5

Elements of Embodied Cognitive Science

5.0 Chapter Overview

One of the key reactions against classical cognitive science was connectionism. A second reaction against the classical approach has also emerged. This second reaction is called embodied cognitive science, and the purpose of this chapter is to introduce its key elements.

Embodied cognitive science explicitly abandons the disembodied mind that serves as the core of classical cognitive science. It views the purpose of cognition not as building representations of the world, but instead as directing actions upon the world. As a result, the structure of an agent's body and how this body can sense and act upon the world become core elements. Embodied cognitive science emphasizes the embodiment and situatedness of agents.

Embodied cognitive science's emphasis on embodiment, situatedness, and action upon the world is detailed in the early sections of the chapter. This emphasis leads to a number of related elements: feedback between agents and environments, stigmergic control of behaviour, affordances and enactive perception, and cognitive scaffolding. In the first half of this chapter these notions are explained, showing how they too can be traced back to some of the fundamental assumptions of cybernetics. Also illustrated is how such ideas are radical departures from the ideas emphasized by classical cognitive scientists.

Not surprisingly, such differences in fundamental ideas lead to embodied cognitive science adopting methodologies that are atypical of classical cognitive science. Reverse engineering is replaced with forward engineering, as typified by behaviour-based robotics. These methodologies use an agent's environment to increase or leverage its abilities, and in turn they have led to novel accounts of complex human activities. For instance, embodied cognitive science can construe social interactions either as sense-act cycles in a social environment or as mediated by simulations that use our own brains or bodies as physical stand-ins for other agents.

In spite of such differences, it is still the case that there are structural similarities between embodied cognitive science and the other two approaches that have been introduced in the preceding chapters. The current chapter ends with a consideration of embodied cognitive science in light of Chapter 2's multiple levels of investigation, which were earlier used as a context in which to consider the research of both classical and of connectionist cognitive science.

5.1 Abandoning Methodological Solipsism

The goal of Cartesian philosophy was to provide a core of incontestable truths to serve as an anchor for knowledge (Descartes, 1960, 1996). Descartes believed that he had achieved this goal. However, the cost of this accomplishment was a fundamental separation between mind and body. Cartesian dualism disembodied the mind, because Descartes held that the mind's existence was independent of the existence of the body.

> I am not that structure of limbs which is called a human body, I am not even some thin vapor which permeates the limbs—a wind, fire, air, breath, or whatever I depict in my imagination, for these are things which I have supposed to be nothing. (Descartes, 1996, p. 18)

Cartesian dualism permeates a great deal of theorizing about the nature of mind and self, particularly in our current age of information technology. One such theory is posthumanism (Dewdney, 1998; Hayles, 1999). Posthumanism results when the content of information is more important than the physical medium in which it is represented, when consciousness is considered to be epiphenomenal, and when the human body is simply a prosthetic. Posthumanism is rooted in the pioneering work of cybernetics (Ashby, 1956, 1960; MacKay, 1969; Wiener, 1948), and is sympathetic to such futuristic views as uploading our minds into silicon bodies (Kurzweil, 1999, 2005; Moravec, 1988, 1999), because, in this view, the nature of the body is irrelevant to the nature of the mind. Hayles uncomfortably notes that a major implication of posthumanism is its "systematic devaluation of materiality and embodiment" (Hayles, 1999, p. 48); "because we are essentially information, we

can do away with the body" (Hayles, 1999, p. 12).

Some would argue that similar ideas pervade classical cognitive science. American psychologist Sylvia Scribner wrote that cognitive science "is haunted by a metaphysical spectre. The spectre goes by the familiar name of Cartesian dualism, which, in spite of its age, continues to cast a shadow over inquiries into the nature of human nature" (Scribner & Tobach, 1997, p. 308).

In Chapter 3 we observed that classical cognitive science departed from the Cartesian approach by seeking materialist explanations of cognition. Why then should it be haunted by dualism?

To answer this question, we examine how classical cognitive science explains, for instance, how a single agent produces different behaviours. Because classical cognitive science appeals to the representational theory of mind (Pylyshyn, 1984), it must claim that different behaviours must ultimately be rooted in different mental representations.

If different behaviours are caused by differences between representations, then classical cognitive science must be able to distinguish or individuate representational states. How is this done? The typical position adopted by classical cognitive science is called methodological solipsism (Fodor, 1980). Methodological solipsism individuates representational states only in terms of their relations to other representational states. Relations of the states to the external world—the agent's environment—are not considered. "Methodological solipsism in psychology is the view that psychological states should be construed without reference to anything beyond the boundary of the individual who has those states" (Wilson, 2004, p. 77).

The methodological solipsism that accompanies the representational theory of mind is an example of the classical sandwich (Hurley, 2001). The classical sandwich is the view that links between a cognitive agent's perceptions and a cognitive agent's actions must be mediated by internal thinking or planning. In the classical sandwich, models of cognition take the form of sense-think-act cycles (Brooks, 1999; Clark, 1997; Pfeifer & Scheier, 1999). Furthermore, these theories tend to place a strong emphasis on the purely mental part of cognition—the thinking—and at the same time strongly de-emphasize the physical—the action. In the classical sandwich, perception, thinking, and action are separate and unequal.

> On this traditional view, the mind passively receives sensory input from its environment, structures that input in cognition, and then marries the products of cognition to action in a peculiar sort of shotgun wedding. Action is a by-product of genuinely mental activity. (Hurley, 2001, p. 11)

Although connectionist cognitive science is a reaction against classical cognitivism, this reaction does not include a rejection of the separation of perception and action via internal representation. Artificial neural networks typically have undeveloped models of perception (i.e., input unit encodings) and action (i.e., output

unit encodings), and in modern networks communication between the two must be moderated by representational layers of hidden units.

> Highly artificial choices of input and output representations and poor choices of problem domains have, I believe, robbed the neural network revolution of some of its initial momentum. . . . The worry is, in essence, that a good deal of the research on artificial neural networks leaned too heavily on a rather classical conception of the nature of the problems. (Clark, 1997, p. 58)

The purpose of this chapter is to introduce embodied cognitive science, a fairly modern reaction against classical cognitive science. This approach is an explicit rejection of methodological solipsism. Embodied cognitive scientists argue that a cognitive theory must include an agent's environment as well as the agent's experience of that environment (Agre, 1997; Chemero, 2009; Clancey, 1997; Clark, 1997; Dawson, Dupuis, & Wilson, 2010; Dourish, 2001; Gibbs, 2006; Johnson, 2007; Menary, 2008; Pfeifer & Scheier, 1999; Shapiro, 2011; Varela, Thompson, & Rosch, 1991). They recognize that this experience depends on how the environment is sensed, which is situation; that an agent's situation depends upon its physical nature, which is embodiment; and that an embodied agent can act upon and change its environment (Webb & Consi, 2001). The embodied approach replaces the notion that cognition is representation with the notion that cognition is the control of actions upon the environment. As such, it can also be viewed as a reaction against a great deal of connectionist cognitive science.

In embodied cognitive science, the environment contributes in such a significant way to cognitive processing that some would argue that an agent's mind has leaked into the world (Clark, 1997; Hutchins, 1995; Menary, 2008, 2010; Noë, 2009; Wilson, 2004). For example, research in behaviour-based robotics eliminates resource-consuming representations of the world by letting the world serve as its own representation, one that can be accessed by a situated agent (Brooks, 1999). This robotics tradition has also shown that nonlinear interactions between an embodied agent and its environment can produce surprisingly complex behaviour, even when the internal components of an agent are exceedingly simple (Braitenberg, 1984; Grey Walter, 1950a, 1950b, 1951, 1963; Webb & Consi, 2001).

In short, embodied cognitive scientists argue that classical cognitive science's reliance on methodological solipsism—its Cartesian view of the disembodied mind—is a deep-seated error. "Classical rule-and-symbol-based AI may have made a fundamental error, mistaking the cognitive profile of the agent plus the environment for the cognitive profile of the naked brain" (Clark, 1997, p. 61).

In reacting against classical cognitive science, the embodied approach takes seriously the idea that Simon's (1969) parable of the ant might also be applicable to human cognition: "A man, viewed as a behaving system, is quite simple. The apparent complexity of his behavior over time is largely a reflection of the complexity

of the environment in which he finds himself" (p. 25). However, when it comes to specifics about applying such insight, embodied cognitive science is frustratingly fractured. "Embodied cognition, at this stage in its very brief history, is better considered a *research program* than a well-defined theory" (Shapiro, 2011, p. 2). Shapiro (2011) went on to note that this is because embodied cognitive science "exhibits much greater latitude in its subject matter, ontological commitment, and methodology than does standard cognitive science" (p. 2).

Shapiro (2011) distinguished three key themes that are present, often to differing degrees, in a variety of theories that belong to embodied cognitive science. The first of Shapiro's themes is *conceptualization*. According to this theme, the concepts that an agent requires to interact with its environment depend on the form of the agent's body. If different agents have different bodies, then their understanding or engagement with the world will differ as well. We explore the theme of conceptualization later in this chapter, in the discussion of concepts such as *umwelten*, affordances, and enactive perception.

Shapiro's (2011) second theme of embodied cognitive science is *replacement*: "An organism's body in interaction with its environment replaces the need for representational processes thought to have been at the core of cognition" (p. 4). The theme of replacement is central to the idea of cognitive scaffolding, in which agents exploit environmental resources for problem representation and solution.

> The biological brain takes all the help it can get. This help includes the use of external physical structures (both natural and artifactual), the use of language and cultural institutions, and the extensive use of other agents. (Clark, 1997, p. 80)

Shapiro's (2011) third theme of embodied cognitive science is *constitution*. According to this theme, the body or the world has more than a causal role in cognition—they are literally constituents of cognitive processing. The constitution hypothesis leads to one of the more interesting and radical proposals from embodied cognitive science, the extended mind. According to this hypothesis, which flies in the face of the Cartesian mind, the boundary of the mind is not the skin or the skull (Clark, 1997, p. 53): "Mind is a leaky organ, forever escaping its 'natural' confines and mingling shamelessly with body and with world."

One reason that Shapiro (2011) argued that embodied cognitive science is not a well-defined theory, but is instead a more ambiguous research program, is because these different themes are endorsed to different degrees by different embodied cognitive scientists. For example, consider the replacement hypothesis. On the one hand, some researchers, such as behaviour-based roboticists (Brooks, 1999) or radical embodied cognitive scientists (Chemero, 2009), are strongly anti-representational; their aim is to use embodied insights to expunge representational issues from cognitive science. On the other hand, some other researchers, such as philosopher Andy Clark (1997), have a more moderate view in which both

representational and non-representational forms of cognition might be present in the same agent.

Shapiro's (2011) three themes of conceptualization, replacement, and constitution characterize important principles that are the concern of the embodied approach. These principles also have important effects on the practice of embodied cognitive science. Because of their concern with environmental contributions to behavioural complexity, embodied cognitive scientists are much more likely to practise forward engineering or synthetic psychology (Braitenberg, 1984; Dawson, 2004; Dawson, Dupuis, & Wilson, 2010; Pfeifer & Scheier, 1999). In this approach, devices are first constructed and placed in an environment, to examine what complicated or surprising behaviours might emerge. Thus while in reverse engineering behavioural observations are the source of models, in forward engineering models are the source of behaviour to observe. Because of their concern about how engagement with the world is dependent upon the physical nature and abilities of agents, embodied cognitive scientists actively explore the role that embodiment plays in cognition. For instance, their growing interest in humanoid robots is motivated by the realization that human intelligence and development require human form (Breazeal, 2002; Brooks et al., 1999).

In the current chapter we introduce some of the key elements that characterize embodied cognitive science. These ideas are presented in the context of reactions against classical cognitive science in order to highlight their innovative nature. However, it is important to keep potential similarities between embodied cognitive science and the other two approaches in mind; while they are not emphasized here, the possibility of such similarities is a central theme of Part II of this book.

5.2 Societal Computing

The travelling salesman problem is a vital optimization problem (Gutin & Punnen, 2002; Lawler, 1985). It involves determining the order in which a salesman should visit a sequence of cities, stopping at each city only once, such that the shortest total distance is travelled. The problem is tremendously important: a modern bibliography cites 500 studies on how to solve it (Laporte & Osman, 1995).

One reason for the tremendous amount of research on the travelling salesman problem is that its solution can be applied to a dizzying array of real-world problems and situations (Punnen, 2002), including scheduling tasks, minimizing interference amongst a network of transmitters, data analysis in psychology, X-ray crystallography, overhauling gas turbine engines, warehouse order-picking problems, and wallpaper cutting. It has also attracted so much attention because it is difficult. The travelling salesman problem is an NP-complete problem (Kirkpatrick, Gelatt, & Vecchi, 1983), which means that as the number of cities involved in the

salesman's tour increases linearly, the computational effort for finding the shortest route increases exponentially.

Because of its importance and difficulty, a number of different approaches to solving the travelling salesman problem have been explored. These include a variety of numerical optimization algorithms (Bellmore & Nemhauser, 1968). Some other algorithms, such as simulated annealing, are derived from physical metaphors (Kirkpatrick, Gelatt, & Vecchi, 1983). Still other approaches are biologically inspired and include neural networks (Hopfield & Tank, 1985; Siqueira, Steiner, & Scheer, 2007), genetic algorithms (Braun, 1991; Fogel, 1988), and molecular computers built using DNA molecules (Lee et al., 2004).

Given the difficulty of the travelling salesman problem, it might seem foolish to suppose that cognitively simple agents are capable of solving it. However, evidence shows that a colony of ants is capable of solving a version of this problem, which has inspired new algorithms for solving the travelling salesman problem (Dorigo & Gambardella, 1997)!

One study of the Argentine ant *Iridomyrmex humilis* used a system of bridges to link the colony's nest to a food supply (Goss et al., 1989). The ants had to choose between two different routes at two different locations in the network of bridges; some of these routes were shorter than others. When food was initially discovered, ants traversed all of the routes with equal likelihood. However, shortly afterwards, a strong preference emerged: almost all of the ants chose the path that produced the shortest journey between the nest and the food.

The ants' solution to the travelling salesmen problem involved an interaction between the world and a basic behaviour: as *Iridomyrmex humilis* moves, it deposits a pheromone trail; the potency of this trail fades over time. An ant that by chance chooses the shortest path will add to the pheromone trail at the decision points sooner than will an ant that has taken a longer route. This means that as other ants arrive at a decision point they will find a stronger pheromone trail in the shorter direction, they will be more likely to choose this direction, and they will also add to the pheromone signal.

> Each ant that passes the choice point modifies the following ant's probability of choosing left or right by adding to the pheromone on the chosen path. This positive feedback system, after initial fluctuation, rapidly leads to one branch being 'selected.' (Goss et al., 1989, p. 581)

The ability of ants to choose shortest routes does not require a great deal of individual computational power. The solution to the travelling salesman problem emerges from the actions of the ant colony as a whole.

> The selection of the shortest branch is not the result of individual ants comparing the different lengths of each branch, but is instead a collective and self-organizing

> process, resulting from the interactions between the ants marking in both directions. (Goss et al., 1989, p. 581)

5.3 Stigmergy and Superorganisms

To compute solutions to the travelling salesman problem, ants from a colony interact with and alter their environment in a fairly minimal way: they deposit a pheromone trail that can be later detected by other colony members. However, impressive examples of richer interactions between social insects and their world are easily found.

For example, wasps are social insects that house their colonies in nests of intricate structure that exhibit, across species, tremendous variability in size, shape, and location (Downing & Jeanne, 1986). The size of nests ranges from a mere dozen to nearly a million cells or combs (Theraulaz, Bonabeau, & Deneubourg, 1998). The construction of some nests requires that specialized labour be coordinated (Jeanne, 1996, p. 473): "In the complexity and regularity of their nests and the diversity of their construction techniques, wasps equal or surpass many of the ants and bees."

More impressive nests are constructed by other kinds of insect colonies, such as termites, whose vast mounds are built over many years by millions of individual insects. A typical termite mound has a height of 2 metres, while some as high as 7 metres have been observed (von Frisch, 1974). Termite mounds adopt a variety of structural innovations to control their internal temperature, including ventilation shafts or shape and orientation to minimize the effects of sun or rain. Such nests,

> seem [to be] evidence of a master plan which controls the activities of the builders and is based on the requirements of the community. How this can come to pass within the enormous complex of millions of blind workers is something we do not know. (von Frisch, 1974, p. 150)

How do colonies of simple insects, such as wasps or termites, coordinate the actions of individuals to create their impressive, intricate nests? "One of the challenges of insect sociobiology is to explain how such colony-level behavior emerges from the individual decisions of members of the colony" (Jeanne, 1996, p. 473).

One theoretical approach to this problem is found in the pioneering work of entomologist William Morton Wheeler, who argued that biology had to explain how organisms cope with complex and unstable environments. With respect to social insects, Wheeler (1911) proposed that a colony of ants, considered as a whole, is actually an organism, calling the colony-as-organism the superorganism: "The animal colony is a true organism and not merely the analogue of the person" (p. 310).

Wheeler (1926) agreed that the characteristics of a superorganism must emerge from the actions of its parts, that is, its individual colony members. However, Wheeler also argued that higher-order properties could not be reduced to properties

of the superorganism's components. He endorsed ideas that were later popularized by Gestalt psychology, such as the notion that the whole is not merely the sum of its parts (Koffka, 1935; Köhler, 1947).

> The unique qualitative character of organic wholes is due to the peculiar non-additive relations or interactions among their parts. In other words, the whole is not merely a sum, or resultant, but also an emergent novelty, or creative synthesis. (Wheeler, 1926, p. 433)

Wheeler's theory is an example of holism (Sawyer, 2002), in which the regularities governing a whole system cannot be easily reduced to a theory that appeals to the properties of the system's parts. Holistic theories have often been criticized as being nonscientific (Wilson & Lumsden, 1991). The problem with these theories is that in many instances they resist traditional, reductionist approaches to defining the laws responsible for emerging regularities. "Holism is an idea that has haunted biology and philosophy for nearly a century, without coming into clear focus" (Wilson & Lumsden, 1991, p. 401).

Theorists who rejected Wheeler's proposal of the superorganism proposed alternative theories that reduced colonial intelligence to the actions of individual colony members. A pioneer of this alternative was a contemporary of Wheeler, French biologist Etienne Rabaud. "His entire work on insect societies was an attempt to demonstrate that each individual insect in a society behaves as if it were alone" (Theraulaz & Bonabeau, 1999). Wilson and Lumsden adopted a similar position:

> It is tempting to postulate some very complex force distinct from individual repertories and operating at the level of the colony. But a closer look shows that the superorganismic order is actually a straightforward summation of often surprisingly simple individual responses. (Wilson & Lumsden, 1991, p. 402)

Of interest to embodied cognitive science are theories which propose that dynamic environmental control guides the construction of the elaborate nests.

The first concern of such a theory is the general account that it provides of the behaviour of each individual. For example, consider one influential theory of wasp behaviour (Evans, 1966; Evans & West-Eberhard, 1970), in which a hierarchy of internal drives serves to release behaviours. For instance, high-level drives might include mating, feeding, and brood-rearing. Such drives set in motion lower-level sequences of behaviour, which in turn might activate even lower-level behavioural sequences. In short, Evans views wasp behaviour as being rooted in innate programs, where a program is a set of behaviours that are produced in a particular sequence, and where the sequence is dictated by the control of a hierarchical arrangement of drives. For example, a brood-rearing drive might activate a drive for capturing prey, which in turn activates a set of behaviours that produces a hunting flight.

Critically, though, Evans' programs are also controlled by releasing stimuli that are *external* to the wasp. In particular, one behaviour in the sequence is presumed to produce an environmental signal that serves to initiate the next behaviour in the sequence. For instance, in Evans' (1966) model of the construction of a burrow by a solitary digger wasp, the digging behaviour of a wasp produces loosened soil, which serves as a signal for the wasp to initiate scraping behaviour. This behaviour in turn causes the burrow to be clogged, which serves as a signal for clearing behaviour. Having a sequence of behaviours under the control of both internal drives and external releasers provides a balance between rigidity and flexibility; the internal drives serve to provide a general behavioural goal, while variations in external releasers can produce variations in behaviours: e.g., resulting in an atypical nest structure when nest damage elicits a varied behavioural sequence. "Each element in the 'reaction chain' is dependent upon that preceding it as well as upon certain factors in the environment (often gestalts), and each act is capable a certain latitude of execution" (p. 144).

If an individual's behaviour is a program whose actions are under some environmental control (Evans, 1966; Evans & West-Eberhard, 1970), then it is a small step to imagine how the actions of one member of a colony can affect the later actions of other members, even in the extreme case where there is absolutely no direct communication amongst colony members; an individual in the colony simply changes the environment in such a way that new behaviours are triggered by other colony members.

This kind of theorizing is prominent in modern accounts of nest construction by social paper wasps (Theraulaz & Bonabeau, 1999). A nest for such wasps consists of a lattice of cells, where each cell is essentially a comb created from a hexagonal arrangement of walls. When a large nest is under construction, where will new cells be added?

Theraulaz and Bonabeau (1999) answered this question by assuming that the addition of new cells was under environmental control. They hypothesized that an individual wasp's decision about where to build a new cell wall was driven by its perception of existing walls. Their theory consisted of two simple rules. First, if there is a location on the nest in which three walls of a cell already existed, then this was proposed as a stimulus to cause a wasp to add another wall here with high probability. Second, if only two walls already existed as part of a cell, this was also a stimulus to add a wall, but this stimulus produced this action with a much lower probability.

The crucial characteristic of this approach is that behaviour is controlled, and the activities of the members of a colony are coordinated, by a dynamic environment. That is, when an individual is triggered to add a cell wall to the nest, then the nest structure changes. Such changes in nest appearance in turn affect the behaviour of other wasps, affecting choices about the locations where walls will be added

next. Theraulaz and Bonabeau (1999) created a nest building simulation that only used these two rules, and demonstrated that it created simulated nests that were very similar in structure to real wasp nests.

In addition to adding cells laterally to the nest, wasps must also lengthen existing walls to accommodate the growth of larvae that live inside the cells. Karsai (1999) proposed another environmentally controlled model of this aspect of nest building. His theory is that wasps perceive the relative difference between the longest and the shortest wall of a cell. If this difference was below a threshold value, then the cell was untouched. However, if this difference exceeded a certain threshold, then this would cause a wasp to lengthen the shortest wall. Karsai used a computer simulation to demonstrate that this simple model provided an accurate account of the three-dimensional growth of a wasp nest over time.

The externalization of control illustrated in theories of wasp nest construction is called stigmergy (Grasse, 1959). The term comes from the Greek *stigma*, meaning "sting," and *ergon,* meaning "work," capturing the notion that the environment is a stimulus that causes particular work, or behaviour, to occur. It was first used in theories of termite mound construction proposed by French zoologist Pierre-Paul Grassé (Theraulaz & Bonabeau, 1999). Grassé demonstrated that the termites themselves do not coordinate or regulate their building behaviour, but that this is instead controlled by the mound structure itself.

Stigmergy is appealing because it can explain how very simple agents create extremely complex products, particularly in the case where the final product, such as a termite mound, is extended in space and time far beyond the life expectancy of the organisms that create it. As well, it accounts for the building of large, sophisticated nests without the need for a complete blueprint and without the need for direct communication amongst colony members (Bonabeau et al., 1998; Downing & Jeanne, 1988; Grasse, 1959; Karsai, 1999; Karsai & Penzes, 1998; Karsai & Wenzel, 2000; Theraulaz & Bonabeau, 1995).

Stigmergy places an emphasis on the importance of the environment that is typically absent in the classical sandwich that characterizes theories in both classical and connectionist cognitive science. However, early classical theories were sympathetic to the role of stigmergy (Simon, 1969). In Simon's famous parable of the ant, observers recorded the path travelled by an ant along a beach. How might we account for the complicated twists and turns of the ant's route? Cognitive scientists tend to explain complex behaviours by invoking complicated representational mechanisms (Braitenberg, 1984). In contrast, Simon (1969) noted that the path might result from simple internal processes reacting to complex external forces—the various obstacles along the natural terrain of the beach: "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. 24).

Similarly, Braitenberg (1984) argued that when researchers explain behaviour by appealing to internal processes, they ignore the environment: "When we analyze a mechanism, we tend to overestimate its complexity" (p. 20). He suggested an alternative approach, synthetic psychology, in which simple agents (such as robots) are built and then observed in environments of varying complexity. This approach can provide cognitive science with more powerful, and much simpler, theories by taking advantage of the fact that not all of the intelligence must be placed inside an agent.

Embodied cognitive scientists recognize that the external world can be used to scaffold cognition and that working memory—and other components of a classical architecture—have leaked into the world (Brooks, 1999; Chemero, 2009; Clark, 1997, 2003; Hutchins, 1995; Pfeifer & Scheier, 1999). In many respect, embodied cognitive science is primarily a reaction against the overemphasis of internal processing that is imposed by the classical sandwich.

5.4 Embodiment, Situatedness, and Feedback

Theories that incorporate stigmergy demonstrate the plausibility of removing central cognitive control; perhaps embodied cognitive science could replace the classical sandwich's sense-think-act cycle with sense-act reflexes.

> The realization was that the so-called central systems of intelligence—or core AI as it has been referred to more recently—was perhaps an unnecessary illusion, and that all the power of intelligence arose from the coupling of perception and actuation systems. (Brooks, 1999, p. viii)

For a stigmergic theory to have any power at all, agents must exhibit two critical abilities. First, they must be able to sense their world. Second, they must be able to physically act upon the world. For instance, stigmergic control of nest construction would be impossible if wasps could neither sense local attributes of nest structure nor act upon the nest to change its appearance.

In embodied cognitive science, an agent's ability to sense its world is called situatedness. For the time being, we will simply equate situatedness with the ability to sense. However, situatedness is more complicated than this, because it depends critically upon the physical nature of an agent, including its sensory apparatus and its bodily structure. These issues will be considered in more detail in the next section.

In embodied cognitive science, an agent's ability to act upon and alter its world depends upon its embodiment. In the most general sense, to say that an agent is embodied is to say that it is an artifact, that it has physical existence. Thus while neither a thought experiment (Braitenberg, 1984) nor a computer simulation (Wilhelms & Skinner, 1990) for exploring a Braitenberg vehicle are embodied, a

physical robot that acts like a Braitenberg vehicle (Dawson, Dupuis, & Wilson, 2010) *is* embodied. The physical structure of the robot itself is important in the sense that it is a source of behavioural complexity. Computer simulations of Braitenberg vehicles are idealizations in which all motors and sensors work perfectly. This is impossible in a physically realized robot. In an embodied agent, one motor will be less powerful than another, or one sensor may be less effective than another. Such differences will alter robot behaviour. These imperfections are another important source of behavioural complexity, but are absent when such vehicles are created in simulated and idealized worlds.

However, embodiment is more complicated than mere physical existence. Physically existing agents can be embodied to different degrees (Fong, Nourbakh sh, & Dautenhahn, 2003). This is because some definitions of embodiment relate to the extent to which an agent can alter its environment. For instance, Fong, Nourbakhsh, & Dautenhahn (2003, p. 149) argued that "embodiment is grounded in the relationship between a system and its environment. The more a robot can perturb an environment, and be perturbed by it, the more it is embodied." As a result, not all robots are equally embodied (Dawson, Dupuis, & Wilson, 2010). A robot that is more strongly embodied than another is a robot that is more capable of affecting, and being affected by, its environment.

The power of embodied cognitive science emerges from agents that are both situated and embodied. This is because these two characteristics provide a critical source of nonlinearity called feedback (Ashby, 1956; Wiener, 1948). Feedback occurs when information about an action's effect on the world is used to inform the progress of that action. As Ashby (1956, p. 53) noted, "'feedback' exists between two parts when each affects the other," when "circularity of action exists between the parts of a dynamic system."

Wiener (1948) realized that feedback was central to a core of problems involving communication, control, and statistical mechanics, and that it was crucial to both biological agents and artificial systems. He provided a mathematical framework for studying communication and control, defining the discipline that he called cybernetics. The term *cybernetics* 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" (Wiener, 1948, p. 11). Interestingly, engine governors make frequent appearances in formal discussions of the embodied approach (Clark, 1997; Port & van Gelder, 1995b; Shapiro, 2011).

The problem with the nonlinearity produced by feedback is that it makes computational analyses extraordinarily difficult. This is because the mathematics of feedback relationships between even small numbers of components is essentially

intractable. For instance, Ashby (1956) realized that feedback amongst a machine that only consisted of four simple components could not analyzed:

> 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 everyone 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. (Ashby, 1956, p. 54)

For this reason, embodied cognitive science is often practised using forward engineering, which is a kind of synthetic methodology (Braitenberg, 1984; Dawson, 2004; Pfeifer & Scheier, 1999). That is, researchers do not take a complete agent and reverse engineer it into its components. Instead, they take a small number of simple components, compose them into an intact system, set the components in motion in an environment of interest, and observe the resulting behaviours.

For instance, Ashby (1960) investigated the complexities of his four-component machine not by dealing with intractable mathematics, but by building and observing a working device, the Homeostat. It comprised four identical machines (electrical input-output devices), incorporated mutual feedback, and permitted him to observe the behaviour, which was the movement of indicators for each machine. Ashby discovered that the Homeostat could learn; he reinforced its responses by physically manipulating the dial of one component to "punish" an incorrect response (e.g., for moving one of its needles in the incorrect direction). Ashby also found that the Homeostat could adapt to two different environments that were alternated from trial to trial. This knowledge was unattainable from mathematical analyses. "A better demonstration can be given by a machine, built so that we know its nature exactly and on which we can observe what will happen in various conditions" (p. 99).

Braitenberg (1984) has argued that an advantage of forward engineering is that it will produce theories that are simpler than those that will be attained by reverse engineering. This is because when complex or surprising behaviours emerge, pre-existing knowledge of the components—which were constructed by the researcher—can be used to generate simpler explanations of the behaviour.

> 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. (Braitenberg, 1984, p. 20)

Braitenberg called this the law of uphill analysis and downhill synthesis.

Another way in which to consider the law of uphill analysis and downhill synthesis is to apply Simon's (1969) parable of the ant. If the environment is taken

seriously as a contributor to the complexity of the behaviour of a situated and embodied agent, then one can take advantage of the agent's world and propose less complex internal mechanisms that still produce the desired intricate results. This idea is central to the replacement hypothesis that Shapiro (2011) has argued is a fundamental characteristic of embodied cognitive science.

5.5 *Umwelten*, Affordances, and Enactive Perception

The situatedness of an agent is not merely perception; the nature of an agent's perceptual apparatus is a critical component of situatedness. Clearly agents can only experience the world in particular ways because of limits, or specializations, in their sensory apparatus (Uexküll, 2001). Ethologist Jakob von Uexküll coined the term *umwelt* to denote the "island of the senses" produced by the unique way in which an organism is perceptually engaged with its world. Uexküll realized that because different organisms experience the world in different ways, they can live in the same world but at the same time exist in different *umwelten*. Similarly, the ecological theory of perception (Gibson, 1966, 1979) recognized that one could not separate the characteristics of an organism from the characteristics of its environment. "It is often neglected that the words *animal* and *environment* make an inseparable pair" (Gibson, 1979, p. 8).

The inseparability of animal and environment can at times even be rooted in the structure of an agent's body. For instance, bats provide a prototypical example of an active-sensing system (MacIver, 2008) because they emit a high-frequency sound and detect the location of targets by processing the echo. The horizontal position of a target (e.g., a prey insect) is uniquely determined by the difference in time between the echo's arrival to the left and right ears. However, this information is not sufficient to specify the vertical position of the target. The physical nature of bat ears solves this problem. The visible external structure (the pinna and the tragus) of the bat's ear has an extremely intricate shape. As a result, returning echoes strike the ear at different angles of entry. This provides additional auditory cues that vary systematically with the vertical position of the target (Wotton, Haresign, & Simmons, 1995; Wotton & Simmons, 2000). In other words, the bat's body—in particular, the shape of its ears—is critical to its *umwelt*.

Passive and active characteristics of an agent's body are central to theories of perception that are most consistent with embodied cognitive science (Gibson, 1966, 1979; Noë, 2004). This is because embodied cognitive science has arisen as part of a reaction against the Cartesian view of mind that inspired classical cognitive science. In particular, classical cognitive science inherited Descartes' notion (Descartes, 1960, 1996) of the disembodied mind that had descended from Descartes' claim of *Cogito ergo sum*. Embodied cognitive scientists have been

strongly influenced by philosophical positions which arose as reactions against Descartes, such as Martin Heidegger's *Being and Time* (Heidegger, 1962), originally published in 1927. Heidegger criticized Descartes for adopting many of the terms of older philosophies but failing to recognize a critical element, their interactive relationship to the world: "The ancient way of interpreting the Being of entities is oriented towards the 'world' or 'Nature' in the widest sense" (Heidegger, 1962, p. 47). Heidegger argued instead for Being-in-the-world as a primary mode of existence. Being-in-the-world is not just being spatially located in an environment, but is a mode of existence in which an agent is actively engaged with entities in the world.

Dawson, Dupuis, and Wilson (2010) used a passive dynamic walker to illustrate this inseparability of agent and environment. A passive dynamic walker is an agent that walks without requiring active control: its walking gait is completely due to gravity and inertia (McGeer, 1990). Their simplicity and low energy requirements have made them very important models for the development of walking robots (Alexander, 2005; Collins et al., 2005; Kurz et al., 2008; Ohta, Yamakita, & Furuta, 2001; Safa, Saadat, & Naraghi, 2007; Wisse, Schwab, & van der Helm, 2004). Dawson, Dupuis, and Wilson constructed a version of McGeer's (1990) original walker from LEGO. The walker itself was essentially a straight-legged hinge that would walk down an inclined ramp. However, the ramp had to be of a particular slope and had to have properly spaced platforms with gaps in between to permit the agent's legs to swing. Thus the LEGO hinge that Dawson, Dupuis, and Wilson (2010) built had the disposition to walk, but it required a specialized environment to have this disposition realized. The LEGO passive dynamic walker is only a walker when it interacts with the special properties of its ramp. Passive dynamic walking is not a characteristic of a device, but is instead a characteristic of a device being in a particular world.

Being-in-the-world is related to the concept of affordances developed by psychologist James J. Gibson (Gibson, 1979). In general terms, the affordances of an object are the possibilities for action that a particular object permits a particular agent. "The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill" (p. 127). Again, affordances emerge from an integral relationship between an object's properties and an agent's abilities to act.

> Note that the four properties listed—horizontal, flat, extended, and rigid—would be *physical* properties of a surface if they were measured with the scales and standard units used in physics. As an affordance of support for a species of animal, however, they have to be measured *relative to the animal*. They are unique for that animal. They are not just abstract physical properties. (p. 127)

Given that affordances are defined in terms of an organism's potential actions, it is not surprising that action is central to Gibson's (1966, 1979) ecological approach

to perception. Gibson (1966, p. 49) noted that "when the 'senses' are considered as active systems they are classified by modes of activity not by modes of conscious quality." Gibson's emphasis on action and the world caused his theory to be criticized by classical cognitive science (Fodor & Pylyshyn, 1981). Perhaps it is not surprising that the embodied reaction to classical cognitive science has been accompanied by a modern theory of perception that has descended from Gibson's work: the enactive approach to perception (Noë, 2004).

Enactive perception reacts against the traditional view that perception is constructing internal representations of the external world. Enactive perception argues instead that the role of perception is to access information in the world when it is needed. That is, perception is not a representational process, but is instead a sensorimotor skill (Noë, 2004). "Perceiving is a way of acting. Perception is not something that happens to us, or in us. It is something we do" (p. 1).

Action plays multiple central roles in the theory of enactive perception (Noë, 2004). First, the purpose of perception is not viewed as building internal representations of the world, but instead as controlling action on the world. Second, and related to the importance of controlling action, our perceptual understanding of objects is sensorimotor, much like Gibson's (1979) notion of affordance. That is, we obtain an understanding of the external world that is related to its changes in appearance that would result by changing our position—by acting on an object, or by moving to a new position. Third, perception is to be an intrinsically exploratory process. As a result, we do not construct complete visual representations of the world. Instead, perceptual objects are virtual—we have access to properties in the world when needed, and only through action.

> Our sense of the perceptual presence of the cat as a whole now does not require us to be committed to the idea that we represent the whole cat in consciousness at once. What it requires, rather, is that we take ourselves to have *access*, now, to the whole cat. The cat, the tomato, the bottle, the detailed scene, all are present perceptually in the sense that they are perceptually accessible to us. (Noë, 2004, p. 63)

Empirical support for the virtual presence of objects is provided by the phenomenon of change blindness. Change blindness occurs when a visual change occurs in plain sight of a viewer, but the viewer does not notice the change. For instance, in one experiment (O'Regan et al., 2000), subjects inspect an image of a Paris street scene. During this inspection, the colour of a car in the foreground of the image changes, but a subject does not notice this change! Change blindness supports the view that representations of the world are not constructed. "The upshot of this is that *all* detail is present in experience not as represented, but rather as accessible" (Noë, 2004, p. 193). Accessibility depends on action, and action also depends on embodiment. "To perceive like us, it follows, you must have a body like ours" (p. 25).

5.6 Horizontal Layers of Control

Classical cognitive science usually assumes that the primary purpose of cognition is planning (Anderson, 1983; Newell, 1990); this planning is used to mediate perception and action. As a result, classical theories take the form of the sense-think-act cycle (Pfeifer & Scheier, 1999). Furthermore, the "thinking" component of this cycle is emphasized far more than either the "sensing" or the "acting." "One problem with psychology's attempt at cognitive theory has been our persistence in thinking about cognition without bringing in perceptual and motor processes" (Newell, 1990, p. 15).

Embodied cognitive science (Agre, 1997; Brooks, 1999, 2002; Chemero, 2009; Clancey, 1997; Clark, 1997, 2003, 2008; Pfeifer & Scheier, 1999; Robbins & Aydede, 2009; Shapiro, 2011; Varela, Thompson, & Rosch, 1991) recognizes the importance of sensing and acting, and reacts against central cognitive control. Its more radical proponents strive to completely replace the sense-think-act cycle with sense-act mechanisms.

This reaction is consistent with several themes in the current chapter: the importance of the environment, degrees of embodiment, feedback between the world and the agent, and the integral relationship between an agent's body and its *umwelt*. Given these themes, it becomes quite plausible to reject the proposal that cognition is used to plan, and to posit instead that the purpose of cognition is to guide action:

> The brain should not be seen as primarily a locus of inner *descriptions* of external states of affairs; rather, it should be seen as a locus of internal *structures* that act as operators upon the world via their role in determining actions. (Clark, 1997, 47)

Importantly, these structures do not stand between sensing and acting, but instead provide direct links between them.

The action-based reaction against classical cognitivism is typified by pioneering work in behaviour-based robotics (Brooks, 1989, 1991, 1999, 2002; Brooks & Flynn, 1989). Roboticist Rodney Brooks construes the classical sandwich as a set of vertical processing layers that separate perception and action. His alternative is a hierarchical arrangement of horizontal processing layers that directly connect perception and action.

Brooks' action-based approach to behaviour is called the subsumption architecture (Brooks, 1999). The subsumption architecture is a set of modules. However, these modules are somewhat different in nature than those that were discussed in Chapter 3 (see also Fodor, 1983). This is because each module in the subsumption architecture can be described as a sense-act mechanism. That is, every module can have access to sensed information, as well as to actuators. This means that modules in the subsumption architecture do not separate perception from action. Instead, each module is used to control some action on the basis of sensed information.

The subsumption architecture arranges modules hierarchically. Lower-level

modules provide basic, general-purpose, sense-act functions. Higher-level modules provide more complex and more specific sense-act functions that can exploit the operations of lower-level operations. For instance, in an autonomous robot the lowest-level module might simply activate motors to move a robot forward (e.g., Dawson, Dupuis. & Wilson, 2010, Chapter 7). The next level might activate a steering mechanism. This second level causes the robot to wander by taking advantage of the movement provided by the lower level. If the lower level were not operating, then wandering would not occur: because although the steering mechanism was operating, the vehicle would not be moving forward.

Vertical sense-act modules, which are the foundation of the subsumption architecture, also appear to exist in the human brain (Goodale, 1988, 1990, 1995; Goodale & Humphrey, 1998; Goodale, Milner, Jakobson, & Carey, 1991; Jakobson et al., 1991).

There is a long-established view that two distinct physiological pathways exist in the human visual system (Livingstone & Hubel, 1988; Maunsell & Newsome, 1987; Ungerleider & Mishkin, 1982): one, the ventral stream, for processing the appearance of objects; the other, the dorsal stream, for processing their locations. In short, in object perception the ventral stream delivers the "what," while the dorsal stream delivers the "where." This view is supported by double dissociation evidence observed in clinical patients: brain injuries can cause severe problems in seeing motion but leave form perception unaffected, or vice versa (Botez, 1975; Hess, Baker, & Zihl, 1989; Zihl, von Cramon, & Mai, 1983).

There has been a more recent reconceptualization of this classic distinction: the duplex approach to vision (Goodale & Humphrey, 1998), which maintains the physiological distinction between the ventral and dorsal streams but reinterprets their functions. In the duplex theory, the ventral stream creates perceptual representations, while the dorsal stream mediates the visual control of action.

> The functional distinction is not between 'what' and 'where,' but between the way in which the visual information about a broad range of object parameters are transformed either for perceptual purposes or for the control of goal-directed actions. (Goodale & Humphrey, 1998, p. 187)

The duplex theory can be seen as representational theory that is elaborated in such a way that fundamental characteristics of the subsumption architecture are present. These results can be used to argue that the human brain is not completely structured as a "classical sandwich." On the one hand, in the duplex theory the purpose of the ventral stream is to create a representation of the perceived world (Goodale & Humphrey, 1998). On the other hand, in the duplex theory the purpose of the dorsal stream is the control of action, because it functions to convert visual information directly into motor commands. In the duplex theory, the ventral stream is strikingly similar to the vertical layers of the subsumption architecture.

Double dissociation evidence from cognitive neuroscience has been used to support the duplex theory. The study of one brain-injured subject (Goodale et al., 1991) revealed normal basic sensation. However, the patient could not describe the orientation or shape of any visual contour, no matter what visual information was used to create it. While this information could not be consciously reported, it was available, and could control actions. The patient could grasp objects, or insert objects through oriented slots, in a fashion indistinguishable from control subjects, even to the fine details that are observed when such actions are initiated and then carried out. This pattern of evidence suggests that the patient's ventral stream was damaged, but that the dorsal stream was unaffected and controlled visual actions. "At some level in normal brains the visual processing underlying 'conscious' perceptual judgments must operate separately from that underlying the 'automatic' visuomotor guidance of skilled actions of the hand and limb" (p. 155).

Other kinds of brain injuries produce a very different pattern of abnormalities, establishing the double dissociation that supports the duplex theory. For instance, damage to the posterior parietal cortex—part of the dorsal stream—can cause optic ataxia, in which visual information cannot be used to control actions towards objects presented in the part of the visual field affected by the brain injury (Jakobson et al., 1991). Optic ataxia, however, does not impair the ability to perceive the orientation and shapes of visual contours.

Healthy subjects can also provide support for the duplex theory. For instance, in one study subjects reached toward an object whose position changed during a saccadic eye movement (Pelisson et al., 1986). As a result, subjects were not conscious of the target's change in location. Nevertheless, they compensated to the object's new position when they reached towards it. "No perceptual change occurred, while the hand pointing response was shifted systematically, showing that different mechanisms were involved in visual perception and in the control of the motor response" (p. 309). This supports the existence of "horizontal" sense-act modules in the human brain.

5.7 Mind in Action

Shakey was a 1960s robot that used a variety of sensors and motors to navigate through a controlled indoor environment (Nilsson, 1984). It did so by uploading its sensor readings to a central computer that stored, updated, and manipulated a model of Shakey's world. This representation was used to develop plans of action to be put into effect, providing the important filling for Shakey's classical sandwich.

Shakey impressed in its ability to navigate around obstacles and move objects to desired locations. However, it also demonstrated some key limitations of the classical sandwich. In particular, Shakey was extremely slow. Shakey typically required

several hours to complete a task (Moravec, 1999), because the internal model of its world was computationally expensive to create and update. The problem with the sense-think-act cycle in robots like Shakey is that by the time the (slow) thinking is finished, the resulting plan may fail because the world has changed in the meantime.

The subsumption architecture of behaviour-based robotics (Brooks, 1999, 2002) attempted to solve such problems by removing the classical sandwich; it was explicitly anti-representational. The logic of this radical move was that the world was its own best representation (Clark, 1997).

Behaviour-based robotics took advantage of Simon's (1969) parable of the ant, reducing costly and complex internal representations by recognizing that the external world is a critical contributor to behaviour. Why expend computational resources on the creation and maintenance of an internal model of the world, when externally the world was already present, open to being sensed and to being acted upon? Classical cognitive science's emphasis on internal representations and planning was a failure to take this parable to heart.

Interestingly, action was more important to earlier cognitive theories. Take, for example, Piaget's theory of cognitive development (Inhelder & Piaget, 1958, 1964; Piaget, 1970a, 1970b, 1972; Piaget & Inhelder, 1969). According to this theory, in their early teens children achieve the stage of formal operations. Formal operations describe adult-level cognitive abilities that are classical in the sense that they involve logical operations on symbolic representations. Formal operations involve completely abstract thinking, where relationships between propositions are considered.

However, Piagetian theory departs from classical cognitive science by including actions in the world. The development of formal operations begins with the sensorimotor stage, which involves direct interactions with objects in the world. In the next preoperational stage these objects are internalized as symbols. The preoperational stage is followed by concrete operations. When the child is in the stage of concrete operations, symbols are manipulated, but not in the abstract: concrete operations are applied to "manipulable objects (effective or immediately imaginable manipulations), in contrast to operations bearing on propositions or simple verbal statements (logic of propositions)" (Piaget, 1972, p. 56). In short, Piaget rooted fully representational or symbolic thought (i.e., formal operations) in the child's physical manipulation of his or her world. "The starting-point for the understanding, even of verbal concepts, is still the actions and operations of the subject" (Inhelder & Piaget, 1964, p. 284).

For example, classification and seriation (i.e., grouping and ordering entities) are operations that can be formally specified using logic or mathematics. One goal of Piagetian theory is to explain the development of such abstract competence. It does so by appealing to basic actions on the world experienced prior to the stage of formal

operations, "actions which are quite elementary: putting things in piles, separating piles into lots, making alignments, and so on" (Inhelder & Piaget, 1964, p. 291).

Other theories of cognitive development share the Piagetian emphasis on the role of the world, but elaborate the notion of what aspects of the world are involved (Vygotsky, 1986). Vygotsky (1986), for example, highlighted the role of social systems—a different conceptualization of the external world—in assisting cognitive development. Vygotsky used the term *zone of proximal development* to define the difference between a child's ability to solve problems without aid and their ability to solve problems when provided support or assistance. Vygotsky was strongly critical of instructional approaches that did not provide help to children as they solved problems.

Vygotsky (1986) recognized that sources of support for development were not limited to the physical world. He expanded the notion of worldly support to include social and cultural factors: "The true direction of the development of thinking is not from the individual to the social, but from the social to the individual" (p. 36). For example, to Vygotsky language was a tool for supporting cognition:

> Real concepts are impossible without words, and thinking in concepts does not exist beyond verbal thinking. That is why the central moment in concept formation, and its generative cause, is a specific use of words as functional 'tools.' (Vygotsky, 1986, p. 107)

Clark (1997, p. 45) wrote: "We may often solve problems by 'piggy-backing' on reliable environmental properties. This exploitation of external structure is what I mean by the term scaffolding." Cognitive scaffolding—the use of the world to support or extend thinking—is characteristic of theories in embodied cognitive science. Clark views scaffolding in the broad sense of a world or structure that descends from Vygotsky's theory:

> Advanced cognition depends crucially on our abilities to dissipate reasoning: to diffuse knowledge and practical wisdom through complex social structures, and to reduce the loads on individual brains by locating those brains in complex webs of linguistic, social, political, and institutional constraints. (Clark, 1997, p. 180)

While the developmental theories of Piaget and Vygotsky are departures from typical classical cognitive science in their emphasis on action and scaffolding, they are very traditional in other respects. American psychologist Sylvia Scribner pointed out that these two theorists, along with Newell and Simon, shared Aristotle's "preoccupation with modes of thought central to theoretical inquiry—with logical operations, scientific concepts, and problem solving in symbolic domains," maintaining "Aristotle's high esteem for theoretical thought and disregard for the practical" (Scribner & Tobach, 1997, p. 338).

Scribner's own work (Scribner & Tobach, 1997) was inspired by Vygotskian theory but aimed to extend its scope by examining practical cognition. Scribner described her research as the study of mind in action, because she viewed cognitive processes as being embedded with human action in the world. Scribner's studies analyzed "the characteristics of memory and thought as they function in the larger, purposive activities which cultures organize and in which individuals engage" (p. 384). In other words, the everyday cognition studied by Scribner and her colleagues provided ample evidence of cognitive scaffolding: "Practical problem solving is an open system that includes components lying outside the formal problem—objects and information in the environment and goals and interests of the problem solver" (pp. 334–335).

One example of Scribner's work on mind in action was the observation of problem-solving strategies exhibited by different types of workers at a dairy (Scribner & Tobach, 1997). It was discovered that a reliable difference between expert and novice dairy workers was that the former were more versatile in finding solutions to problems, largely because expert workers were much more able to exploit environmental resources. "The physical environment did not determine the problem-solving process but . . . was drawn into the process through worker initiative" (p. 377).

For example, one necessary job in the dairy was assembling orders. This involved using a computer printout of a wholesale truck driver's order for products to deliver the next day, to fetch from different areas in the dairy the required number of cases and partial cases of various products to be loaded onto the driver's truck. However, while the driver's order was placed in terms of individual units (e.g., particular numbers of quarts of skim milk, of half-pints of chocolate milk, and so on), the computer printout converted these individual units into "case equivalents." For example, one driver might require 20 quarts of skim milk. However, one case contains only 16 quarts. The computer printout for this part of the order would be 1 + 4, indicating one full case plus 4 additional units.

Scribner found differences between novice and expert product assemblers in the way in which these mixed numbers from the computer printout were converted into gathered products. Novice workers would take a purely mental arithmetic approach. As an example, consider the following protocol obtained from a novice worker:

> It was one case minus six, so there's two, four, six, eight, ten, sixteen (determines how many in a case, points finger as she counts). So there should be ten in here. Two, four, six, ten (counts units as she moves them from full to empty). One case minus six would be ten. (Scribner & Tobach, 1997, p. 302)

In contrast, expert workers were much more likely to scaffold this problem solving by working directly from the visual appearance of cases, as illustrated in a very different protocol:

> I walked over and I visualized. I knew the case I was looking at had ten out of it, and I only wanted eight, so I just added two to it. I don't never count when I'm making the order, I do it visual, a visual thing you know. (Scribner & Tobach, 1997, p. 303)

It was also found that expert workers flexibly alternated the distribution of scaffolded and mental arithmetic, but did so in a systematic way: when more mental arithmetic was employed, it was done to decrease the amount of physical exertion required to complete the order. This led to Scribner postulating a law of mental effort: "In product assembly, mental work will be expended to save physical work" (Scribner & Tobach, 1997, p. 348).

The law of mental effort was the result of Scribner's observation that expert workers in the dairy demonstrated marked diversity and flexibility in their solutions to work-related problems. Intelligent agents may be flexible in the manner in which they allocate resources between sense-act and sense-think-act processing. Both types of processes may be in play simultaneously, but they may be applied in different amounts when the same problem is encountered at different times and under different task demands (Hutchins, 1995).

Such flexible information processing is an example of *bricolage* (Lévi-Strauss, 1966). A *bricoleur* is an "odd job man" in France.

> The '*bricoleur*' is adept at performing a large number of diverse tasks; but, unlike the engineer, he does not subordinate each of them to the availability of raw materials and tools conceived and procured for the purpose of the project. His universe of instruments is closed and the rules of his game are always to make do with 'whatever is at hand.' (Lévi-Strauss, 1966, p. 17)

Bricolage seems well suited to account for the flexible thinking of the sort described by Scribner. Lévi-Strauss (1966) proposed *bricolage* as an alternative to formal, theoretical thinking, but cast it in a negative light: "The '*bricoleur*' is still someone who works with his hands and uses devious means compared to those of a craftsman" (pp. 16–17). Devious means are required because the *bricoleur* is limited to using only those components or tools that are at hand. "The engineer is always trying to make his way out of and go beyond the constraints imposed by a particular state of civilization while the '*bricoleur*' by inclination or necessity always remains within them" (p. 19).

Recently, researchers have renewed interest in *bricolage* and presented it in a more positive light than did Lévi-Strauss (Papert, 1980; Turkle, 1995). To Turkle (1995), *bricolage* was a sort of intuition, a mental tinkering, a dialogue mediated by

a virtual interface that was increasingly important with the visual GUIs of modern computing devices.

> As the computer culture's center of gravity has shifted from programming to dealing with screen simulations, the intellectual values of *bricolage* have become far more important. . . . Playing with simulation encourages people to develop the skills of the more informal soft mastery because it is so easy to run 'What if?' scenarios and tinker with the outcome. (Turkle, 1995, p. 52)

Papert (1980) argued that *bricolage* demands greater respect because it may serve as "a model for how scientifically legitimate theories are built" (p. 173).

The *bricolage* observed by Scribner and her colleagues when studying mind in action at the dairy revealed that practical cognition is flexibly and creatively scaffolded by an agent's environment. However, many of the examples reported by Scribner suggest that this scaffolding involves using the environment as an external representation or memory of a problem. That the environment can be used in this fashion, as an externalized extension of memory, is not surprising. Our entire print culture—the use of handwritten notes, the writing of books—has arisen from a technology that serves as an extension of memory (McLuhan, 1994, p. 189): "Print provided a vast new memory for past writings that made a personal memory inadequate."

However, the environment can also provide a more intricate kind of scaffolding. In addition to serving as an external store of information, it can also be exploited to manipulate its data. For instance, consider a naval navigation task in which a ship's speed is to be computed by measuring of how far the ship has travelled over a recent interval of time (Hutchins, 1995). An internal, representational approach to performing this computation would be to calculate speed based on internalized knowledge of algebra, arithmetic, and conversions between yards and nautical miles. However, an easier external solution is possible. A navigator is much more likely to draw a line on a three-scale representation called a nomogram. The top scale of this tool indicates duration, the middle scale indicates distance, and the bottom scale indicates speed. The user marks the measured time and distance on the first two scales, joins them with a straight line, and reads the speed from the intersection of this line with the bottom scale. Thus the answer to the problem isn't as much computed as it is inspected. "Much of the computation was done by the tool, or by its designer. The person somehow could succeed by doing less because the tool did more" (Hutchins, 1995, p. 151).

Classical cognitive science, in its championing of the representational theory of mind, demonstrates a modern persistence of the Cartesian distinction between mind and body. Its reliance on mental representation occurs at the expense of ignoring potential contributions of both an agent's body and world. Early representational theories were strongly criticized because of their immaterial nature.

For example, consider the work of Edward Tolman (1932, 1948). Tolman appealed to representational concepts to explain behaviour, such as his proposal that rats navigate and locate reinforcers by creating and manipulating a cognitive map. The mentalistic nature of Tolman's theories was a source of harsh criticism:

> Signs, in Tolman's theory, occasion in the rat realization, or cognition, or judgment, or hypotheses, or abstraction, but they do not occasion action. In his concern with what goes on in the rat's mind, Tolman has neglected to predict what the rat will do. So far as the theory is concerned the rat is left buried in thought; if he gets to the food-box at the end that is his concern, not the concern of the theory. (Guthrie, 1935, p. 172)

The later successes, and current dominance, of cognitive theory make such criticisms appear quaint. But classical theories are nonetheless being rigorously reformulated by embodied cognitive science.

Embodied cognitive scientists argue that classical cognitive science, with its emphasis on the disembodied mind, has failed to capture important aspects of thinking. For example, Hutchins (1995, p. 171) noted that "by failing to understand the source of the computational power in our interactions with simple 'unintelligent' physical devices, we position ourselves well to squander opportunities with so-called intelligent computers." Embodied cognitive science proposes that the modern form of dualism exhibited by classical cognitive science is a mistake. For instance, Scribner hoped that her studies of mind in action conveyed "a conception of mind which is not hostage to the traditional cleavage between the mind and the hand, the mental and the manual" (Scribner & Tobach, 1997, p. 307).

5.8 The Extended Mind

In preceding pages of this chapter, a number of interrelated topics that are central to embodied cognitive science have been introduced: situation and embodiment, feedback between agents and environments, stigmergic control of behaviour, affordances and enactive perception, and cognitive scaffolding. These topics show that embodied cognitive science places much more emphasis on body and world, and on sense and action, than do other "flavours" of cognitive science.

This change in emphasis can have profound effects on our definitions of *mind* or *self* (Bateson, 1972). For example, consider this famous passage from anthropologist Gregory Bateson:

> But what about 'me'? Suppose I am a blind man, and I use a stick. I go tap, tap, tap. Where do *I* start? Is my mental system bounded at the handle of the stick? Is it bounded by my skin? (Bateson, 1972, p. 465)

The embodied approach's emphasis on agents embedded in their environments leads to a radical and controversial answer to Bateson's questions, in the form of the extended mind (Clark, 1997, 1999, 2003, 2008; Clark & Chalmers, 1998; Menary, 2008, 2010; Noë, 2009; Rupert, 2009; Wilson, 2004, 2005). According to the extended mind hypothesis, the mind and its information processing are not separated from the world by the skull. Instead, the mind interacts with the world in such a way that information processing is both part of the brain and part of the world—the boundary between the mind and the world is blurred, or has disappeared.

Where is the mind located? The traditional view—typified by the classical approach introduced in Chapter 3—is that thinking is inside the individual, and that sensing and acting involve the world outside. However, if cognition is scaffolded, then some thinking has moved from inside the head to outside in the world. "It is the human brain *plus* these chunks of external scaffolding that finally constitutes the smart, rational inference engine we call mind" (Clark, 1997, p. 180). As a result, Clark (1997) described the mind as a leaky organ, because it has spread from inside our head to include whatever is used as external scaffolding.

The extended mind hypothesis has enormous implications for the cognitive sciences. The debate between classical and connectionist cognitive science does not turn on this issue, because both approaches are essentially representational. That is, both approaches tacitly endorse the classical sandwich; while they have strong disagreements about the nature of representational processes in the filling of the sandwich, neither of these approaches views the mind as being extended. Embodied cognitive scientists who endorse the extended mind hypothesis thus appear to be moving in a direction that strongly separates the embodied approach from the other two. It is small comfort to know that all cognitive scientists might agree that they are in the business of studying the mind, when they can't agree upon what minds are.

For this reason, the extended mind hypothesis has increasingly been a source of intense philosophical analysis and criticism (Adams & Aizawa, 2008; Menary, 2010; Robbins & Aydede, 2009). Adams and Aizawa (2008) are strongly critical of the extended mind hypothesis because they believe that it makes no serious attempt to define the "mark of the cognitive," that is, the principled differences between cognitive and non-cognitive processing:

> If just any sort of information processing is cognitive processing, then it is not hard to find cognitive processing in notebooks, computers and other tools. The problem is that this theory of the cognitive is wildly implausible and evidently not what cognitive psychologists intend. A wristwatch is an information processor, but not a cognitive agent. What the advocates of extended cognition need, but, we argue, do not have, is a plausible theory of the difference between the cognitive and

> the non-cognitive that does justice to the subject matter of cognitive psychology. (Adams & Aizawa, 2008, p. 11)

A variety of other critiques can be found in various contributions to Robbins and Aydede's (2009) *Cambridge Handbook of Situated Cognition*. Prinz made a pointed argument that the extended mind has nothing to contribute to the study of consciousness. Rupert noted how the notion of innateness poses numerous problems for the extended mind. Warneken and Tomasello examined cultural scaffolding, but they eventually adopted a position where these cultural tools have been internalized by agents. Finally, Bechtel presented a coherent argument from the philosophy of biology that there is good reason for the skull to serve as the boundary between the world and the mind. Clearly, the degree to which extendedness is adopted by situated researchers is far from universal.

In spite of the currently unresolved debate about the plausibility of the extended mind, the extended mind hypothesis is an idea that is growing in popularity in embodied cognitive science. Let us briefly turn to another implication that this hypothesis has for the practice of cognitive science.

The extended mind hypothesis is frequently applied to single cognitive agents. However, this hypothesis also opens the door to co-operative or public cognition in which a group of agents are embedded in a shared environment (Hutchins, 1995). In this situation, more than one cognitive agent can manipulate the world that is being used to support the information processing of other group members.

Hutchins (1995) provided one example of public cognition in his description of how a team of individuals is responsible for navigating a ship. He argued that "organized groups may have cognitive properties that differ from those of the individuals who constitute the group" (p. 228). For instance, in many cases it is very difficult to translate the heuristics used by a solo navigator into a procedure that can be implemented by a navigation team.

Collective intelligence—also called swarm intelligence or co-operative computing—is also of growing importance in robotics. Entomologists used the concept of the superorganism (Wheeler, 1911) to explain how entire colonies could produce more complex results (such as elaborate nests) than one would predict from knowing the capabilities of individual colony members. Swarm intelligence is an interesting evolution of the idea of the superorganism; it involves a collective of agents operating in a shared environment. Importantly, a swarm's components are only involved in local interactions with each other, resulting in many advantages (Balch & Parker, 2002; Sharkey, 2006).

For instance, a computing swarm is scalable—it may comprise varying numbers of agents, because the same control structure (i.e., local interactions) is used regardless of how many agents are in the swarm. For the same reason, a computing swarm is flexible: agents can be added or removed from the swarm without

reorganizing the entire system. The scalability and flexibility of a swarm make it robust, as it can continue to compute when some of its component agents no longer function properly. Notice how these advantages of a swarm of agents are analogous to the advantages of connectionist networks over classical models, as discussed in Chapter 4.

Nonlinearity is also a key ingredient of swarm intelligence. For a swarm to be considered intelligent, the whole must be greater than the sum of its parts. This idea has been used to identify the presence of swarm intelligence by relating the amount of work done by a collective to the number of agents in the collection (Beni & Wang, 1991). If the relationship between work accomplished and number of agents is linear, then the swarm is not considered to be intelligent. However, if the relationship is nonlinear—for instance, exponentially increasing—then swarm intelligence is present. The nonlinear relationship between work and numbers may itself be mediated by other nonlinear relationships. For example, Dawson, Dupuis, and Wilson (2010) found that in collections of simple LEGO robots, the presence of additional robots influenced robot paths in an arena in such a way that a sorting task was accomplished far more efficiently.

While early studies of robot collectives concerned small groups of homogenous robots (Gerkey & Mataric, 2004), researchers are now more interested in complex collectives consisting of different types of machines for performing diverse tasks at varying locations or times (Balch & Parker, 2002; Schultz & Parker, 2002). This leads to the problem of coordinating the varying actions of diverse collective members (Gerkey & Mataric, 2002, 2004; Mataric, 1998). One general approach to solving this coordination problem is intentional co-operation (Balch & Parker, 2002; Parker, 1998, 2001), which uses direct communication amongst robots to prevent unnecessary duplication (or competition) between robot actions. However, intentional co-operation comes with its own set of problems. For instance, communication between robots is costly, particularly as more robots are added to a communicating team (Kube & Zhang, 1994). As well, as communication makes the functions carried out by individual team members more specialized, the robustness of the robot collective is jeopardized (Kube & Bonabeau, 2000). Is it possible for a robot collective to coordinate its component activities, and solve interesting problems, in the absence of direction communication?

The embodied approach has generated a plausible answer to this question via stigmergy (Kube & Bonabeau, 2000). Kube and Bonabeau (2000) demonstrated that the actions of a large collective of robots could be stigmergically coordinated so that the collective could push a box to a goal location in an arena. Robots used a variety of sensors to detect (and avoid) other robots, locate the box, and locate the goal location. A subsumption architecture was employed to instantiate a fairly simple set of sense-act reflexes. For instance, if a robot detected that is was in contact with

the box and could see the goal, then box-pushing behaviour was initiated. If it was in contact with the box but could not see the goal, then other movements were triggered, resulting in the robot finding contact with the box at a different position.

This subsumption architecture caused robots to seek the box, push it towards the goal, and do so co-operatively by avoiding other robots. Furthermore, when robot activities altered the environment, this produced corresponding changes in behaviour of other robots. For instance, a robot pushing the box might lose sight of the goal because of box movement, and it would therefore leave the box and use its other exploratory behaviours to come back to the box and push it from a different location. "Cooperation in some tasks is possible without direct communication" (Kube & Bonabeau, 2000, p. 100). Importantly, the solution to the box-pushing problem required such co-operation, because the box being manipulated was too heavy to be moved by a small number of robots!

The box-pushing research of Kube and Bonabeau (2000) is an example of stigmergic processing that occurs when two or more individuals collaborate on a task using a shared environment. Hutchins (1995) brought attention to less obvious examples of public cognition that exploit specialized environmental tools. Such scaffolding devices cannot be dissociated from culture or history. For example, Hutchins noted that navigation depends upon centuries-old mathematics of chart projections, not to mention millennia-old number systems.

These observations caused Hutchins (1995) to propose an extension of Simon's (1969) parable of the ant. Hutchins argued that rather than watching an individual ant on the beach, we should arrive at a beach after a storm and watch generations of ants at work. As the ant colony matures, the ants will appear smarter, because their behaviours are more efficient. But this is because,

> the environment is not the same. Generations of ants have left their marks on the beach, and now a dumb ant has been made to appear smart through its simple interaction with the residua of the history of its ancestor's actions. (Hutchins, 1995, p. 169)

Hutchins' (1995) suggestion mirrored concerns raised by Scribner's studies of mind in action. She observed that the diversity of problem solutions generated by dairy workers, for example, was due in part to social scaffolding.

> We need a greater understanding of the ways in which the institutional setting, norms and values of the work group and, more broadly, cultural understandings of labor contribute to the reorganization of work tasks in a given community. (Scribner & Tobach, 1997, p. 373)

Furthermore, Scribner pointed out that the traditional methods used by classical researchers to study cognition were not suited for increasing this kind of

understanding. The extended mind hypothesis leads not only to questions about the nature of mind, but also to the questions about the methods used to study mentality.

5.9 The Roots of Forward Engineering

The most typical methodology to be found in classical cognitive science is reverse engineering. Reverse engineering involves observing the behaviour of an intact system in order to infer the nature and organization of the system's internal processes. Most cognitive theories are produced by using a methodology called functional analysis (Cummins, 1975, 1983), which uses experimental results to iteratively carve a system into a hierarchy of functional components until a basic level of subfunctions, the cognitive architecture, is reached.

A practical problem with functional analysis or reverse engineering is the frame of reference problem (Pfeifer & Scheier, 1999). This problem arises during the distribution of responsibility for the complexity of behaviour between the internal processes of an agent and the external influences of its environment. Classical cognitive science, a major practitioner of functional analysis, endorses the classical sandwich; its functional analyses tend to attribute behavioural complexity to the internal processes of an agent, while at the same time ignoring potential contributions of the environment. In other words, the frame of reference problem is to ignore Simon's (1969) parable of the ant.

Embodied cognitive scientists frequently adopt a different methodology, forward engineering. In forward engineering, a system is constructed from a set of primitive functions of interest. The system is then observed to determine whether it generates surprising or complicated behaviour. "Only about 1 in 20 'gets it'—that is, the idea of thinking about psychological problems by inventing mechanisms for them and then trying to see what they can and cannot do" (Minsky, personal communication, 1995). This approach has also been called synthetic psychology (Braitenberg, 1984). Reverse engineers collect data to create their models; in contrast, forward engineers build their models first and use them as primary sources of data (Dawson, 2004).

We noted in Chapter 3 that classical cognitive science has descended from the seventeenth-century rationalist philosophy of René Descartes (1960, 1996). It was observed in Chapter 4 that connectionist cognitive science descended from the early eighteenth-century empiricism of John Locke (1977), which was itself a reaction against Cartesian rationalism. The synthetic approach seeks "understanding by building" (Pfeifer & Scheier, 1999), and as such permits us to link embodied cognitive science to another eighteenth-century reaction against Descartes, the philosophy of Giambattista Vico (Vico, 1990, 1988, 2002).

Vico based his philosophy on the analysis of word meanings. He argued that the Latin term for truth, *verum*, had the same meaning as the Latin term *factum*, and therefore concluded that "it is reasonable to assume that the ancient sages of Italy entertained the following beliefs about the true: 'the true is precisely what is made'" (Vico, 1988, p. 46). This conclusion led Vico to his argument that humans could only understand the things that they made, which is why he studied societal artifacts, such as the law.

Vico's work provides an early motivation for forward engineering: "To know (*scire*) is to put together the elements of things" (Vico, 1988, p. 46). Vico's account of the mind was a radical departure from Cartesian disembodiment. To Vico, the Latins "thought every work of the mind was sense; that is, whatever the mind does or undergoes derives from contact with bodies" (p. 95). Indeed, Vico's *verum-factum* principle is based upon embodied mentality. Because the mind is "immersed and buried in the body, it naturally inclines to take notice of bodily things" (p. 97).

While the philosophical roots of forward engineering can be traced to Vico's eighteenth-century philosophy, its actual practice—as far as cognitive science is concerned—did not emerge until cybernetics arose in the 1940s. One of the earliest examples of synthetic psychology was the Homeostat (Ashby, 1956, 1960), which was built by cyberneticist William Ross Ashby in 1948. The Homeostat was a system that changed its internal states to maximize stability amongst the interactions between its internal components and the environment. William Grey Walter (1963, p. 123) noted that it was "like a fireside cat or dog which only stirs when disturbed, and then methodically finds a comfortable position and goes to sleep again."

Ashby's (1956, 1960) Homeostat illustrated the promise of synthetic psychology. The feedback that Ashby was interested in could not be analyzed mathematically; it was successfully studied synthetically with Ashby's device. Remember, too, that when the Homeostat was created, computer simulations of feedback were still in the future.

As well, it was easier to produce interesting behaviour in the Homeostat than it was to analyze it. This is because the secret to its success was a large number of potential internal states, which provided many degrees of freedom for producing stability. At the same time, this internal variability was an obstacle to traditional analysis. "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'" (Grey Walter, 1963, p. 124).

Concerns about this characteristic of the Homeostat inspired the study of the first autonomous robots, created by cyberneticist William Grey Walter (1950a, 1950b, 1951, 1963). The first two of these machines were constructed in 1948 (de Latil, 1956); comprising surplus war materials, their creation was clearly an act of *bricolage*. "The first model of this species was furnished with pinions from

old clocks and gas meters" (Grey Walter, 1963, p. 244). By 1951, these two had been replaced by six improved machines (Holland, 2003a), two of which are currently displayed in museums.

The robots came to be called Tortoises because of their appearance: they seemed to be toy tractors surrounded by a tortoise-like shell. Grey Walter viewed them as an artificial life form that he classified as *Machina speculatrix*. *Machina speculatrix* was a reaction against the internal variability in Ashby's Homeostat. The goal of Grey Walter's robotics research was to explore the degree to which one could produce complex behaviour from such very simple devices (Boden, 2006). When Grey Walter modelled behaviour he "was determined to wield Occam's razor. That is, he aimed to posit as simple a mechanism as possible to explain apparently complex behaviour. And simple, here, meant simple" (Boden, 2006, p. 224). Grey Walter restricted a Tortoise's internal components to "two functional elements: two miniature radio tubes, two sense organs, one for light and the other for touch, and two effectors or motors, one for crawling and the other for steering" (Grey Walter, 1950b, p. 43).

The interesting behaviour of the Tortoises was a product of simple reflexes that used detected light (via a light sensor mounted on the robot's steering column) and obstacles (via movement of the robot's shell) to control the actions of the robot's two motors. Light controlled motor activity as follows. In dim light, the Tortoise's drive motor would move the robot forward, while the steering motor slowly turned the front wheel. Thus in dim light the Tortoise "explored." In moderate light, the drive motor continued to run, but the steering motor stopped. Thus in moderate light the Tortoise "approached." In bright light, the drive motor continued to run, but the steering motor ran at twice the normal speed, causing marked oscillatory movements. Thus in bright light the Tortoise "avoided."

The motors were affected by the shell's sense of touch as follows. When the Tortoise's shell was moved by an obstacle, an oscillating signal was generated that first caused the robot to drive fast while slowly turning, and then to drive slowly while quickly turning. The alternation of these behaviours permitted the Tortoise to escape from obstacles. Interestingly, when movement of the Tortoise shell triggered such behaviour, signals from the photoelectric cell were rendered inoperative for a few moments. Thus Grey Walter employed a simple version of what later would be known as Brooks' (1999) subsumption architecture: a higher layer of touch processing could inhibit a lower layer of light processing.

In accordance with forward engineering, after Grey Walter constructed his robots, he observed their behaviour by recording the paths that they took in a number of simple environments. He preserved a visual record of their movement by using time-lapse photography; because of lights mounted on the robots, their paths were literally traced on each photograph (Holland, 2003b). Like the paths on the beach

traced in Simon's (1969) parable of the ant, the photographs recorded Tortoise behaviour that was "remarkably unpredictable" (Grey Walter, 1950b, p. 44).

Grey Walter observed the behaviours of his robots in a number of different environments. For example, in one study the robot was placed in a room where a light was hidden from view by an obstacle. The Tortoise began to explore the room, bumped into the obstacle, and engaged in its avoidance behaviour. This in turn permitted the robot to detect the light, which it approached. However, it didn't collide with the light. Instead the robot circled it cautiously, veering away when it came too close. "Thus the machine can avoid the fate of the moth in the candle" (Grey Walter, 1963, p. 128).

When the environment became more complicated, so too did the behaviours produced by the Tortoise. If the robot was confronted with two stimulus lights instead of one, it would first be attracted to one, which it circled, only to move away and circle the other, demonstrating an ability to choose: it solved 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). If a mirror was placed in its environment, the mirror served as an obstacle, but it reflected the light mounted on the robot, which was an attractant. The resulting dynamics produced the so-called "mirror dance" in which the robot,

> lingers before a mirror, flickering, twittering and jigging like a clumsy Narcissus. The behaviour 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)

In less controlled or open-ended environments, the behaviour that was produced was lifelike in its complexity. The Tortoises produced "the exploratory, speculative behaviour that is so characteristic of most animals" (Grey Walter, 1950b, p. 43). Examples of such behaviour were recounted by cyberneticist Pierre de Latil (1956):

> Elsie moved to and fro just like a real animal. A kind of head at the end of a long neck towered over the shell, like a lighthouse on a promontory and, like a lighthouse; it veered round and round continuously. (de Latil, 1956, p. 209)

The *Daily Mail* reported that,

> the toys possess the senses of sight, hunger, touch, and memory. They can walk about the room avoiding obstacles, stroll round the garden, climb stairs, and feed themselves by automatically recharging six-volt accumulators from the light in the room. And they can dance a jig, go to sleep when tired, and give an electric shock if disturbed when they are not playful. (Holland, 2003a, p. 2090)

Grey Walter released the Tortoises to mingle with the audience at a 1955 meeting of

the British Association (Hayward, 2001): "The tortoises, with their in-built attraction towards light, moved towards the pale stockings of the female delegates whilst avoiding the darker legs of the betrousered males" (p. 624).

Grey Walter was masterfully able to promote his work to the general public (Hayward, 2001; Holland, 2003a). However, he worried that public reception of his machines would decrease their scientific importance. History has put such concerns to rest; Grey Walter's pioneering research has influenced many modern researchers (Reeve & Webb, 2003). Grey Walter's,

> ingenious devices were seriously intended as working models for understanding biology: a 'mirror for the brain' that could both generally enrich our understanding of principles of behavior (such as the complex outcome of combining simple tropisms) and be used to test specific hypotheses (such as Hebbian learning). (Reeve & Webb, 2003, p. 2245)

5.10 Reorientation without Representation

The robotics work of Grey Walter has been accurately described as an inspiration to modern studies of autonomous systems (Reeve & Webb, 2003). Indeed, the kind of research conducted by Grey Walter seems remarkably similar to the "new wave" of behaviour-based or biologically inspired robotics (Arkin, 1998; Breazeal, 2002; Sharkey, 1997; Webb & Consi, 2001).

In many respects, this represents an important renaissance of Grey Walter's search for "mimicry of life" (Grey Walter, 1963, p. 114). Although the Tortoises were described in his very popular 1963 book *The Living Brain*, they essentially disappeared from the scientific picture for about a quarter of a century. Grey Walter was involved in a 1970 motorcycle accident that ended his career; after this accident, the whereabouts of most of the Tortoises was lost. One remained in the possession of his son after Grey Walter's death in 1977; it was located in 1995 after an extensive search by Owen Holland. This discovery renewed interest in Grey Walter's work (Hayward, 2001; Holland, 2003a, 2003b), and has re-established its important place in modern research.

The purpose of the current section is to briefly introduce one small segment of robotics research that has descended from Grey Walter's pioneering work. In Chapter 3, we introduced the reorientation task that is frequently used to study how geometric and feature cues are used by an agent to navigate through its world. We also described a classical theory, the geometric module (Cheng, 1986; Gallistel, 1990), which has been used to explain some of the basic findings concerning this task. In Chapter 4, we noted that the reorientation task has also been approached from the perspective of connectionist cognitive science. A simple

artificial neural network, the perceptron, has been offered as a viable alternative to classical theory (Dawson et al., 2010). In this section we briefly describe a third approach to the reorientation task, because embodied cognitive science has studied it in the context of behaviour-based robotics.

Classical and connectionist cognitive science provide very different accounts of the co-operative and competitive interactions between geometric and featural cues when an agent attempts to relocate the target location in a reorientation arena. However, these different accounts are both representational. One of the themes pervading embodied cognitive science is a reaction against representational explanations of intelligent behaviour (Shapiro, 2011). One field that has been a test bed for abandoning internal representations is known as new wave robotics (Sharkey, 1997).

New wave roboticists strive to replace representation with reaction (Brooks, 1999), to use sense-act cycles in the place of representational sense-think-act processing. This is because "embodied and situated systems can solve rather complicated tasks without requiring internal states or internal representations" (Nolfi & Floreano, 2000, p. 93). One skill that has been successfully demonstrated in new wave robotics is navigation in the context of the reorientation task (Lund & Miglino, 1998).

The Khepera robot (Bellmore & Nemhauser, 1968; Boogaarts, 2007) is a standard platform for the practice of new wave robotics. It has the appearance of a motorized hockey puck, uses two motor-driven wheels to move about, and has eight sensors distributed around its chassis that allow it to detect the proximity of obstacles. Roboticists have the goal of combining the proximity detector signals to control motor speed in order to produce desired dynamic behaviour. One approach to achieving this goal is to employ evolutionary robotics (Nolfi & Floreano, 2000). Evolutionary robotics involves using a genetic algorithm (Holland, 1992; Mitchell, 1996) to find a set of weights between each proximity detector and each motor.

In general, evolutionary robotics proceeds as follows (Nolfi & Floreano, 2000). First, a fitness function is defined, to evaluate the quality of robot performance. Evolution begins with an initial population of different control systems, such as different sets of sensor-to-motor weights. The fitness function is used to assess each of these control systems, and those that produce higher fitness values "survive." Survivors are used to create the next generation of control systems via prescribed methods of "mutation." The whole process of evaluate-survive-mutate is iterated; average fitness is expected to improve with each new generation. The evolutionary process ends when improvements in fitness stabilize. When evolution stops, the result is a control system that should be quite capable of performing the task that was evaluated by the fitness function.

Lund and Miglino (1998) used this procedure to evolve a control system that enabled Khepera robots to perform the reorientation task in a rectangular arena

without feature cues. Their goal was to see whether a standard result—rotational error—could be produced in an agent that did not employ the geometric module, and indeed which did not represent arena properties at all. Lund and Miglino's fitness function simply measured a robot's closeness to the goal location. After 30 generations of evolution, they produced a system that would navigate a robot to the goal location from any of 8 different starting locations with a 41 percent success rate. Their robots also produced rotational error, for they incorrectly navigated to the corner 180° from the goal in another 41 percent of the test trials. These results were strikingly similar to those observed when rats perform reorientation in featureless rectangular arenas (e.g., Gallistel, 1990).

Importantly, the control system that was evolved by Lund and Miglino (1998) was simply a set of weighted connections between proximity detectors and motors, and not an encoding of arena shape.

> The geometrical properties of the environment can be assimilated in the sensory-motor schema of the robot behavior without any explicit representation. In general, our work, in contrast with traditional cognitive models, shows how environmental knowledge can be reached without any form of direct representation. (Lund and Miglino, 1998, p. 198)

If arena shape is not explicitly represented, then how does the control system developed by Lund and Miglino (1998) produce reorientation task behaviour? When the robot is far enough from the arena walls that none of the sensors are detecting an obstacle, the controller weights are such that the robot moves in a gentle curve to the left. As a result, it never encounters a short wall when it leaves from any of its eight starting locations! When a long wall is (inevitably) encountered, the robot turns left and follows the wall until it stops in a corner. The result is that the robot will be at either the target location or its rotational equivalent.

The control system evolved by Lund and Miglino (1998) is restricted to rectangular arenas of a set size. If one of their robots is placed in an arena of even a slightly different size, its performance suffers (Nolfi, 2002). Nolfi used a much longer evolutionary process (500 generations), and also placed robots in different sized arenas, to successfully produce devices that would generate typical results not only in a featureless rectangular arena, but also in arenas of different dimensions. Again, these robots did so without representing arena shape or geometry.

Nolfi's (2002) more general control system worked as follows. His robots would begin by moving forwards and avoiding walls, which would eventually lead them into a corner. When facing a corner, signals from the corner's two walls caused the robot to first turn to orient itself at an angle of 45° from one of the corner's walls. Then the robot would make an additional turn that was either clockwise or counterclockwise, depending upon whether the sensed wall was to the robot's left or the right.

The final turn away from the corner necessarily pointed the robot in a direction that would cause it to follow a long wall, because sensing a wall at 45° is an indirect measurement of wall length:

> If the robot finds a wall at about 45° on its left side and it previously left a corner, it means that the actual wall is one of the two longer walls. Conversely, if it encounters a wall at 45° on its right side, the actual wall is necessarily one of the two shorter walls. What is interesting is that the robot "measures" the relative length of the walls through action (i.e., by exploiting sensory–motor coordination) and it does not need any internal state to do so. (Nolfi, 2002, p. 141)

As a result, the robot sensed the long wall in a rectangular arena without representing wall length. It followed the long wall, which necessarily led the robot to either the goal corner or the corner that results in a rotational error, regardless of the actual dimensions of the rectangular arena.

Robots simpler than the Khepera can also perform the reorientation task, and they can at the same time generate some of its core results. The subsumption architecture has been used to design a simple LEGO robot, antiSLAM (Dawson, Dupuis, & Wilson, 2010), that demonstrates rotational error and illustrates how a new wave robot can combine geometric and featural cues, an ability not included in the evolved robots that have been discussed above.

The ability of autonomous robots to navigate is fundamental to their success. In contrast to the robots described in the preceding paragraphs, one of the major approaches to providing such navigation is called SLAM, which is an acronym for a representational approach named "simultaneous localization and mapping" (Jefferies & Yeap, 2008). Representationalists assumed that agents navigate their environment by sensing their current location and referencing it on some internal map. How is such navigation to proceed if an agent is placed in a novel environment for which no such map exists? SLAM is an attempt to answer this question. It proposes methods that enable an agent to build a new map of a novel environment and at the same time use this map to determine the agent's current location.

The representational assumptions that underlie approaches such as SLAM have recently raised concerns in some researchers who study animal navigation (Alerstam, 2006). To what extent might a completely reactive, sense-act robot be capable of demonstrating interesting navigational behaviour? The purpose of antiSLAM (Dawson, Dupuis, & Wilson, 2010) was to explore this question in an incredibly simple platform—the robot's name provides some sense of the motivation for its construction.

AntiSLAM is an example of a Braitenberg Vehicle 3 (Braitenberg, 1984), because it uses six different sensors, each of which contributes to the speed of two motors that propel and steer it. Two are ultrasonic sensors that are used as sonar to detect obstacles, two are rotation detectors that are used to determine when the

robot has stopped moving, and two are light sensors that are used to attract the robot to locations of bright illumination. The sense-act reflexes of antiSLAM were not evolved but were instead created using the subsumption architecture.

The lowest level of processing in antiSLAM is "drive," which essentially uses the outputs of the ultrasonic sensors to control motor speed. The closer to an obstacle a sensor gets, the slower is the speed of the one motor that the sensor helps to control. The next level is "escape." When both rotation sensors are signaling that the robot is stationary (i.e., stopped by an obstacle detected by both sensors), the robot executes a turn to point itself in a different direction. The next level up is "wall following": motor speed is manipulated in such a way that the robot has a strong bias to keep closer to a wall on the right than to a wall on the left. The highest level is "feature," which uses two light sensors to contribute to motor speed in such a way that it approaches areas of brighter light.

AntiSLAM performs complex, lifelike exploratory behaviour when placed in general environments. It follows walls, steers itself around obstacles, explores regions of brighter light, and turns around and escapes when it finds itself stopped in a corner or in front of a large obstacle.

When placed in a reorientation task arena, antiSLAM generates behaviours that give it the illusion of representing geometric and feature cues (Dawson, Dupuis, & Wilson, 2010). It follows walls in a rectangular arena, slowing to a halt when enters a corner. It then initiates a turning routine to exit the corner and continue exploring. Its light sensors permit it to reliably find a target location that is associated with particular geometric and local features. When local features are removed, it navigates the arena using geometric cues only, and it produces rotational errors. When local features are moved (i.e., an incorrect corner is illuminated), its choice of locations from a variety of starting points mimics the same combination of geometric and feature cues demonstrated in experiments with animals. In short, it produces some of the key features of the reorientation task—however, it does so without creating a cognitive map, and even without representing a goal. Furthermore, observations of antiSLAM's reorientation task behaviour indicated that a crucial behavioural measure, the path taken by an agent as it moves through the arena, is critical. Such paths are rarely reported in studies of reorientation.

The reorienting robots discussed above are fairly recent descendants of Grey Walter's (1963) Tortoises, but their more ancient ancestors are the eighteenth-century life-mimicking, clockwork automata (Wood, 2002). These devices brought into sharp focus the philosophical issues concerning the comparison of man and machine that was central to Cartesian philosophy (Grenville, 2001; Wood, 2002). Religious tensions concerning the mechanistic nature of man, and the spiritual nature of clockwork automata, were soothed by dualism: automata and animals

were machines. Men too were machines, but unlike automata, they also had souls. It was the appearance of clockwork automata that led to their popularity, as well as to their conflicts with the church. "Until the scientific era, what seemed most alive to people was what most *looked* like a living being. The vitality accorded to an object was a function primarily of its form" (Grey Walter, 1963, p. 115).

In contrast, Grey Walter's Tortoises were not attempts to reproduce appearances, but were instead simulations of more general and more abstract abilities central to biological agents,

> exploration, curiosity, free-will in the sense of unpredictability, goal-seeking, self-regulation, avoidance of dilemmas, foresight, memory, learning, forgetting, association of ideas, form recognition, and the elements of social accommodation. Such is life. (Grey Walter, 1963, p. 120)

By situating and embodying his machines, Grey Walter invented a new kind of scientific tool that produced behaviours that were creative and unpredictable, governed by nonlinear relationships between internal mechanisms and the surrounding, dynamic world.

Modern machines that mimic lifelike behaviour still raise serious questions about what it is to be human. To Wood (2002, p. xxvii) all automata were presumptions "that life can be simulated by art or science or magic. And embodied in each invention is a riddle, a fundamental challenge to our perception of what makes us human." The challenge is that if the lifelike behaviours of the Tortoises and their descendants are merely feedback loops between simple mechanisms and their environments, then might the same be true of human intelligence?

This challenge is reflected in some of roboticist Rodney Brooks' remarks in Errol Morris' 1997 documentary *Fast, Cheap & Out of Control*. Brooks begins by describing one of his early robots: "To an observer it appears that the robot has intentions and it has goals and it is following people and chasing prey. But it's just the interaction of lots and lots of much simpler processes." Brooks then considers extending this view to human cognition: "Maybe that's all there is. Maybe a lot of what humans are doing could be explained this way."

But as the segment in the documentary proceeds, Brooks, the pioneer of behaviour-based robotics, is reluctant to believe that humans are similar types of devices:

> When I think about it, I can almost see myself as being made up of thousands and thousands of little agents doing stuff almost independently. But at the same time I fall back into believing the things about humans that we all believe about humans and living life that way. Otherwise I analyze it too much; life becomes almost meaningless. (Morris, 1997)

Conflicts like those voiced by Brooks are brought to the forefront when embodied

cognitive science ventures to study humanoid robots that are designed to exploit social environments and interactions (Breazeal, 2002; Turkle, 2011).

5.11 Robotic Moments in Social Environments

The embodied approach has long recognized that an agent's environment is much more that a static array of stimuli (Gibson, 1979; Neisser, 1976; Scribner & Tobach, 1997; Vygotsky, 1986). "The richest and most elaborate affordances of the environment are provided by other animals and, for us, other people" (Gibson, 1979, p. 135). A social environment is a rich source of complexity and ranges from dynamic interactions with other agents to cognitive scaffolding provided by cultural conventions. "All higher mental processes are primarily social phenomena, made possible by cognitive tools and characteristic situations that have evolved in the course of history" (Neisser, 1976, p. 134).

In the most basic sense of *social*, multiple agents in a shared world produce a particularly complex source of feedback between each other's actions. "What the other animal affords the observer is not only behaviour but also social interaction. As one moves so does the other, the one sequence of action being suited to the other in a kind of behavioral loop" (Gibson, 1979, p. 42).

Grey Walter (1963) explored such behavioural loops when he placed two Tortoises in the same room. Mounted lights provided particularly complex stimuli in this case, because robot movements would change the position of the two lights, which in turn altered subsequent robot behaviours. In describing a photographic record of one such interaction, Grey Walter called the social dynamics of his machines,

> the formation of a cooperative and a competitive society. . . . When the two creatures are released at the same time in the dark, each is attracted by the other's headlight but each in being attracted extinguishes the source of attraction to the other. The result is a stately circulating movement of minuet-like character; whenever the creatures touch they become obstacles and withdraw but are attracted again in rhythmic fashion. (Holland, 2003a, p. 2104)

Similar behavioural loops have been exploited to explain the behaviour of larger collections of interdependent agents, such as flocks of flying birds or schools of swimming fish (Nathan & Barbosa, 2008; Reynolds, 1987). Such an aggregate presents itself as another example of a superorganism, because the synchronized movements of flock members give "the strong impression of intentional, centralized control" (Reynolds, 1987, p. 25). However, this impression may be the result of local, stigmergic interactions in which an environment chiefly consists of other flock members in an agent's immediate vicinity.

In his pioneering work on simulating the flight of a flock of artificial birds, called boids, Reynolds (1987) created lifelike flocking behaviour by having each independently flying boid adapt its trajectory according to three simple rules: avoid collision with nearby flock mates, match the velocity of nearby flock mates, and stay close to nearby flock mates. A related model (Couzin et al., 2005) has been successfully used to predict movement of human crowds (Dyer et al., 2008; Dyer et al., 2009; Faria et al., 2010).

However, many human social interactions are likely more involved than the simple behavioural loops that defined the social interactions amongst Grey Walter's (1963) Tortoises or the flocking behaviour of Reynolds' (1987) boids. These interactions are possibly still behavioural loops, but they may be loops that involve processing special aspects of the social environment. This is because it appears that the human brain has a great deal of neural circuitry devoted to processing specific kinds of social information.

Social cognition is fundamentally involved with how we understand others (Lieberman, 2007). One key avenue to such understanding is our ability to use and interpret facial expressions (Cole, 1998; Etcoff & Magee, 1992). There is a long history of evidence that indicates that our brains have specialized circuitry for processing faces. Throughout the eighteenth and nineteenth centuries, there were many reports of patients whose brain injuries produced an inability to recognize faces but did not alter the patients' ability to identify other visual objects. This condition was called prosopagnosia, for "face blindness," by German neuroscientist Joachim Bodamer in a famous 1947 manuscript (Ellis & Florence, 1990). In the 1980s, recordings from single neurons in the monkey brain revealed cells that appeared to be tailored to respond to specific views of monkey faces (Perrett, Mistlin, & Chitty, 1987; Perrett, Rolls, & Caan, 1982). At that time, though, it was unclear whether analogous neurons for face processing were present in the human brain.

Modern brain imaging techniques now suggest that the human brain has an elaborate hierarchy of co-operating neural systems for processing faces and their expressions (Haxby, Hoffman, & Gobbini, 2000, 2002). Haxby, Hoffman, and Gobbini (2000, 2002) argue for the existence of multiple, bilateral brain regions involved in different face perception functions. Some of these are core systems that are responsible for processing facial invariants, such as relative positions of the eyes, nose, and mouth, which are required for recognizing faces. Others are extended systems that process dynamic aspects of faces in order to interpret, for instance, the meanings of facial expressions. These include subsystems that co-operatively account for lip reading, following gaze direction, and assigning affect to dynamic changes in expression.

Facial expressions are not the only source of social information. Gestures and actions, too, are critical social stimuli. Evidence also suggests that mirror neurons in

the human brain (Gallese et al., 1996; Iacoboni, 2008; Rizzolatti & Craighero, 2004; Rizzolatti, Fogassi, & Gallese, 2006) are specialized for both the generation and interpretation of gestures and actions.

Mirror neurons were serendipitously discovered in experiments in which motor neurons in region F5 were recorded when monkeys performed various reaching actions (Di Pellegrino et al., 1992). By accident, it was discovered that many of the neurons that were active when a monkey performed an action also responded when similar actions were *observed* being performed by another:

> After the initial recording experiments, we incidentally observed that some experimenter's actions, such as picking up the food or placing it inside the testing box, activated a relatively large proportion of F5 neurons in the absence of any overt movement of the monkey. (Di Pellegrino et al., 1992, p. 176)

The chance discovery of mirror neurons has led to an explosion of research into their behaviour (Iacoboni, 2008). It has been discovered that when the neurons fire, they do so for the entire duration of the observed action, not just at its onset. They are grasp specific: some respond to actions involving precision grips, while others respond to actions involving larger objects. Some are broadly tuned, in the sense that they will be triggered when a variety of actions are observed, while others are narrowly tuned to specific actions. All seem to be tuned to object-oriented action: a mirror neuron will respond to a particular action on an object, but it will fail to respond to the identical action if no object is present.

While most of the results described above were obtained from studies of the monkey brain, there is a steadily growing literature indicating that the human brain also has a mirror system (Buccino et al., 2001; Iacoboni, 2008).

Mirror neurons are not solely concerned with hand and arm movements. For instance, some monkey mirror neurons respond to mouth movements, such as lip smacking (Ferrari et al., 2003). Similarly, the human brain has a mirror system for the act of touching (Keysers et al., 2004). Likewise, another part of the human brain, the insula, may be a mirror system for emotion (Wicker et al., 2003). For example, it generates activity when a subject experiences disgust, and also when a subject observes the facial expressions of someone else having a similar experience.

Two decades after its discovery, extensive research on the mirror neuron system has led some researchers to claim that it provides the neural substrate for social cognition and imitative learning (Gallese & Goldman, 1998; Gallese, Keysers, & Rizzolatti, 2004; Iacoboni, 2008), and that disruptions of this system may be responsible for autism (Williams et al., 2001). The growing understanding of the mirror system and advances in knowledge about the neuroscience of face perception have heralded a new interdisciplinary research program, called social cognitive neuroscience (Blakemore, Winston, & Frith, 2004; Lieberman, 2007; Ochsner & Lieberman, 2001).

> It may once have seemed foolhardy to work out connections between fundamental neurophysiological mechanisms and highly complex social behaviour, let alone to decide whether the mechanisms are specific to social processes. However . . . neuroimaging studies have provided some encouraging examples. (Blakemore, Winston, & Frith, 2004, p. 216)

The existence of social cognitive neuroscience is a consequence of humans evolving, embodied and situated, in a social environment that includes other humans and their facial expressions, gestures, and actions. The modern field of sociable robotics (Breazeal, 2002) attempts to develop humanoid robots that are also socially embodied and situated. One purpose of such robots is to provide a medium for studying human social cognition via forward engineering.

A second, applied purpose of sociable robotics is to design robots to work co-operatively with humans by taking advantage of a shared social environment. Breazeal (2002) argued that because the human brain has evolved to be expert in social interaction, "if a technology behaves in a socially competent manner, we evoke our evolved social machinery to interact with it" (p. 15). This is particularly true if a robot's socially competent behaviour is mediated by its humanoid embodiment, permitting it to gesture or to generate facial expressions. "When a robot holds our gaze, the hardwiring of evolution makes us think that the robot is interested in us. When that happens, we feel a possibility for deeper connection" (Turkle, 2011, p. 110). Sociable robotics exploits the human mechanisms that offer this deeper connection so that humans won't require expert training in interacting with sociable robots.

A third purpose of sociable robotics is to explore cognitive scaffolding, which in this literature is often called leverage, in order to extend the capabilities of robots. For instance, many of the famous platforms of sociable robotics—including Cog (Brooks et al., 1999; Scassellati, 2002), Kismet (Breazeal, 2002, 2003, 2004), Domo (Edsinger-Gonzales & Weber, 2004), and Leanardo (Breazeal, Gray, & Berlin, 2009)—are humanoid in form and are social learners—their capabilities advance through imitation and through interacting with human partners. Furthermore, the success of the robot's contribution to the shared social environment leans heavily on the contributions of the human partner. "Edsinger thinks of it as getting Domo to do more 'by leveraging the people.' Domo needs the help. It understands very little about any task as a whole" (Turkle, 2011, p. 157).

The leverage exploited by a sociable robot takes advantage of behavioural loops mediated by the expressions and gestures of both robot and human partner. For example, consider the robot Kismet (Breazeal, 2002). Kismet is a sociable robotic "infant," a dynamic, mechanized head that participates in social interactions. Kismet has auditory and visual perceptual systems that are designed to perceive social cues provided by a human "caregiver." Kismet can also deliver such social cues

by changing its facial expression, directing its gaze to a location in a shared environment, changing its posture, and vocalizing.

When Kismet is communicating with a human, it uses the interaction to fulfill internal drives or needs (Breazeal, 2002). Kismet has three drives: a social drive to be in the presence of and stimulated by people, a stimulation drive to be stimulated by the environment in general (e.g., by colourful toys), and a fatigue drive that causes the robot to "sleep." Kismet sends social signals to satisfy these drives. It can manipulate its facial expression, vocalization, and posture to communicate six basic emotions: anger, disgust, fear, joy, sorrow, and surprise. These expressions work to meet the drives by manipulating the social environment in such a way that the environment changes to satisfy Kismet's needs.

For example, an unfulfilled social drive causes Kismet to express sadness, which initiates social responses from a caregiver. When Kismet perceives the caregiver's face, it wiggles its ears in greeting, and initiates a playful dialog to engage the caregiver. Kismet will eventually habituate to these interactions and then seek to fulfill a stimulation drive by coaxing the caregiver to present a colourful toy. However, if this presentation is too stimulating—if the toy is presented too closely or moved too quickly—the fatigue drive will produce changes in Kismet's behaviour that attempt to decrease this stimulation. If the world does not change in the desired way, Kismet will end the interaction by "sleeping." "But even at its worst, Kismet gives the appearance of trying to relate. At its best, Kismet appears to be in continuous, expressive conversation" (Turkle, 2011, p. 118).

Kismet's behaviour leads to lengthy, dynamic interactions that are realistically social. A young girl interacting with Kismet "becomes increasingly happy and relaxed. Watching girl and robot together, it is easy to see Kismet as increasingly happy and relaxed as well. Child and robot are a happy couple" (Turkle, 2011, p. 121). Similar results occur when adults converse with Kismet. "One moment, Rich plays at a conversation with Kismet, and the next, he is swept up in something that starts to feel real" (p. 154).

Even the designer of a humanoid robot can be "swept up" by their interactions with it. Domo (Edsinger-Gonzales & Weber, 2004) is a limbed humanoid robot that is intended to be a physical helper, by performing such actions as placing objects on shelves. It learns to behave by physically interacting with a human teacher. These physical interactions give even sophisticated users—including its designer, Edsinger—a strong sense that Domo is a social creature. Edsinger finds himself vacillating back and forth between viewing Domo as a creature or as being merely a device that he has designed.

> For Edsinger, this sequence—experiencing Domo as having desires and then talking himself out of the idea—becomes familiar. For even though he is Domo's programmer, the robot's behaviour has not become dull or predictable.

> Working together, Edsigner and Domo appear to be learning from each other. (Turkle, 2011, p. 156)

That sociable robots can generate such strong reactions within humans is potentially concerning. The feeling of the uncanny occurs when the familiar is presented in unfamiliar form (Freud, 1976). The uncanny results when standard categories used to classify the world disappear (Turkle, 2011). Turkle (2011) called one such instance, when a sociable robot is uncritically accepted as a creature, the robotic moment. Edsinger's reactions to Domo illustrated its occurrence: "And this is where we are in the robotic moment. One of the world's most sophisticated robot 'users' cannot resist the idea that pressure from a robot's hand implies caring" (p. 160).

At issue in the robotic moment is a radical recasting of the posthuman (Hayles, 1999). "The boundaries between people and things are shifting" (Turkle, 2011, p. 162). The designers of sociable robots scaffold their creations by taking advantage of the expert social abilities of humans. The robotic moment, though, implies a dramatic rethinking of what such human abilities entail. Might human social interactions be reduced to mere sense-act cycles of the sort employed in devices like Kismet? "To the objection that a robot can only seem to care or understand, it has become commonplace to get the reply that people, too, may only seem to care or understand" (p. 151).

In Hayles' (1999) definition of posthumanism, the body is dispensable, because the essence of humanity is information. But this is an extremely classical view. An alternative, embodied posthumanism is one in which the mind is dispensed with, because what is fundamental to humanity is the body and its engagement with reality. "From its very beginnings, artificial intelligence has worked in this space between a mechanical view of people and a psychological, even spiritual, view of machines" (Turkle, 2011, p. 109). The robotic moment leads Turkle to ask "What will love be? And what will it mean to achieve ever-greater intimacy with our machines? Are we ready to see ourselves in the mirror of the machine and to see love as our performances of love?" (p. 165).

5.12 The Architecture of Mind Reading

Social interactions involve coordinating the activities of two or more agents. Even something as basic as a conversation between two people is highly coordinated, with voices, gestures, and facial expressions used to orchestrate joint actions (Clark, 1996). Fundamental to coordinating such social interactions is our ability to predict the actions, interest, and emotions of others. Generically, the study of the ability to make such predictions is called the study of theory of mind, because many theorists argue that these predictions are rooted in our assumption that others, like us, have minds or mental states. As a result, researchers call our ability to foretell

others' actions mind reading or mentalizing (Goldman, 2006). "*Having* a mental state and *representing* another individual as having such a state are entirely different matters. The latter activity, *mentalizing* or *mind reading*, is a second-order activity: It is mind thinking about minds" (p. 3).

There are three general, competing theories about how humans perform mind reading (Goldman, 2006). The first is rationality theory, a version of which was introduced in Chapter 3 in the form of the intentional stance (Dennett, 1987). According to rationality theory, mind reading is accomplished via the ascription of contents to the putative mental states of others. In addition, we assume that other agents are rational. As a result, future behaviours are predicted by inferring what future behaviours follow rationally from the ascribed contents. For instance, if we ascribe to someone the belief that piano playing can only be improved by practising daily, and we also ascribe to them the desire to improve at piano, then according to rationality theory it would be natural to predict that they would practise piano daily.

A second account of mentalizing is called theory-theory (Goldman, 2006). Theory-theory emerged from studies of the development of theory of mind (Gopnik & Wellman, 1992; Wellman, 1990) as well as from research on cognitive development in general (Gopnik & Meltzoff, 1997; Gopnik, Meltzoff, & Kuhl, 1999). Theory-theory is the position that our understanding of the world, including our understanding of other people in it, is guided by naïve theories (Goldman, 2006). These theories are similar in form to the theories employed by scientists, because a naïve theory of the world will—eventually—be revised in light of conflicting evidence.

> Babies and scientists share the same basic cognitive machinery. They have similar programs, and they reprogram themselves in the same way. They formulate theories, make and test predictions, seek explanations, do experiments, and revise what they know in the light of new evidence. (Gopnik, Meltzoff, & Kuhl, 1999, p. 161)

There is no special role for a principle of rationality in theory-theory, which distinguishes it from rationality theory (Goldman, 2006). However, it is clear that both of these approaches to mentalizing are strikingly classical in nature. This is because both rely on representations. One senses the social environment, then thinks (by applying rationality or by using a naïve theory), and then finally predicts future actions of others. A third theory of mind reading, simulation theory, has emerged as a rival to theory-theory, and some of its versions posit an embodied account of mentalizing.

Simulation theory is the view that people mind read by replicating or emulating the states of others (Goldman, 2006). In simulation theory, "mindreading includes a crucial role for putting oneself in others' shoes. It may even be part of the brain's design to generate mental states that match, or resonate with, states of people one is observing" (p. 4).

The modern origins of simulation theory rest in two philosophical papers from the 1980s, one by Gordon (1986) and one by Heal (1986). Gordon (1986) noted that the starting point for explaining how we predict the behaviour of others should be investigating our ability to predict our own actions. We can do so with exceedingly high accuracy because "our declarations of immediate intention are causally tied to some actual precursor of behavior: perhaps tapping into the brain's updated behavioral 'plans' or into 'executive commands' that are about to guide the relevant motor sequences" (p. 159).

For Gordon (1986), our ability to accurately predict our own behaviour was a kind of practical reasoning. He proceeded to argue that such reasoning could also be used in attempts to predict others. We could predict others, or predict our own future behaviour in hypothetical situations, by *simulating* practical reasoning.

> To simulate the appropriate practical reasoning I can engage in a kind of *pretend-play:* pretend that the indicated conditions *actually obtain,* with all other conditions remaining (so far as is logically possible and physically probable) as they presently stand; then continuing the make-believe try to 'make up my mind' what to do given these (modified) conditions. (Gordon, 1986, p. 160)

A key element of such "pretend play" is that behavioural output is taken offline.

Gordon's proposal causes simulation theory to depart from the other two theories of mind reading by reducing its reliance on ascribed mental contents. For Gordon (1986, p. 162), when someone simulates practical reasoning to make predictions about someone else, "they are 'putting themselves in the other's shoes' in one sense of that expression: that is, they project themselves into the other's *situation,* but without any attempt to project themselves into, as we say, the other's 'mind.'" Heal (1986) proposed a similar approach, which she called replication.

A number of different variations of simulation theory have emerged (Davies & Stone, 1995a, 1995b), making a definitive statement of its fundamental characteristics problematic (Heal, 1996). Some versions of simulation theory remain very classical in nature. For instance, simulation could proceed by setting the values of a number of variables to define a situation of interest. These values could then be provided to a classical reasoning system, which would use these represented values to make plausible predictions.

> Suppose I am interested in predicting someone's action. . . . I place myself in what I take to be his initial state by imagining the world as it would appear from his point of view and I then deliberate, reason and reflect to see what decision emerges. (Heal, 1996, p. 137)

Some critics of simulation theory argue that it is just as Cartesian as other mind reading theories (Gallagher, 2005). For instance, Heal's (1986) notion of replication exploits shared mental abilities. For her, mind reading requires only the assumption

that others "are like me in being thinkers, that they possess the same fundamental cognitive capacities and propensities that I do" (p. 137).

However, other versions of simulation theory are far less Cartesian or classical in nature. Gordon (1986, pp. 17–18) illustrated such a theory with an example from Edgar Allen Poe's *The Purloined Letter*:

> When I wish to find out how wise, or how stupid, or how good, or how wicked is any one, or what are his thoughts at the moment, I fashion the expression of my face, as accurately as possible, in accordance with the expression of his, and then wait to see what thoughts or sentiments arise in my mind or heart, as if to match or correspond with the expression. (Gordon, 1986, pp. 17–18)

In Poe's example, mind reading occurs not by using our reasoning mechanisms to take another's place, but instead by exploiting the fact that we share similar bodies. Songwriter David Byrne (1980) takes a related position in *Seen and Not Seen*, in which he envisions the implications of people being able to mould their appearance according to some ideal: "they imagined that their personality would be forced to change to fit the new appearance. . . .This is why first impressions are often correct." Social cognitive neuroscience transforms such views from art into scientific theory.

> Ultimately, subjective experience is a biological data format, a highly specific mode of presenting about the world, and the Ego is merely a complex physical event—an activation pattern in your central nervous system. (Metzinger, 208, p. 208)

Philosopher Robert Gordon's version of simulation theory (Gordon, 1986, 1992, 1995, 1999, 2005a, 2005b, 2007, 2008) provides an example of a radically embodied theory of mind reading. Gordon (2008, p. 220) could "see no reason to hold on to the assumption that our psychological competence is chiefly dependent on the application of concepts of mental states." This is because his simulation theory exploited the body in exactly the same way that Brooks' (1999) behaviour-based robots exploited the world: as a replacement for representation (Gordon, 1999). "One's own behavior control system is employed as a manipulable model of other such systems. . . . Because one human behavior control system is being used to model others, general information about such systems is unnecessary" (p. 765).

What kind of evidence exists to support a more embodied or less Cartesian simulation theory? Researchers have argued that simulation theory is supported by the discovery of the brain mechanisms of interest to social cognitive neuroscience (Lieberman, 2007). In particular, it has been argued that mirror neurons provide the neural substrate that instantiates simulation theory (Gallese & Goldman, 1998): "[Mirror neuron] activity seems to be nature's way of getting the observer into the same 'mental shoes' as the target—exactly what the conjectured simulation heuristic aims to do" (p. 497–498).

Importantly, the combination of the mirror system and simulation theory implies that the "mental shoes" involved in mind reading are not symbolic representations. They are instead motor representations; they are actions-on-objects as instantiated by the mirror system. This has huge implications for theories of social interactions, minds, and selves:

> Few great social philosophers of the past would have thought that social understanding had anything to do with the pre-motor cortex, and that 'motor ideas' would play such a central role in the emergence of social understanding. Who could have expected that shared thought would depend upon shared 'motor representations'? (Metzinger, 2009, p. 171)

If motor representations are the basis of social interactions, then simulation theory becomes an account of mind reading that stands as a reaction against classical, representational theories. Mirror neuron explanations of simulation theory replace sense-think-act cycles with sense-act reflexes in much the same way as was the case in behaviour-based robotics. Such a revolutionary position is becoming commonplace for neuroscientists who study the mirror system (Metzinger, 2009).

Neuroscientist Vittorio Gallese, one of the discoverers of mirror neurons, provides an example of this radical position:

> Social cognition is not only social metacognition, that is, explicitly thinking about the contents of some else's mind by means of abstract representations. We can certainly explain the behavior of others by using our complex and sophisticated mentalizing ability. My point is that most of the time in our daily social interactions, we do not need to do this. We have a much more direct access to the experiential world of the other. This dimension of social cognition is embodied, in that it mediates between our multimodal experiential knowledge of our own lived body and the way we experience others. (Metzinger, 2009, p. 177)

Cartesian philosophy was based upon an extraordinary act of skepticism (Descartes, 1996). In his search for truth, Descartes believed that he could not rely on his knowledge of the world, or even of his own body, because such knowledge could be illusory.

> I shall think that the sky, the air, the earth, colors, shapes, sounds, and all external things are merely the delusions of dreams which he [a malicious demon] has devised to ensnare my judgment. I shall consider myself as not having hands or eyes, or flesh, or blood or senses, but as falsely believing that I have all these things. (Descartes, 1996, p. 23)

The disembodied Cartesian mind is founded on the myth of the external world.

Embodied theories of mind invert Cartesian skepticism. The body and the world are taken as fundamental; it is the mind or the holistic self that has become the myth. However, some have argued that our notion of a holistic internal self

is illusory (Clark, 2003; Dennett, 1991, 2005; Metzinger, 2009; Minsky, 1985, 2006; Varela, Thompson, & Rosch, 1991). "We are, in short, in the grip of a seductive but quite untenable illusion: the illusion that the mechanisms of mind and self can ultimately unfold only on some privileged stage marked out by the good old-fashioned skin-bag" (Clark, 2003, p. 27).

5.13 Levels of Embodied Cognitive Science

Classical cognitive scientists investigate cognitive phenomena at multiple levels (Dawson, 1998; Marr, 1982; Pylyshyn, 1984). Their materialism commits them to exploring issues concerning implementation and architecture. Their view that the mind is a symbol manipulator leads them to seek the algorithms responsible for solving cognitive information problems. Their commitment to logicism and rationality has them deriving formal, mathematical, or logical proofs concerning the capabilities of cognitive systems.

Embodied cognitive science can also be characterized as adopting these same multiple levels of investigation. Of course, this is not to say that there are not also interesting technical differences between the levels of investigation that guide embodied cognitive science and those that characterize classical cognitive science.

By definition, embodied cognitive science is committed to providing implementational accounts. Embodied cognitive science is an explicit reaction against Cartesian dualism and its modern descendant, methodological solipsism. In its emphasis on environments and embodied agents, embodied cognitive science is easily as materialist as the classical approach. Some of the more radical positions in embodied cognitive science, such as the myth of the self (Metzinger, 2009) or the abandonment of representation (Chemero, 2009), imply that implementational accounts may be even more critical for the embodied approach than is the case for classical researchers.

However, even though embodied cognitive science shares the implementational level of analysis with classical cognitive science, this does not mean that it interprets implementational evidence in the same way. For instance, consider single cell recordings from visual neurons. Classical cognitive science, with its emphasis on the creation of internal models of the world, views such data as providing evidence about what kinds of visual features are detected, to be later combined into more complex representations of objects (Livingstone & Hubel, 1988). In contrast, embodied cognitive scientists see visual neurons as being involved not in modelling, but instead in controlling action. As a result, single cell recordings are more likely to be interpreted in the context of ideas such as the affordances of ecological perception (Gibson, 1966, 1979; Noë, 2004). "Our brain does not simply register a chair, a teacup, an apple; it immediately represents the seen object as what I could do with

it—as an affordance, a set of possible behaviors" (Metzinger, 2009, p. 167). In short, while embodied and classical cognitive scientists seek implementational evidence, they are likely to interpret it very differently.

The materialism of embodied cognitive science leads naturally to proposals of functional architectures. An architecture is a set of primitives, a physically grounded toolbox of core processes, from which cognitive phenomena emerge. Explicit statements of primitive processes are easily found in embodied cognitive science. For example, it is common to see subsumption architectures explicitly laid out in accounts of behaviour-based robots (Breazeal, 2002; Brooks, 1999, 2002; Kube & Bonabeau, 2000; Scassellati, 2002).

Of course, the primitive components of a typical subsumption architecture are designed to mediate actions on the world, not to aid in the creation of models of it. As a result, the assumptions underlying embodied cognitive science's primitive sense-act cycles are quite different from those underlying classical cognitive science's primitive sense-think-act processing.

As well, embodied cognitive science's emphasis on the fundamental role of an agent's environment can lead to architectural specifications that can dramatically differ from those found in classical cognitive science. For instance, a core aspect of an architecture is control—the mechanisms that choose which primitive operation or operations to execute at any given time. Typical classical architectures will internalize control; for example, the central executive in models of working memory (Baddeley, 1986). In contrast, in embodied cognitive science an agent's environment is critical to control; for example, in architectures that exploit stigmergy (Downing & Jeanne, 1988; Holland & Melhuish, 1999; Karsai, 1999; Susi & Ziemke, 2001; Theraulaz & Bonabeau, 1999). This suggests that the notion of the extended mind is really one of an extended architecture; control of processing can reside outside of an agent.

When embodied cognitive scientists posit an architectural role for the environment, as is required in the notion of stigmergic control, this means that an agent's physical body must also be a critical component of an embodied architecture. One reason for this is that from the embodied perspective, an environment cannot be defined in the absence of an agent's body, as in proposing affordances (Gibson, 1979). A second reason for this is that if an embodied architecture defines sense-act primitives, then the available actions that are available are constrained by the nature of an agent's embodiment. A third reason for this is that some environments are explicitly defined, at least in part, by bodies. For instance, the social environment for a sociable robot such as Kismet (Breazeal, 2002) includes its moveable ears, eyebrows, lips, eyelids, and head, because it manipulates these bodily components to coordinate its social interactions with others.

Even though an agent's body can be part of an embodied architecture does not mean that this architecture is not functional. The key elements of Kismet's expressive features are shape and movement; the fact that Kismet is not flesh is irrelevant because its facial features are defined in terms of their function.

> In the robotic moment, what you are made of—silicon, metal, flesh—pales in comparison with how you behave. In any given circumstance, some people and some robots are competent and some not. Like people, any particular robot needs to be judged on its own merits. (Turkle, 2011, p. 94)

That an agent's body can be part of a functional architecture is an idea that is foreign to classical cognitive science. It also leads to an architectural complication that may be unique to embodied cognitive science. Humans have no trouble relating to, and accepting, sociable robots that are obviously toy creatures, such as Kismet or the robot dog Aibo (Turkle, 2011). In general, as the appearance and behaviour of such robots becomes more lifelike, their acceptance will increase.

However, as robots become closer in resemblance to humans, they produce a reaction called the uncanny valley (MacDorman & Ishiguro, 2006; Mori, 1970). The uncanny valley is seen in a graph that plots human acceptance of robots as a function of robot appearance. The uncanny valley is the part of the graph in which acceptance, which has been steadily growing as appearance grows more lifelike, suddenly plummets when a robot's appearance is "almost human"—that is, when it is realistically human, but can still be differentiated from biological humans.

The uncanny valley is illustrated in the work of roboticist Hiroshi Ishiguro, who,

> built androids that reproduced himself, his wife, and his five-year old daughter. The daughter's first reaction when she saw her android clone was to flee. She refused to go near it and would no longer visit her father's laboratory. (Turkle, 2011, p. 128)

Producing an adequate architectural component—a body that avoids the uncanny valley—is a distinctive challenge for embodied cognitive scientists who ply their trade using humanoid robots.

In embodied cognitive science, functional architectures lead to algorithmic explorations. We saw that when classical cognitive science conducts such explorations, it uses reverse engineering to attempt to infer the program that an information processor uses to solve an information processing problem. In classical cognitive science, algorithmic investigations almost always involve observing behaviour, often at a fine level of detail. Such behavioural observations are the source of relative complexity evidence, intermediate state evidence, and error evidence, which are used to place constraints on inferred algorithms.

Algorithmic investigations in classical cognitive science are almost exclusively focused on unseen, internal processes. Classical cognitive scientists use behavioural observations to uncover the algorithms hidden within the "black box" of an agent.

Embodied cognitive science does not share this exclusive focus, because it attributes some behavioural complexities to environmental influences. Apart from this important difference, though, algorithmic investigations—specifically in the form of behavioural observations—are central to the embodied approach. Descriptions of behaviour are the primary product of forward engineering; examples in behaviour-based robotics span the literature from time lapse photographs of Tortoise trajectories (Grey Walter, 1963) to modern reports of how, over time, robots sort or rearrange objects in an enclosure (Holland & Melhuish, 1999; Melhuish et al., 2006; Scholes et al., 2004; Wilson et al., 2004). At the heart of such behavioural accounts is acceptance of Simon's (1969) parable of the ant. The embodied approach cannot understand an architecture by examining its inert components. It must see what emerges when this architecture is embodied in, situated in, and interacting with an environment.

When embodied cognitive science moves beyond behaviour-based robotics, it relies on some sorts of behavioural observations that are not employed as frequently in classical cognitive science. For example, many embodied cognitive scientists exhort the phenomenological study of cognition (Gallagher, 2005; Gibbs, 2006; Thompson, 2007; Varela, Thompson, & Rosch, 1991). Phenomenology explores how people experience their world and examines how the world is meaningful to us via our experience (Brentano, 1995; Husserl, 1965; Merleau-Ponty, 1962).

Just as enactive theories of perception (Noë, 2004) can be viewed as being inspired by Gibson's (1979) ecological account of perception, phenomenological studies within embodied cognitive science (Varela, Thompson, & Rosch, 1991) are inspired by the philosophy of Maurice Merleau-Ponty (1962). Merleau-Ponty rejected the Cartesian separation between world and mind: "Truth does not 'inhabit' only 'the inner man,' or more accurately, there is no inner man, man is in the world, and only in the world does he know himself" (p. xii). Merleau-Ponty strove to replace this Cartesian view with one that relied upon embodiment. "We shall need to reawaken our experience of the world as it appears to us in so far as we are in the world through our body, and in so far as we perceive the world with our body" (p. 239).

Phenomenology with modern embodied cognitive science is a call to further pursue Merleau-Ponty's embodied approach.

> What we are suggesting is a change in the nature of reflection from an abstract, disembodied activity to an embodied (mindful), open-ended reflection. By embodied, we mean reflection in which body and mind have been brought together. (Varela, Thompson, & Rosch, 1991, p. 27)

However, seeking evidence from such reflection is not necessarily straightforward (Gallagher, 2005). For instance, while Gallagher acknowledges that the body is critical in its shaping of cognition, he also notes that many aspects of our bodily

interaction with the world are not available to consciousness and are therefore difficult to study phenomenologically.

Embodied cognitive science's interest in phenomenology is an example of a reaction against the formal, disembodied view of the mind that classical cognitive science has inherited from Descartes (Devlin, 1996). Does this imply, then, that embodied cognitive scientists do not engage in the formal analyses that characterize the computational level of analysis? No. Following the tradition established by cybernetics (Ashby, 1956; Wiener, 1948), which made extensive use of mathematics to describe feedback relations between physical systems and their environments, embodied cognitive scientists too are engaged in computational investigations. Again, though, these investigations deviate from those conducted within classical cognitive science. Classical cognitive science used formal methods to develop proofs about what information processing problem was being solved by a system (Marr, 1982), with the notion of "information processing problem" placed in the context of rule-governed symbol manipulation. Embodied cognitive science operates in a very different context, because it has a different notion of information processing. In this new context, cognition is not modelling or planning, but is instead coordinating action (Clark, 1997).

When cognition is placed in the context of coordinating action, one key element that must be captured by formal analyses is that actions unfold in time. It has been argued that computational analyses conducted by classical researchers fail to incorporate the temporal element (Port & van Gelder, 1995a): "Representations are static structures of discrete symbols. Cognitive operations are transformations from one static symbol structure to the next. These transformations are discrete, effectively instantaneous, and sequential" (p. 1). As such, classical analyses are deemed by some to be inadequate. When embodied cognitive scientists explore the computational level, they do so with a different formalism, called dynamical systems theory (Clark, 1997; Port & van Gelder, 1995b; Shapiro, 2011).

Dynamical systems theory is a mathematical formalism that describes how systems change over time. In this formalism, at any given time a system is described as being in a state. A state is a set of variables to which values are assigned. The variables define all of the components of the system, and the values assigned to these variables describe the characteristics of these components (e.g., their features) at a particular time. At any moment of time, the values of its components provide the position of the system in a state space. That is, any state of a system is a point in a multidimensional space, and the values of the system's variables provide the coordinates of that point.

The temporal dynamics of a system describe how its characteristics change over time. These changes are captured as a path or trajectory through state space. Dynamical systems theory provides a mathematical description of such trajectories,

usually in the form of differential equations. Its utility was illustrated in Randall Beer's (2003) analysis of an agent that learns to categorize objects, of circuits for associative learning (Phattanasri, Chiel, & Beer, 2007), and of a walking leg controlled by a neural mechanism (Beer, 2010).

While dynamical systems theory provides a medium in which embodied cognitive scientists can conduct computational analyses, it is also intimidating and difficult. "A common criticism of dynamical approaches to cognition is that they are practically intractable except in the simplest cases" (Shapiro, 2011, pp. 127–128). This was exactly the situation that led Ashby (1956, 1960) to study feedback between multiple devices synthetically, by constructing the Homeostat. This does not mean, however, that computational analyses are impossible or fruitless. On the contrary, it is possible that such analyses can co-operate with the synthetic exploration of models in an attempt to advance both formal and behavioural investigations (Dawson, 2004; Dawson, Dupuis, & Wilson, 2010).

In the preceding paragraphs we presented an argument that embodied cognitive scientists study cognition at the same multiple levels of investigation that characterize classical cognitive science. Also acknowledged is that embodied cognitive scientists are likely to view each of these levels slightly differently than their classical counterparts. Ultimately, that embodied cognitive science explores cognition at these different levels of analysis also implies that embodied cognitive scientists are also committed to the notion of validating their theories by seeking strong equivalence. It stands to reason that the validity of a theory created within embodied cognitive science would be best established by showing that this theory is supported at all of the different levels of investigation.

5.14 What Is Embodied Cognitive Science?

To review, the central claim of classical cognitive science is that cognition is computation, where computation is taken to be the manipulation of internal representations. From this perspective, classical cognitive science construes cognition as an iterative sense-think-act cycle. The "think" part of this cycle is emphasized, because it is responsible for modelling and planning. The "thinking" also stands as a required mentalistic buffer between sensing and acting, producing what is known as the classical sandwich (Hurley, 2001). The classical sandwich represents a modern form of Cartesian dualism, in the sense that the mental (thinking) is distinct from the physical (the world that is sensed, and the body that can act upon it) (Devlin, 1996).

Embodied cognitive science, like connectionist cognitive science, arises from the view that the core logicist assumptions of classical cognitive science are not adequate to explain human cognition (Dreyfus, 1992; Port & van Gelder, 1995b; Winograd & Flores, 1987b).

> The lofty goals of artificial intelligence, cognitive science, and mathematical linguistics that were prevalent in the 1950s and 1960s (and even as late as the 1970s) have now given way to the realization that the 'soft' world of people and societies is almost certainly not amenable to a precise, predictive, mathematical analysis to anything like the same degree as is the 'hard' world of the physical universe. (Devlin, 1996, p. 344)

As such a reaction, the key elements of embodied cognitive science can be portrayed as an inversion of elements of the classical approach.

While classical cognitive science abandons Cartesian dualism in one sense, by seeking materialist explanations of cognition, it remains true to it in another sense, through its methodological solipsism (Fodor, 1980). Methodological solipsism attempts to characterize and differentiate mental states without appealing to properties of the body or of the world (Wilson, 2004), consistent with the Cartesian notion of the disembodied mind.

In contrast, embodied cognitive science explicitly rejects methodological solipsism and the disembodied mind. Instead, embodied cognitive science takes to heart the message of Simon's (1969) parable of the ant by recognizing that crucial contributors to behavioural complexity include an organism's environment and bodily form. Rather than creating formal theories of disembodied minds, embodied cognitive scientists build embodied and situated agents.

Classical cognitive science adopts the classical sandwich (Hurley, 2001), construing cognition as an iterative sense-think-act cycle. There are no direct links between sensing and acting from this perspective (Brooks, 1991); a planning process involving the manipulation of internal models stands as a necessary intermediary between perceiving and acting.

In contrast, embodied cognitive science strives to replace sense-think-act processing with sense-act cycles that bypass representational processing. Cognition is seen as the control of direct action upon the world rather than the reasoning about possible action. While classical cognitive science draws heavily from the symbol-manipulating examples provided by computer science, embodied cognitive science steps further back in time, taking its inspiration from the accounts of feedback and adaptation provided by cybernetics (Ashby, 1956, 1960; Wiener, 1948).

Shapiro (2011) invoked the theme of conceptualization to characterize embodied cognitive science because it saw cognition as being directed action on the world. Conceptualization is the view that the form of an agent's body determines the concepts that it requires to interact with the world. Conceptualization is also a view that draws from embodied and ecological accounts of perception (Gibson, 1966, 1979; Merleau-Ponty, 1962; Neisser, 1976); such theories construed perception as being the result of action and as directing possible actions (affordances) on the world.

As such, the perceptual world cannot exist independently of a perceiving agent; *umwelten* (Uexküll, 2001) are defined in terms of the agent as well.

The relevance of the world to embodied cognitive science leads to another of its characteristics: Shapiro's (2011) notion of replacement. Replacement is the view that an agent's direct actions on the world can replace internal models, because the world can serve as its own best representation. The replacement theme is central to behaviour-based robotics (Breazeal, 2002; Brooks, 1991, 1999, 2002; Edsinger-Gonzales & Weber, 2004; Grey Walter, 1963; Sharkey, 1997), and leads some radical embodied cognitive scientists to argue that the notion of internal representations should be completely abandoned (Chemero, 2009). Replacement also permits theories to include the co-operative interaction between and mutual support of world and agent by exploring notions of cognitive scaffolding and leverage (Clark, 1997; Hutchins, 1995; Scribner & Tobach, 1997).

The themes of conceptualization and replacement emerge from a view of cognition that is radically embodied, in the sense that it cannot construe cognition without considering the rich relationships between mind, body, and world. This also leads to embodied cognitive science being characterized by Shapiro's (2011) third theme, constitution. This theme, as it appears in embodied cognitive science, is the extended mind hypothesis (Clark, 1997, 1999, 2003, 2008; Clark & Chalmers, 1998; Menary, 2008, 2010; Noë, 2009; Rupert, 2009; Wilson, 2004, 2005). According to the extended mind hypothesis, the world and body are literally constituents of cognitive processing; they are not merely causal contributors to it, as is the case in the classical sandwich.

Clearly embodied cognitive science has a much different view of cognition than is the case for classical cognitive science. This in turn leads to differences in the way that cognition is studied.

Classical cognitive science studies cognition at multiple levels: computational, algorithmic, architectural, and implementational. It typically does so by using a top-down strategy, beginning with the computational and moving "down" towards the architectural and implementational (Marr, 1982). This top-down strategy is intrinsic to the methodology of reverse engineering or functional analysis (Cummins, 1975, 1983). In reverse engineering, the behaviour of an intact system is observed and manipulated in an attempt to decompose it into an organized system of primitive components.

We have seen that embodied cognitive science exploits the same multiple levels of investigation that characterize classical cognitive science. However, embodied cognitive science tends to replace reverse engineering with an inverse, bottom-up methodology, as in forward engineering or synthetic psychology (Braitenberg, 1984; Dawson, 2004; Dawson, Dupuis, & Wilson, 2010; Pfeifer & Scheier, 1999). In forward engineering, a set of interesting primitives is assembled into a working system.

This system is then placed in an interesting environment in order to see what it can and cannot do. In other words, forward engineering starts with implementational and architectural investigations. Forward engineering is motivated by the realization that an agent's environment is a crucial contributor to behavioural complexity, and it is an attempt to leverage this possibility. As a result, some have argued that this approach can lead to simpler theories than is the case when reverse engineering is adopted (Braitenberg, 1984).

Shapiro (2011) has noted that it is too early to characterize embodied cognitive science as a unified school of thought. The many different variations of the embodied approach, and the important differences between them, are beyond the scope of the current chapter. A more accurate account of the current state of embodied cognitive science requires exploring an extensive and growing literature, current and historical (Agre, 1997; Arkin, 1998; Bateson, 1972; Breazeal, 2002; Chemero, 2009; Clancey, 1997; Clark, 1997, 2003, 2008; Dawson, Dupuis, & Wilson, 2010; Dourish, 2001; Gallagher, 2005; Gibbs, 2006; Gibson, 1979; Goldman, 2006; Hutchins, 1995; Johnson, 2007; Menary, 2010; Merleau-Ponty, 1962; Neisser, 1976; Noë, 2004, 2009; Pfeifer & Scheier, 1999; Port & van Gelder, 1995b; Robbins & Aydede, 2009; Rupert, 2009; Shapiro, 2011; Varela, Thompson, & Rosch, 1991; Wilson, 2004; Winograd & Flores, 1987b).