text into sequences of phonemes. First, the LTS thread separates the text stream into clauses, and normalizes the text by applying special processing rules to idiosyncratic text entries (numbers, abbreviations, and so on). Second, the remaining unprocessed test items are converted into phonemes in one of two ways. First, a word is looked up to see if it exists in a pronunciation dictionary of common words. (If this first lookup fails, the word will be tested for an English suffix. If the suffix is found, it will be removed, and the remaining word stem will be looked up in the dictionary again.) Second, if the word is not found in that dictionary, then it is converted into speech by applying a set of phonological rules that decompose the text into a sequence of morphemes. The phonological representation of the text that is generated by this two-stage process is then converted into audible speech by applying a set of transition rules to it, and then applying digital speech synthesis. During this stage of processing, the LTS thread will identify syllables in the morpheme sequences, and mark some of them for additional stress to make the ultimate speech output as natural sounding as possible. Also, the LTS thread will identify the context in which a particular phoneme is found (i.e., surrounding phonemes). This is because the pronunciation of some speech sounds will change as a function of context. The LTS thread has a series of rules that instantiate these context-dependent alterations.

While DECtalk exhibits outstanding performance, this is accomplished with considerable cost. (Hallahan, 1996) notes that the program is the product of over 30 man-years of development, and consists of around 160,000 lines of code. This large amount of code is required because there is a considerable amount of specific knowledge that is built into the program. For instance, the LTS thread alone has more than 1,500 rules of pronunciation. Even with this large number of rules, it still requires a dictionary of exceptional words that has over 15,000 entries. On older hardware, running DECtalk at settings that produced medium quality output resulted in its using 69% of a CPU's processing resources. Producing the highest quality output consumed 89% of the CPU's resources. DECtalk has only become more portable recently because of advances in CPU design.

NETtalk is a connectionist network that was intended to replace much of the LTS thread in DECtalk. Rather than handcrafting a large number of rules, and a dictionary of exceptional words, NETtalk was intended to be a fairly small program that learned to convert text into speech (Sejnowski & Rosenberg, 1988). The network had 7 groups of 29 input units per group to represent text, 80 hidden units, and 26 output units that represented phonemes for a total of 309 units and 18,629 weighted connections. Text was moved through an input "window", so that the network was trained to pronounce the text in the middle of the "window", while at the same time being aware of the text's context (i.e., the text on either side of the "window", which had either just been pronounced or was to be pronounced next). The network was trained on two different texts. One was phonetic transcription from the informal speech of a child. The other was over 20,000 different words from a dictionary. Training was accomplished using the generalized delta rule (Rumelhart, Hinton, & Williams, 1986). By the time the network had learned about 5000 words, its performance was nearly perfect, and its performance generalized quite well to words that it had not seen previously. Interestingly, the network was able to perform at this high level without requiring a large separate lookup table as is used in DECtalk.

NETtalk was explicitly designed to exhibit some of the functionality of DECtalk. It was not intended to have any implications at all for psychology or cognitive science. However, during training, NETtalk's output was channeled into audio hardware. Sejnowski and Rosenberg (1988) noted, "during the early stages of learning in NETtalk, the sounds produced by the network are uncannily similar to early speech sounds in children" (p. 670). They use this surprising finding to hypothesize that NETtalk might have discovered representations that are particularly efficient for use by a parallel networks, and that these representations may be similar to those employed by humans. They go on to suggest that the developmental regularities that have been observed in NETtalk and other networks (e.g., Elman et al., 1996; Rumelhart & McClelland, 1986) "may be a general property of incremental learning in networks with distributed representations" (p. 672). In other words, even though NETtalk was only intended as a particular feat of engineering, its surprising emergent behavior suggested that it might shed light on some topics of psychological interest.

7.5.2 Cricket Phonotaxis

A second example comes from the study of cricket phonotaxis. This section briefly reviews the central points of Webb's synthetic study of this phenomenon (Webb, 1996).

Phonotaxis, the ability to identify a particular sound and move towards it, is fundamental to a female cricket's choosing of a mate (Webb, 1996). A male cricket will generate a song as a series of syllables produced at a specific frequency and with a specific rhythm. A female cricket can use these properties to isolate the song of a male cricket of her own species from any other sound. After selecting the song, the female cricket will move towards the male producing it, even under conditions in which other males of the same species are chirping at the same time. The mechanisms underlying cricket phonotaxis are not yet completely understood.

Sounds from the world provide external stimulation to a cricket's eardrums, which are mounted on its forelegs. Sound also travels inside the cricket's body to the ears through a tracheal tube that connects the two ears to each other and to openings on the cricket's body called spinnacles. These internal and external sounds travel different distances, and therefore arrive at the same ear at different times, resulting in their being out of phase. The amount of phase shift depends upon the direction of the sound source. In general, the cricket's eardrum that is closer to the sound source will have higher amplitude of vibration.

What mechanisms are responsible for converting differences between eardrum vibration amplitudes into movements in the direction of the detected sound? Each eardrum stimulates a neuron that encodes amplitude. The larger the amplitude, the higher will be the spike train frequency of the neuron, and the sooner will it start to respond.

There are two theories of how the responses of the two neurons are used to direct the cricket's locomotion. One popular theory is that the cricket turns in the direction of the side with the neuron that is firing more frequently. However, this account would work for any sound, and thus requires postulating additional neural mechanisms for picking out the song with the correct rhythm.

A second, simpler theory is that with each sound burst the cricket turns in the direction of the side whose neuron begins to fire first. In other words, this theory ignores spike train frequency. This second theory has the advantage that it does not require additional rhythm-detecting circuitry, because changes in the rhythm of the detected song will naturally alter the onset of neural firing. However, it is not clear that this simple theory is sufficient to account for the regularities of cricket phonotaxis.

Webb (1996) adopted the synthetic approach to evaluate the adequacy of this second theory. Se constructed a LEGO robot with specialized electronics that mimicked the functionality of the neural circuits in the cricket's auditory system. The robot had two wheels driving it from the rear, each rotated by its own engine. When both motors were running, they pushed the robot forward. The robot was programmed to stop the engine of the side whose "ear circuit" reached threshold first. This resulted in the robot turning in that side's direction – with the aim of having it turn in the direction of the detected song.

The "ear circuitry" of the robot was optimally sensitive to a sound that had a specific frequency and rhythm. Webb (1996) began to test the adequacy of the theory by placing it at one side of an arena, and placing a speaker on the other side. She recorded the trajectory taken by the robot when sounds were broadcast from the speaker. When the sound was of the optimal frequency and rhythm, Webb found that the robot followed a zigzag path towards the speaker that was very similar to the trajectory taken by a female cricket. When the properties of the sound deviated from the optimal, the phonotactic behavior of the robot became far less successful. For example, when the syllable rate of the sound was increased, the robot drove through the arena in predominately straight lines. When the syllable rate was decreased, the robot followed a curved path towards the speaker, but rarely reached the speaker's actual location. These robot behaviors began to establish the adequacy of the second theory. "I discovered afterward that real crickets, too, tend to take curved paths at slower rates while failing more completely for faster rates. So the robot not only succeeds like a cricket but tends to fail like one too" (p. 98).

Female crickets will choose between songs generated by two different males of the same species, usually moving to the louder of the two songs. Webb (1996) realized that she had not explicitly programmed this ability into her robot. Nevertheless, she decided to see what the robot would do in an arena in which two speakers were present, and in which the same sound was being played through both. "To my surprise, the robot seemed to have no problem making up its mind (so to speak) and went almost directly to one speaker or the other" (p. 99). This suggests that the simple theory of phonotaxis may not only explain the general phenomenon of song isolation, but might also account for how a female cricket chooses one mate over another. "Again it appears that it is the interaction of the robot's uncomplicated mechanisms with particular sound fields that produces this interesting – and useful – behavior."

Webb (1996) used this experimental situation to generate a sound scenario that was completely unnatural. She alternated the location of the sound's generation between the two speakers in the arena. Under these conditions, the robot becomes confused, and moves between the two sounds. Experiments with actual crickets presented with this laboratory situation produced very similar results.

These kinds of results provide yet another demonstration of the advantages of the synthetic approach. Webb (1996) explicitly avoided building complicated capacities into her robot, and did not expect that the robot's behavior would be rich and varied. However, when this simple device was situated in the appropriate environment, its performance exceeded her expectations. "It shows that a rather competent and complex performance can come from a simple control mechanism, provided it interacts in the right way with its environment" (p. 99).

7.5.3 Stigmergy And Group Behavior

If you visit the website for the Collective Robotic Intelligence Project (CRIP) at the University of Alberta (http://www.cs.ualberta.ca/~kube/research.html), then you will have an opportunity to view some interesting video footage of a small collection of autonomous robots engaged in very complicated group behavior. Six small, cylindrically shaped robots move in an arena. In the middle of the arena is a brightly lit box. At the start of the video, four of the robots move directly to the box, while two others wander to one side of the arena. Of the robots that reach the box first, three line up side by side against it, while the fourth pauses behind this group, and then moves away. The three robots attempt to push the box, fail, and then break formation. Two return to a different position on the box, and are then joined by one of the other robots that had originally wandered off. When these three robots come in contact with the box, it begins to slide and turn. The movement of the box causes the three robots to break formation, but soon they return to push again. In a moment, the three other robots join them; the six robots jockey for position near one corner of the box, and push it quite quickly into a corner of the arena that is brightly lit with an overhead spotlight.

This video illustrates performance on a box transport task, which is one of the benchmark tests used to study cooperative behavior in robots. For each robot, the goal of this task is to locate the brightly lit box and to push it into a goal location, which is also brightly lit. Each robot is equipped with light sensors that point forward (for locating the box) and upward (for locating the goal). Once the robot detects a side of the box, it determines if the box is between the robot and the goal location. If it is, the robot pushes against the box. If it is not, the robot attempts to find a different position against the box. In many cases, this will result in the robot losing sight of the box, and having to search for it again. The robot can also lose sight of the box if another robot comes between it and the box.

The box transport task is designed to assess cooperative behavior, because the box is weighted so that at least two robots are required to move it. In order to succeed, the robots must position themselves along the box so that more than one of them push at the same time, and so that they are all pushing in a fairly consistent direction.

The astonishing thing about the behavior that can be seen in the website videos is that while it seems to be highly effective and coordinated, it is accomplished by very simple mechanisms. Furthermore, the robots do not explicitly communicate with each other, and are not centrally controlled. These robots are the culmination of several years of research that began with the study of core abilities in a group of software agents, and evolved into the performance of the physical robots that is illustrated on the website.

Kube and Zhang (1994) used software agents to explore some properties of cooperative behavior that were inspired from the study of social insects. They modeled the sensing and acting of a group of robots totally in a software environment. The simulated robots were provided with three sensors (one for the goal, one for obstacles, and one for other robots), two actuators (left and right wheel motors), and five simple behaviors. The behaviors were constructed using the subsumption architecture of Brooks (e.g., 1999). The default behavior is *find*, which causes the robot to move forward in a large arc. This behavior can be suppressed when the robot detects another; in this case it will change its behavior to *follow* the detected robot. If it gets too close to another robot while following, it will activate its *slow* behavior. If the goal sensor becomes active, then the robot will initiate the goal behavior, which causes it to move towards the goal. This behavior will only be stopped by initiating the *avoid* behavior, which occurs when the robot detects that a collision with another robot is imminent. Note that none of these behaviors involve communicating with other robots to coordinate their attack on a target. The simulation demonstrated the collective box transport behavior of the type that was later produced in real robots that incorporated most of these general behavioral principles (e.g., Kube & Bonabeau, 2000).

How does this cooperative behavior arise in robots that do not communicate directly with one another? The answer to this question again depends upon realizing that the robots (simulated or real) are situated in an environment that they are both sensing and acting upon. By changing the environment (e.g., by pushing the box, or blocking the path of another robot), they change the environment that is sensed by other robots, which in turn alters the behavior of the other robots. This indirect form of communication – accomplishing by directly altering the environment, and therefore indirectly altering the behavior of agents in the environment – is called stigmergy. This term comes from combining the terms *stigma* (wound from a pointed object) and *ergon* (work, product of labor) to produce a term whose meaning is “stimulating product of labor” (Holland & Melhuish, 1999).

Stigmergy was a term coined by French zoologist Pierre-Paul Grassé to explain the nest building behavior of termites (Theraulaz & Bonabeau, 1999). Grassé demonstrated that the termites themselves do not coordinate or regulate their building behavior, but that this is instead controlled by the nest structure itself. The current state of part of the nest stimulates a termite to perform an activity that alters the nest; the alteration in turn triggers a new behavior from either the same termite or from another. Stigmergy also provides an account of the construction of the nests of paper wasps (e.g., Karsai, 1999), and offers an alternative to older theories that attributed a fairly high degree of intelligence or higher-level rules to these insects. Stigmergy is generally viewed as a fairly simple mechanism for producing complex and coordinated performances from a group of agents, but has not been studied extensively. “The potential of stigmergy is still largely untapped in the biology community, in which it originated” (Theraulaz & Bonabeau, 1999 p. 113). Research on collective robotics, such as the box transport research cited above, or studies by Holland and Melhuish (1999) on how robots can exploit stigmergy to sort different objects into clusters, can be viewed as an attempt to increase our understanding of stigmergy, and to identify how it can interact with other principles to organize useful, collective behaviors.

For the purpose of the present chapter, stigmergy is an example of the “law of downhill synthesis”, which we will consider in more detail in the next section. From a robot designer’s point of view, an individual robot is provided with a very basic set of sensorimotor abilities, and is not required to include any facility for communicating directly with other agents. When placed in a complex environment – made particularly complicated by the presence of more than one agent – the result is complex collective behavior. Importantly, this behavior is completely emergent, because none of the capacities built into the robot are explicitly designed to be social or interactive.

7.6 THE LAW OF UPHILL ANALYSIS AND DOWNHILL SYNTHESIS

Brooks (2002) describes the behavior of one of his graduate students interacting with Cog, a humanoid robot with a moving head and arm, and with camera eyes that saccade to objects of interest (Brooks, Breazeal, Marjanovic, Scassellati, & Williamson, 1999). In this interaction, the student first held a whiteboard eraser and shook it. Then Cog would saccade to it, reach for it, and touch it. This sequence of events was then repeated, and it seemed clear that the two were taking turns. However, when this interaction occurred, the capacity for taking turns had not yet been programmed into Cog, and was not planned to be added to the robot for many years. The graduate student “had filled in the behavioral details so that the game of turn-taking with the eraser worked out. But she had done it subconsciously. She had picked up on the dynamics of what Cog could do and embedded them in a more elaborate setting, and Cog had been able to perform at a higher level than its design so far called for” (p. 92).

This anecdote illustrates one theme that we have seen in the historical and modern examples of synthetic research that have been presented in this chapter: the generation of behavior that is more complex than expected from a simple system embedded in an interesting environment. It also provides an example that shows, even subconsciously, that humans may have a natural tendency to be overly generous in assigning complexity to the internal systems of agents that we see in the world, or with which we might interact.

These two points are related to two complementary themes that have been argued to be central characteristics of the synthetic approach (Braitenberg, 1984). The first theme is “downhill synthesis”, which means that it is fairly straightforward to construct simple devices that, when they interact with the environment, produce surprising and interesting emergent behaviors. This theme is evident in the examples that we have seen in this chapter, as well as when we discussed the “thoughtless walkers” in Chapter 6.

The second theme is “uphill analysis”, which Braitenberg (1984) uses as an argument in favor of the synthetic approach, and against an approach in which the behaviors of existing systems are explained via analysis. "It is much more difficult to start from the outside and try to guess internal structure just form the observation of the data. [...] Analysis is more difficult than invention in the sense in which, generally, induction takes more time to perform than deduction: in induction one has to search for the way, whereas in deduction one follows a straightforward path. A psychological consequence of this is the following: when we analyze a mechanisms, we tend to overestimate its complexity". In other words, if the goal of synthetic psychology is to explain how various behaviors arise, then Braitenberg is claiming that the synthetic approach will lead to simpler theories than those arrived at by adopting the analytic perspective. Braitenberg feels strongly enough about this position to proclaim this "the law of uphill analysis and downhill synthesis."

One reason that the law of uphill analysis and downhill synthesis seems to be quite plausible is our sense that if a researcher has constructed a system, then he or she should have an excellent understanding of its inner workings, and therefore should be in an excellent position to offer straightforward mechanistic explanations of complex behavior. Given that the synthetic approach can produce rich and surprising results, this seems to make it an extremely attractive alternative to the more traditional analytic approach. However, it is important to realize that while the law of uphill analysis and downhill synthesis can provide grounds for arguing that the synthetic approach is attractive, it cannot justify abandoning analysis entirely. As a matter of fact, for synthetic psychology to succeed, synthesis and analysis must both be combined in a research program.

7.6.1 From Demonstration To Explanation

Why is analysis a required component of the synthetic approach? To answer this question, let us consider for a moment what the goals of a synthetic research program might be.

Brooks (1999, pp. 96-97) takes great pains to let us know what, in general, behavior-based robotics and, more specifically, his subsumption architecture, is *not*. It is not connectionism, nor neural networks, nor production rules, nor a blackboard control architecture, nor even German philosophy. What then is it?

It could be that behavior-based robotics merely demonstrates that complex behaviors frequently emerge from simple systems. To this point, this chapter could be considered to be a short catalogue of such demonstrations. However, of biologically inspired robots like the one used to study cricket phonotaxis, Webb (2000, p. 545) asks, "such examples of engineering can be attention grabbing, but what is their value for biological science? In particular, beyond the 'gimmick' of resemblance to natural systems, is any deeper understanding of how animals behave brought about by the building of such robot systems?"

The answer to questions like these depends first on determining whether the synthetic approach to robotics is intended to be anything more than attention grabbing demonstrations. Even a cursory glance at the literature would indicate that roboticists are interested in going beyond demonstrations, and coming up with theories of intelligence. For example, Adams, Breazeal, Brooks, and Scasselati (2000, p. 28) note, "just as computer simulations of neural nets have been used to explore and refine models from neuroscience, we can use humanoid robots to investigate and validate models from cognitive science and behavioral science." Webb (2000) argues that biologically inspired robots can be used to test existing hypotheses, to alter assumptions about stimuli and responses when confronted with a real environment, to enforce complete theories (and identify incomplete ones), and to produce novel hypotheses. Pfeifer and Scheier (1999) propose that the goal of embodied cognitive science is to achieve a better understanding of intelligence. "The methodology of embodied cognitive science is synthetic, its goal is understanding by building" (p. 631). With these goals in mind, merely generating complicated behavior is not a sufficient research program. The synthetic approach is in the business of explaining, and not just demonstrating.

If the synthetic approach is to generate new explanations of intelligent behavior, then analysis is going to be required. To see why this is so, imagine that a researcher is constructing autonomous systems according to a scheme similar to that described in Section 7.4.2.3, in which more successful systems are being selected for copying, and in which the copying process can introduce random mutations. (This hypothetical example is not so far fetched, as it captures the spirit of how problems are solved by genetic algorithms (e.g., Holland, 1992; Mitchell, 1996).) Imagine that after this process had been carried out for a certain period of time, one of the constructed systems exhibited a surprising and complicated behavior that was of considerable interest to psychologists. How would this system be used to contribute to psychological theory?

Simply demonstrating the interesting behavior would be interesting, but would not be satisfactory on its own. After all, a psychologists would already know of some other system that generates the behavior (i.e., a person or an animal), and would only be interested in this new system if it shed some light on how these other agents of interest worked. If the new system that demonstrated the behavior did not do this, it would be actually be complicating the situation, because instead of having one unexplained system (the person or animal), we would have two (the person/animal and the new autonomous system). As a result, in order to contribute to psychological theory, there would be a very strong demand for the researcher to explain the behavior of these new system – to say exactly how its inner mechanisms interacted with each other and with the environment to produce the behavior, and how the absence of such interactions resulted in the behavior not appearing in less successful systems.

In this particular hypothetical example, though, synthesis does not imply an easy route to understanding and explanation. The fact that the researcher constructed the system using selection implies that explanation must depend upon a later stage of analysis. This is because the success of this particular system (and the failure of other similar systems) was due to some random mutation that affected its internal mechanisms. This mutation was caused by the researcher, but not intentionally. To explain its behavior, the researcher would have to take the system apart, examine its inner workings, and probably take other systems apart as well to identify the differences between successful and unsuccessful systems. This later stage of analysis, while necessary, is likely to be difficult and intensive. Of vehicles created by natural selection, as is the case in this hypothetical example, Braitenberg (1984) writes "we can imagine that in most cases our analysis of brains in type 6 vehicles would fail altogether: the wiring that produces their behavior may be so complicated and involved that we will never be able to isolate a simple scheme" (p. 28).

Of course, synthetic researchers recognize that the analysis of their creations will be challenging. Nevertheless, they also realize that such analysis is required to generate explanations. For example, Pfeifer and Scheier (1999, p. 131) outline a ten-step research program for conducting experiments with agents. The last three steps of this program are purely analytic. They involve collecting data about the agent's behavior, as well as its internal states; the behavior is then described and analyzed statistically. The ultimate goal of this research program is to "formulate explanations of the agent's behavior".

Webb (2000) provides an additional argument for the need for analysis in her assessment of how biorobotics can contribute to biology. She notes that just because a robot generates the same behavior as an animal, it is not appropriate to conclude that they two systems exploit the same control mechanisms. This is because a standard realization in modeling is that the same behavior can be generated by, in principle, an infinite number of different algorithms (See also Dawson, 1998, Chapters 5 and 6). As a result, a great deal of analysis is required to determine whether the synthetic system and the modeled animal are strongly equivalent. "Proper experimental evaluation is needed to determine fully the real strengths or limitations of the implemented hypothesis. Behavior qualitatively similar to the animal in a few trials, while encouraging, cannot be taken as confirmation, yet too few studies do more" (Webb, 2000, p. 553).

7.6.2 Implications Of Braitenberg's Law

According to Braitenberg's (1984) law of uphill analysis and downhill synthesis, synthesis is much easier than analysis, and is more likely to circumvent the frame-of-reference problem. In other words, the synthetic approach should be capable of generating simpler theories than those that would be generated by the analytic approach. However, we have just seen that synthetic researchers have the goal of generating explanations of intelligence and behavior, and because of this goal realize that analysis is a crucial component of their research program. What then is really implied by the law of uphill analysis and downhill synthesis?

The law of uphill analysis and downhill synthesis is not a claim that analysis should be abandoned, but is instead a claim that the route to understanding and explanation should first involve performing synthesis, and then later conducting analysis. It is this combined approach – with an emphasis on early synthesis -- that holds the promise of generating simpler theories than an approach that exclusively involves analyzing the behavior of existing agents.

One reason for this promise is the fact that, as we have seen repeatedly, the synthetic approach is an explicit attempt to make the most by using the least. Synthetic modelers usually attempt to design fairly simple systems, in the hope that complex behaviors will emerge when they are situated in an environment. A second reason for this promise is that even in cases when researchers may not know precisely how to explain emergent behavior, the fact that they have constructed the model should make analysis easier, because they already have an accurate understanding of the main functional components of their model, and should therefore be in a position to target their analyses efficiently and appropriately.

7.6.3 Towards Synthetic Psychology

In this chapter, we have considered a number of examples of the synthetic approach. We have used these examples to demonstrate that one of the attractions of this approach is the fact that very simple mechanisms can generate complicated behavior when situated in a complex environment. We have seen that this in turn has led to the argument that the synthetic approach will lead to simpler theories than those that can be generated by the analytic approach. However, we have also seen that if the goal of the synthetic approach is to generate explanations, analysis cannot be completely abandoned. Synthetic research will usually involve early stages of synthesis that are followed by later stages of analysis, as was foreshadowed by the SEA approach that was outlined in Chapter 6.

Almost all of the examples of synthetic research that we have considered to this point have involved sensorimotor systems – in particular, robots of one sort or another. One question that needs to be addressed is whether such systems define exclusively the domain of the synthetic approach. Can the synthetic approach be applied to non-robotic systems? A second question that must be dealt with is whether systems of the type that we have been considering, which are predominately anti-representational, are of any interest to psychologists. Is synthetic psychology going to be reduced to studying non-representational systems that act on the world, or can the synthetic approach be applied to systems that use representations and have been of more interest to cognitive psychology and cognitive science? These questions are addressed in the next chapter.

7.7 INSTRUCTIONS, PROGRAMS AND MOVIES

I still have to create the material that will be included here. It will be similar to the media support at the end of Chapter 6, and will show the basic structure of a couple of Lego robots, give a short program listing for them, and will include a couple of QuickTime movies that demonstrate their behavior.

7.8 REFERENCES

Adams, B., Breazeal, C., Brooks, R. A., & Scasselati, B. (2000). Humanoid robots: A new kind of tool. *IEEE Intelligent Systems*, 25-31.

Ashby, W. R. (1956). *An Introduction To Cybernetics*. London: Chapman & Hall.
Ashby, W. R. (1960). *Design For A Brain* (Second Edition ed.). New York, NY: John Wiley & Sons.
Braitenberg, V. (1984). *Vehicles: Explorations In Synthetic Psychology*. Cambridge, MA: MIT Press.
Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, S., & Williamson, M. (1999). The Cog project: Building a humanoid robot. In C. Nehaniv (Ed.), *Computation for Metaphors, Analogy, and Agents* (pp. 52-87). Berlin: Springer-Verlag.
Brooks, R. A. (1999). *Cambrian Intelligence: The Early History Of The New AI*. Cambridge, MA: MIT Press.
Brooks, R. A. (2002). *Flesh And Machines: How Robots Will Change Us*. New York, NY: Pantheon Books.
Dawson, M. R. W. (1998). *Understanding Cognitive Science*. Oxford, UK: Blackwell.
Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness*. Cambridge, MA: MIT Press.
Grey Walter, W. (1950). An imitation of life. *Scientific American, 182*(5), 42-45.
Grey Walter, W. (1951). A machine that learns. *Scientific American, 184*(8), 60-63.
Grey Walter, W. (1963). *The Living Brain*. New York, NY: W.W. Norton & Co.
Hallahan, W. L. (1996). DECtalk software: Text-to-speech technology and implementation. *Digital Technical Journal, 7*, 5-19.
Holland, J. H. (1992). *Adaptation In Natural And Artificial Systems*. Cambridge, MA: MIT Press.
Holland, O., & Melhuish, C. (1999). Stigmergy, self-organization, and sorting in collective robotics. *Artificial Life, 5*, 173-202.
Karsai, I. (1999). Decentralized control of construction behavior in paper wasps: An overview of the stigmergy approach. *Artificial Life, 5*, 117-136.
Kube, C. R., & Bonabeau, E. (2000). Cooperative transport by ants and robots. *Robotics and Autonomous Systems, 30*, 85-101.
Kube, C. R., & Zhang, H. (1994). Collective robotics: From social insects to robots. *Adaptive Behavior, 2*, 189-218.
Mitchell, M. (1996). *An Introduction To Genetic Algorithms*. Cambridge, MA: MIT Press.
Pfeifer, R., & Scheier, C. (1999). *Understanding Intelligence*. Cambridge, MA: MIT Press.
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533-536.
Rumelhart, D. E., & McClelland, J. (1986). On learning the past tenses of English verbs. In J. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing. Volume 2: Psychological And Biological Models* (pp. 216-271). Cambridge, MA: MIT Press.
Sejnowski, T. J., & Rosenberg, C. R. (1988). NETtalk: A parallel network that learns to read aloud. In J. A. Anderson & E. Rosenfeld (Eds.), *Neurocomputing: Foundations Of Research* (pp. 663-672). Cambridge, MA: MIT Press.
Simon, H. A. (1996). *The Sciences Of The Artificial* (Third ed.). Cambridge, MA: MIT Press.
Theraulaz, G., & Bonabeau, E. (1999). A brief history of stigmergy. *Artificial Life, 5*, 97-116.
Webb, B. (1996). A cricket robot. *Scientific American, 275*, 94-99.
Webb, B. (2000). What does robotics offer animal behavior? *Animal Behavior, 60*, 545-558.
Wiener, N. (1948). *Cybernetics: Or Control And Communication In The Animal And The Machine*. Cambridge, MA: MIT Press.

