Explaining the Mind: The Embodied Cognition Challenge

by

Anatoly Zhitnik

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Philosophy

Waterloo, Ontario, Canada, 2008

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

This thesis looks at a relatively new line of research in Cognitive Science – embodied cognition. Its relation to the computational-representational paradigm, primarily symbolicism, is extensively discussed. It is argued that embodied cognition is compatible with the established paradigm but challenges its research focus and traditionally assumed segregation of cognition from bodily and worldly activities

Subsequently the impact of embodied cognition on philosophy of Cognitive Science is considered. The second chapter defends the applicability of *mechanistic explanation* to cases of embodied cognition. Further, it argues that a proposed alternative, dynamic systems theory, is not a substitute to the mechanistic approach.

The last chapter critically examines the thesis that mind is extended beyond the bodily boundary and into the world. It is concluded that arguments in favour of the extended mind thesis are inadequate. Considerations in favour of the orthodox view that the does not "leak" out into the world are also presented.

Acknowledgements

I am greatly indebted to a number of people whose help was instrumental in helping me start and finish this project. First and foremost, I would like to thank my supervisor, Chris Eliasmith, who patiently guided me throughout the writing process and exposed me to the key ideas that helped shape this thesis; his intellectual support was simply invaluable.

I would also like to thank all of the graduate students at the UW Department of Philosophy for their friendship and stimulating conversations. I also wish to extend my thanks to Shawn Warren for, literally, hours of discussions we have had about the mind.

I am especially grateful to my parents for making this possible at all; my sister, Kate, and Mike for letting me stay with them while I was writing this; and, of course, my nephews, Tim and Kai, who have been an endless source of joy, amusement, and affection.

Last, but not least, I want to thank Michelle, without her unwavering support and delicious sandwiches I, undoubtedly, never would have finished this project.

Table of Contents

Introduction

Since the inception of Cognitive Science around the middle of the 20th century, it has been dominated by the computational-representational approach (CR) to understanding the mind. This approach views cognition as a process of transforming representations that carry information relevant to the cognizer such that it can produce flexible and sophisticated behaviour. CR can be considered to be the reigning paradigm in Cognitive Science but there is an on-going debate as to how its core notions, namely computation and representation, should be conceptualized. Recently, a new approach to characterizing cognitive functions has been gaining popularity – embodied cognition (EC). Despite EC's wide appeal, it is notoriously difficult to isolate a core set of principles that the majority of EC researchers would endorse (Ziemke, 2003). Furthermore, even given a common set of principles, there is a further question of the extent to which they are compatible with various conceptualizations of CR. In other words, even if two proponents of EC agree on some key principle, they may still disagree about the implications it has for the CR approach.

In this thesis I focus on EC and its implications for CR and Philosophy of Cognitive Science. In the first chapter I provide an overview of CR, concentrating on the theoretical foundations that drove early artificial intelligence (AI) research. In addition, I am concerned with EC and how the term '*embodiment'* figures in this literature. I attempt to distill the foundational ideas of EC and use that analysis to highlight shortcomings of CR. I argue that EC does not threaten CR as a dominant paradigm but only suggests an alternative research domain and a rethinking of the role and complexity of computations and representations in cognitive

activity. The second chapter discusses the implications of EC for the way researchers analyze and explain the workings of cognitive systems. I defend the value and applicability of *mechanistic explanation* to EC and argue that a proposed alternative, Dynamic Systems Theory (DST), is not a substitute. The last chapter focuses on the implications of EC ideas for Philosophy of Mind. Specifically, I will critically evaluate the thesis that mind and cognition extend beyond the bodily boundaries and into the external environment. I argue that there are no compelling reasons to accept this position and, to the contrary, everything we know about cognition and mechanisms that make intelligent behaviour possible favours the traditional view that mental processes do not extend beyond the confines of the body.

Chapter 1

Minds and Bodies

1.1 Early AI

Metaphors provide a way of comprehending the ill-understood in terms of the familiar. The metaphor for mind as a kind of machine or computer is at the core of the CR paradigm. In order to fully understand the grounds for such a characterization, one ought to look at the theoretical foundations of CR exemplified by early AI research.

The roots of CR can be traced to early work in AI and, by extension, computer science. Pioneer AI researchers like Allen Newell and Herbert Simon took the mind to be a control system for the body, which was primarily responsible for generating appropriate output (behaviour) in response to environmental input. "The mind then is simply the name we give to the control system that has evolved within the organism to carry out the interactions to the benefit of that organism or, ultimately, for the survival of its species" (Newell, 1990., p.43). With such a characterization at hand it became fruitful to see the mind as generating *response functions*; functions of the environmental input and the goals of the organism which result in behaviour.

This immediately suggests that the control system has to, in some sense, have *knowledge* that is relevant to achieving its goals. For instance, if the goal is to find food, the organism has to *know* what is eatable and what is not. Such factual knowledge is complemented by normative knowledge regarding inferences (strictly logical or mere heuristics that proved to be useful) that can made from the known facts. Furthermore, a system

can be seen as more or less intelligent depending on how well it can utilize its knowledge to attain its goals (Newell, 1990). The idea that intelligent behaviour necessarily implicates relevant knowledge has been dubbed the Knowledge Principle (Lenat & Feigenbaum, 1991., p. 186):

> The Knowledge Principle (KP): A system exhibits intelligent understanding and action at a high level of competence primarily because of the *specific* knowledge that it can bring to bear: the concepts, facts, representations, methods, models, metaphors, and heuristics about its domain of endeavor.

The more controversial idea articulated by Lenat and Feigenbaum in the same paper is that knowledge, not reasoning methods, constitutes *the* obstacle on the way to creating artificial intelligence. In saying that they do not mean to downplay the importance of valid reasoning methods, but only to imply that it is a relatively easy task to formalize valid inference rules (logic) and heuristics. In contrast, common sense knowledge[1] and specific facts about a particular domain are much harder to obtain and specify explicitly.[2]

The description of the system at the *knowledge level* is deliberate and guided by pragmatic considerations. There is nothing, in principle, that precludes a different level of description, say, in terms of the physical laws that govern the system. There are, however, at least three virtues of pitching the description of a cognitive system at the "knowledge level": predictive success; the ability to study the system in abstraction from its actual internal workings; and the fact that clear cases of cognition can be usefully characterized as

1 Common sense facts are a multitude and we generally never entertain them explicitly, which makes it harder to figure out what knowledge is actually necessary for normal intelligent activity. For example: doors normally have knobs, cars cannot swim etc.

2 Lenat and Feigenbaum took this theoretical outlook to the extreme with the CYC project – one of the most elaborate AI systems meant to simulate human reasoning. The CYC system consists of a vast knowledge base and an inference engine that includes valid rules of inferences, rules of thumb and various general as well as domain-specific heuristics. Various items of knowledge often had to be coded by hand, and in ten years since its inception in 1984 the project consumed "about two person-centuries of data-entry time" (Clark, 1997, p.2).

knowledge-dependent problem solving activity.

The first virtue is clearly exemplified in the pervasive use of folk psychology. We take an *intentional stance* (Dennett, 1987) towards other agents and are able to make sense of and predict their behaviour (output of their cognitive system) by taking into account only their goals and doxastic status. In fact, taking an intentional stance towards an agent is almost identical to describing them at the *knowledge level.*[3] In Newell's terminology (Newell, 1982., p.98):

> To treat a system at the knowledge level is to treat it as having some knowledge and some goals, and believing it will do whatever is within its power to attain its goals, in so far as its knowledge indicates.

If I know that person *X* wants to catch a bus and that *X* knows the bus leaves at some time *t* I can form a reasonable prediction that *X* will be at the bus stop around time *t*.

The second theoretical payoff is that there is nothing about the concepts of knowledge and goals that prescribes a specific form of implementation: "It is easy to see why describing a system at the knowledge level is useful. The essential feature is that no details of the actual internal processing are required" (Newell, 1990, p.50). Some philosophers saw the abstraction from implementational details as a key to understanding the nature of mental states. Putnam (1975), for example, argued that a theory of mind that identifies mental states with specific brain-states is guilty of chauvinism; it does not allow for creatures with brains different from ours to have mental states. Instead, Putnam argued, one ought to identify a mental state with the *functional* role it serves within the whole cognitive system. Thus, as long as a given system,

3 The most glaring difference, however, is that Newell appeals to an agent's *knowledge* to account for behaviour, whereas Dennett only appeals to *beliefs*. The difference is crucial because knowledge is traditionally identified with justified and, most importantly, *true* beliefs. But since we normally act in accordance with our beliefs (which may or may not be true and, therefore, may or may not constitute knowledge) Dennett's notion is broader than Newell's.

regardless of how it is implemented, has the right functional organization it can have mental states.

Lastly, a view of cognition as a knowledge-based process coheres well with the fact that the most salient examples of cognitive tasks we perform are usefully characterized as species of problem-solving. Even the most rudimentary activity such as unlocking our house door can be construed as a problem that requires specific knowledge (e.g., where the key is), and a plethora of common-sense knowledge about how doors work and what keys normally look like. Examples of high-level cognition such as the ability to play chess, or diagnose an illness, are clearly most dependent on domain-specific knowledge.

Thus, early AI researchers took the mind to be a control system responsible for generating behaviours required to meet the goals of an organism. This was essentially seen as a problem-solving activity that required factual as well as normative knowledge. A description of the mind as a knowledge-dependent problem-solver allowed early cognitive scientists to pursue their work without much concern for details of implementation. Most of the work dealt with the issues of knowledge representation and transformation.

1.2 Representations and Transformations

In order for the control system to perform its functions it has to be able to (a) retrieve the knowledge from the external environment (b) internally represent and store this knowledge for future use and (c) to freely transform and bring it to bear on the task at hand. The role of internal representations is to "stand in" for the states of the world relevant to some particular goal. For instance, consider an organism whose goal is to escape a rapidly approaching

predator. Or, alternatively, think of the organism as facing a*problem* of how to get away, which requires factual information about the environment and the predator in order to be solved. From the perspective of classical AI it is first necessary for an organism to encode sensory stimuli so that the initial location of the predator, trajectory of its path, speed, the nature of the terrain, etc. are internally represented and available for use. Once these representations have been built up, relevant information becomes available for the cognitive system. By transforming the acquired data the system can infer that the predator will be at some location at time *t* – it will have an internal representation of that knowledge. This cycle of encoding stimuli, building up and transformation of internal representations will continue until the system has the requisite information to plan its actions in order to avoid the predator.

As the complexity of the organism grows, the range of goals and possible environmental interactions expands as well. As a result, the mind ought to generate new response functions necessary to cope with the changes and solve new problems – it ought to exhibit adaptive behaviour. Proponents of CR contend that it becomes necessary to represent more features of the environment in greater detail, form new goals and augment the set of possible transformations (Newell, 1990). While it may have initially sufficed to just represent speed, trajectory, and location it may become necessary to be attuned to other environmental features such as colour, smell, or particular sound frequencies in order to better detect a predator. Consequently, a new representational repertoire is only useful in so far as it can be used to attain goals, which may require new sorts of transformations.[4] It may, for example, be necessary to learn new heuristics[5] or inference rules before the available knowledge can be

4 However, it does not imply that cognitive abilities are exercised only to generate outward goal-directed behaviour.

5 For example: "Red things which smell *this* way are dangerous" or "Red things which smell *that* way are eatable"

used to achieve some goal and promote adaptive success.

The fundamental role knowledge allegedly plays in intelligent behaviour places a constraint on how it has to be encoded by a cognitive system. The manner in which knowledge is encoded has to be sufficiently flexible to sustain a rich variety of representations and transformations, be capable of using them to generate novel response functions, and ultimately approximate intelligent behaviour such as problem-solving. According to CR, the challenge for Cognitive Science is to provide a detailed description and construct a physical system that can meet these criteria. Based on empirical results of early AI research, Newell and Simon have suggested a working hypothesis called the Physical Symbol System (PSS) hypothesis: "The necessary and sufficient conditions for a physical system to exhibit general intelligent action is that it be a physical symbol system" (Newell, 1980., p.170). The approach guided by PSS is often referred to as the 'classical view', 'cognitivism', or 'symbolicism'.

The concept of a PSS is closely linked the notion of a formal (symbol) system. A formal system consists of a set of symbols, grammar that dictates how these symbols are to be conjoined, a set of axioms and, finally, inference rules (Haugeland, 1981). Symbols themselves have no intrinsic meaning, but have to be interpreted or mapped onto the real world. Notably, a formal system can be defined and used based solely on syntax, abstracted entirely from the meaning (semantics) of symbols. Paradigmatic examples of formal systems include logic, Euclidian geometry, algebra etc. Any input-output relation in a formal system could be specified in purely mechanistic terms – as an appropriate sequence of symbol manipulations. A proof within, say, first-order logic can be specified as an algorithm – a finite list of inference rules that are to be applied to some original symbolic expression (input) in order to arrive at the

target symbolic expression (output). Moreover, a machine could be constructed that instantiates a specified algorithm by automatically manipulating symbol tokens. Pioneering work in AI has in fact focused on building machines that were designed for a very specific purpose such as playing checkers, proving logic theorems and solving cryptoarithmetic problems. What is notable about all these problems is that they are extremely well-defined. They have *the* correct solution that can in principle be reached by brute search through all the possible combinations. To prove a logic theorem, one simply needs to know the appropriate succession of inference rule that have to be applied to the initial expression. So, given enough computational power and time, one can simply try all the possibilities in a "mindless" manner until the appropriate sequence is found.

The work of Alonzo Church and Alan Turing has provided an invaluable result for Cognitive Science showing that there exists a theoretical construct – a Universal Turing Machine (UTM) - that can produce any function which can be specified by an algorithm. Such functions constitute a class of *computable functions* (Pylyshyn, 1989., p.54):

> Universality implies that a formal symbol processing mechanism can produce *any* arbitrary input-output function that we can specify in sufficient detail. Put in more familiar terms, a universal machine can be *programmed* to compute any formally specified function.

A UTM is an idealized device that is postulated to have infinite memory and infinite amount of time in order to encode, store and manipulate discrete symbols. Digital computers we have today are for all intents and purposes Turing Machines with limited storage. The following lengthy quotation from Pylyshyn nicely captures the motivations for choosing a digital computer as a physical system for modelling human cognition (Pylyshyn, 1989., p.52):

> At the most abstract level, the class of mechanisms called *computers* are the only known mechanisms that are sufficiently plastic in their behaviour to match the plasticity of human cognition. They are also the only known mechanism capable of producing behaviour that can be described as *knowledge dependent*. Because of such properties *computing* remains the primary candidate for meeting the dual needs of explaining cognition in mechanism terms and accounting for certain otherwise problematic aspects of cognition - in particular the fact that behaviour can be systematically influenced by - inducing differences in beliefs or goals.

Thus, the mind is viewed as a program instantiated in the brain. Cognition, then, is a process of transforming symbolic representations by means of computations in order to produce appropriate response functions. From this perspective, the primary challenges confronting Cognitive Science concern knowledge representation and computational methods.

The computational representational (CR) paradigm has been hailed as "the most theoretically and experimentally successful approach to mind ever developed" (Thagard, 2005., p.11). The classical symbolicist version advocated by Newell and Simon, however, has been extensively questioned. For example, the computer metaphor, reliance on serial computation, and the construal of representations as inherently symbolic have each been challenged by the connectionist research program (Rumelhart, 1989; Smolensky, 1987). Connectionism exploits *parallel* and *distributed* ("brain-style") computation and uses *sub-symbolic* representations. In contrast to serial computation, where computations are performed sequentially, parallel and distributed approach relies on performing multiple computations simultaneously. Connectionist representations have been called sub-symbolic because their constituent parts (model neuron units) do not themselves represent anything. However, an interaction of multiple such units gives rise to a representation. Nevertheless, the central CR hypothesis, that mind is best

understood as performing computations over representations, broadly construed, is accepted by connectionist researchers.

An explicit denial of this hypothesis comes from the proponents of dynamicist view of cognition (van Gelder & Port, 1995a). The Dynamicist Hypothesis (DH) "is encapsulated in the simple slogan, *cognitive agents are dynamical systems*" (van Gelder, 1998., p.615), where 'dynamical' simply means 'changing in time'. At the heart of DH is the rejection of the view that cognition is necessarily computational in nature, requires internal representations, and is instantiated by the brain alone. Those who champion DH contend that (van Gelder & Port, 1995b., p.3):

> The cognitive system is not a computer, it is a dynamical system. It is not the brain, inner and encapsulated; rather, it is the whole system comprised of nervous system, body, and environment. The cognitive system... is a structure of mutually and simultaneously influencing *change.*

Proponents of DH contend that mind has to be understood using the vocabulary of Dynamic Systems Theory (DST) and eschew the notions of computation and representation.

Dynamicism has thus far failed to replace CR as the dominant paradigm in cognitive science, however some of its central ideas, especially "situatedness" and "embodiment", figure prominently in embodied cognition (EC) – the primary focus of this thesis. I will now articulate the basics of EC and evaluate its impact on the established CR paradigm. I argue that EC, contrary to claims of DST theorists, does not invalidate the central hypothesis of CR. Rather, I argue that in light of evolutionary considerations, EC suggests an alternative research domain and invites cognitive scientists to rethink the extent to which a cognitive system has to rely on complex representations to produce complex behaviour. In the subsequent chapters I

consider how EC challenges mechanistic explanation and traditionally accepted boundaries of mind.

1.3 Embodied Cognition

The EC research program takes a markedly different approach to understanding the mind than the CR paradigm. The latter conceptualizes the mind as a rule-based information processor instantiated in the brain. Thus, the brain is the seat of intelligence, where cognition is centralized. Anything outside of it is treated only as a source of input. The fact that intelligent creatures have physical bodies that interact with their environmental niche was not thought to make a difference for cognition *per se*. The implicit assumption was that one could abstract away from these bodily elements without adversely affecting theories and explanations. The EC movement calls into question this "isolationist assumption" (Cowart, 2004) regarding the mind, and suggests that a proper account of how we perform cognitive tasks has to make reference to the environment, body and the history, individual and evolutionary, of their dynamic, real-time interaction.

The field of EC is relatively new, and still lacks an agreed upon conceptual foundation. For instance, even though proponents of EC would agree that embodiment is necessary for cognition, the very notion of 'embodiment' remains nebulous. There is no agreed upon definition of 'embodiment' or what aspects thereof are of interest to cognitive science. Notable reviews of EC research such as Anderson (2003) and Wilson (2002) forego providing one altogether. Recent work by Tom Ziemke has attempted to confront the issue directly and focuses on the different notions of embodiment that figure in the EC literature.

In what follows I explicate the basics of EC by focusing on two questions: (1) What does it mean for a cognitive system to be 'embodied'? and (2) What implications do facts about embodiment have for the CR paradigm and cognitive science in general? In order to address (1) I take as my starting point the six notions of embodiment found in Ziemke (2003) and illustrate how they figure in EC research. This discussion naturally leads to a contrast between EC and CR, and illustrates how the former suggests a rethinking of the nature and role of representations in cognitive systems, and the choice of problem domains in AI research.

1.4 Aspects of Embodiment

The PSS hypothesis advocated by symbolicists effectively demarcated the kinds of systems that can be cognitive from the ones that cannot. Proponents of EC seek to do the same by claiming that cognition is necessarily embodied. Getting clear on what embodiment is would effectively supply EC with necessary conditions for cognition. Ziemke lists the following six notions of 'embodiment' that can be found throughout EC literature: (1) structural coupling embodiment; (2) historical embodiment; (3) physical embodiment; (4) organism embodiment; (5) organismic embodiment; and (6) social embodiment. Each one should be seen as highlighting *an aspect of embodiment* rather than being an exhaustive definition of the concept. If a single, robust notion of embodiment is viable it is likely to incorporate insights from all of the variations listed above.

The best way to explain the meaning and significance of these notions is by way of an example. Rodney Brooks' robotics research is perhaps the single example most often associated with EC research. I will discuss Brooks' research in some detail and refer to it in the

subsequent discussion of the aforementioned notions of embodiment.

1.4.1 Autonomous Mobile Robots

Rodney Brooks' work on mobile robots (mobots) has become a cornerstone of EC and constitutes a major departure from the classical approach to AI. Brooks' most famous creation is *Herbert* – a mobot that wanders around the lab, and detects and collects empty soda cans. Perhaps the most notable feature that set Brooks' work apart from earlier attempts to build autonomous agents was the eschewal of detailed world representations. Previous attempts to create mobots had relied on the "*sense-model-plan-act*" (Brooks, 1991b) (SMPA) framework grounded in classical CR view of the mind. In SMPA the relationship between the mind and the external world is mediated by detailed internal representations. Intelligent behaviour relies heavily on knowledge that has to be explicitly represented in order to be useful. SMPA-type mobots are equipped with sensors (i.e, TV camera) that take input from the environment and construct internal world models, which are used by the central processor for subsequent planning and acting. These mobots were typically tested in simplified static environments since it allowed AI researchers to avoid practical issues, such as the need for a vision system that can recognize and track objects in a dynamic environment. When placed in real-world environments early mobots proved to be brittle: uneven lighting, obscured shapes, sensory noise all made it exceedingly difficult to extract required information from a camera image. Brooks noted that even in highly simplified circumstances "[m]uch of the processing time was consumed in the perceptual end of the systems and building up the world models" (Brooks, 1991b., p.570). This often resulted in outdated representations of the world and, therefore, sub-

optimal plans and actions. The SMPA framework led to a “representational bottleneck” which hindered the mobots' ability to act promptly and adequately. Such a state of affairs, Brooks argued, is evolutionarily implausible because cognition is time-pressured. Quick responses to environmental changes are necessary to ensure survival (Brooks, 1991a., p.140):

> When we examine very simple level intelligence we find that explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own model.

Using “world as its own best model” has become a slogan of Brooks’ new approach to robotics. The idea is quite simple. Why bother representing all the details and features of the world when they are always available? Furthermore, the world is always up to date, there is no worry about planning and acting based on outdated descriptions.

Brooks’ mobots eschew the need for explicit representations by using an alternative design methodology called the *subsumption architecture* (SA). SA mobot consists of distinct layers, each devoted to producing a simple behaviour such as detecting an obstacle ahead or moving in a random direction. The layers are directly connected to sensors that take input from the environment and actuators which are responsible for moving the mobot about its environment. Thus, claims Brooks, there is a direct route from perception to action without representations to mediate sensory input and action. Furthermore, layer activity is not determined by a central planning system; instead each layer can override the activity of others based on environmental cues. For instance, consider a mobot with two activity producing layers: obstacle avoidance and random wandering. The first layer is directly connected to a sonar sensor which can detect the presence of an obstacle ahead. Once an obstacle is detected the mobot halts and reorients away from the obstruction. The second layer is responsible for

initiating movement in a random direction. In the absence of obstacles, the first layer is inactive which allows the mobot to engage in random wandering. As it moves through the environment, a sonar sensor detects an obstacle ahead thereby activating the obstacle avoidance layer and inhibiting wandering. This approach, argued Brooks, allows for complex behaviour to *emerge* without a predetermined plan, central control system or symbolic representations. Instead the behaviour arises out of interactions between behaviour producing layers and environmental structures.

1.4.2 Embodiment as Structural Coupling

The first of the six notions I will consider is the notion of embodiment as *structural coupling* between the agent and the environment. Descriptions of cognition as structural coupling can be found throughought EC literature (Clark 1997; Brooks 1991a; Brooks, 1991b) as well as in the literature related to DST (van Gelder & Port 1995a; Thelen, 1995). Quick et. al (1999., p, 3) provide the following definition of this kind of embodiment:

> A system *X* is embodied in an environment *E* if pertubatory channels exist between the two. That is, *X* is embodied in *E* if for every time t at which both *X* and *E* exist, some subset of *E*'s possible states with respect to *X* have the capacity to perturb *X*'s state, and some subset of *X*'s possible states with respect to *E* have the capacity to perturb *E*'s state.

The emphasis is placed on the continuous, mutually-altering interaction between the environment and

the agent.

Embodiment as structural coupling underlies the popular idea in EC that cognition is a *situated* or *embedded* activity. Wilson (2002., p.626) explicates this idea in the following way:

> Simply put, situated cognition is cognition that takes place in the context of task-relevant inputs and outputs. That is, while a cognitive process is being carried out, perceptual information continues to come in that affects processing, and motor activity is executed that affects the environment in task-relevant ways.

Thus, upon carrying out some action in an attempt to solve the problem, the problem itself changes thereby affecting the strategy used to solve it. For instance, consider Brooks' Herbert navigating the busy MIT lab full of dynamic obstacles. Since Herbert's goal is to avoid collision, the choice of a path depends on the locations and movements of all the surrounding objects. Similarly other individuals in the lab choose a path that takes into account location and the trajectory of Herbert's movement. Hence, the path Herbert takes is a function of the paths others around him take and vice versa. There is a dynamic interplay between the nature of the problem and the actions Herbert takes to solve it.

Situatedness is closely linked to another constraint on cognition which was not adequately met by the SMPA-style mobots: cognition has to operate in real time to be effective. More computationally efficient strategies are required to deal with fast paced changes that threaten survival. This suggests that traditional reliance on highly detailed representations (often resulting in computational bottlenecks) may be excessive. The virtue of structural coupling is that it allows for more computationally efficient strategies and is, therefore, more plausible from an evolutionary point of view. The efficiency is gained by relying on reactive, situationally-guided behaviour and "adaptive hookups" (Clark, 1997) to reduce the number of representations and computations required to perform a task. Success with situationally determined reactive behaviours prompted Brooks (1991a) to claim that representations are not required for intelligent action. This radical stance has been questioned extensively (Kirsh 1991;

Clark 1997; Wilson 2002). The gist of the opposition is that reactive behaviours are unlikely to scale up to the level of sophisticated cognition. Wilson (2002., p.626) points out that the idea that cognition is situated may simply be overstated:

> [O]ne of the hallmarks of human cognition is that it can take place *decoupled from any immediate interaction with the environment.* We can lay plans for the future, and think over what has happened in the past… In short, our ability to form mental representations about things that are remote in time and space, which is arguably the *sine qua non* of human thought, in principle cannot yield to a situated cognition analysis.(My emphasis)

Kirsh (1991) compiled a short list of activities that cannot be situationally guided. They are what Clark (1997) has termed "representation-hungry" problems, and involve predicting other people's behaviour, planning for future contingencies and so on.

An additional problem with the notion of structural coupling is that it is inadequate for specifying the kinds of systems that can be cognitive. The problem is that it covers non-cognitive cases and is, therefore, overly broad. All physical objects are situated in some environment and to some extent allow for mutually altering interactions with it. For instance, airflow through a room can perturb the position of a door thereby changing amount of airflow in the room. Hence, there is structural coupling between the airflow and the position of the door since each one takes the state of the other as a parameter. Thus, structural coupling needs to be refined.

Quick et. al (1999., p.4) point out that the notion admits to *degrees* of embodiment that "...could be quantitatively measured, for example, in terms of the total complexity of the dynamical relationship between system and environment over all possible interactions." An inanimate object, such as a door in the example above, is embodied in a minimal and rather

uninteresting way. On the other hand, agents that have the means to actively perceive and act upon their environments in pursuit of their goals exhibit a much greater degree of embodiment (Quick et. al., 1999., p.3):

> A biological system bristling with active and sensitive sensory effector surfaces is likely to be much more tightly coupled to its environment than an inanimate object. Highly complex structurally plastic systems such as organisms with nervous systems are likely to exhibit a particularly complex range of behaviours as a result of coupling between nervous system, body and environment.

A higher degree of complexity allows for more ways to register and respond to changes both external and internal to the agent. There are good reasons to think that a high degree of structural coupling may be a necessary condition for embodiment *per se*.

To support this view I wish to point out that only agents that possess the means to perceive and act will exhibit a sufficiently high degree of embodiment. This is tantamount to saying that sensorimotor capacities are necessary for an agent to be considered embodied and this view is explicitly articulated by a number of authors working in EC. Varela et al. (1991., pp.172-173) for instance say the following about embodiment and its relation to cognition:

> By using the term *embodiment* we mean to highlight two points: first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context.

Esther Thelen, a developmental psychologist, well known for her research on infant development provides the following definition of embodied cognition (Thelen et al., 2001., p.1):

> To say that cognition is embodied means that it arises from bodily

> interactions with the world. From this point of view, cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capacities that are inseparably linked and that together form the matrix within which memory, emotion, language, and all other aspects of life are meshed.

The apparent consensus among EC researchers is such that the role of sensorimotor capacities in cognition cannot be understated. Section 1.4.3 will explore this theme more thoroughly with an emphasis on empirical data that underlies the consensus and its ramifications for CR.

An additional benefit of this refined characterization is that it not only highlights the importance of sensing and acting in cognition but also underscores a critical role that the environment plays in shaping cognitive behaviour – a claim which is definitive of EC. A rock will exhibit more or less the same quantitatively identifiable behaviour in any environment, whether it is the desert, or the bottom of the ocean. In contrast, a highly embodied agent will exhibit novel types of behaviour and coping strategies depending on the environment it is in. For example, contrast Herbert's behaviour in an empty room to that in a cluttered MIT lab. An empty room affords little to no opportunity for any interesting interaction. At best Herbert is likely to wander aimlessly only to change direction upon reaching a wall. To a casual observer this behaviour will appear utterly senseless. In contrast, the complex, dynamic setting of the MIT lab together with Herbert's ability to sense and act result in an emergent behaviour of dynamic obstacle avoidance.[6]

Cognition is a situated activity that takes place in a dynamic environment. As a result, cognition has to operate in real time in order to be effective and to have evolved in the first

6 The idea that complexity of behaviour is at least in part a reflection of the environmental (rather than internal) complexity is definitive of EC, but it is hardly new. It was explicitly articulated as early as 1969 by Herbert Simon: “A man, viewed as a behaving system, is quite simple. The apparent complexity of his behaviour over time is largely a reflection of the complexity of the environment in which he finds himself” (Simon, 1969., p.25).

place. Structural coupling between an agent and its environment allows for more computationally efficient strategies by reducing the need for detailed internal representations and allowing some behaviours to be situationally guided. This, however, requires an agent to possess sensorimotor capacities that would allow it to detect and respond to environmental cues. Therefore, embodiment as structural coupling highlights the importance of sensorimotor capacities for cognition and the role of environment in shaping behaviour.

1.4.3 Historical, Organismoid and Organismic Embodiment

There exists a close relationship between the notions of historical, organismic and organismoid embodiment, the second, fourth and fifth notions of embodiment identified by Ziemke. Historical embodiment brings attention to the fact that cognitive abilities are a "reflection of a *history* of agent-environment interaction and in many cases adaptation" (Ziemke, 2003., p.1307). Organismic and organismoid embodiment stress that the nature of cognition is constrained and determined by the *type* of body that an agent has (whether it has legs, wings or fins) and the fact that organisms are *living* entities – both products of evolution. Given these related themes I find it useful to adjust Ziemke's classification and treat these notions together.

The idea that cognition has evolved because it contributed positively to organisms' adaptive success is uncontentious. Anderson (2007., p.65) states "EC treats cognition as a set of tools evolved by organisms for coping with their environment." Interestingly enough, this characterization is almost identical to that proposed by Allen Newell and quoted earlier: "The mind then is simply the name we give to the control system that has evolved within the organism to carry out the interactions to the benefit of that organism or, ultimately, for the

survival of its species" (Newell, 1990., p.43). In both cases the mind is seen as serving the body, or an agent as a whole, in order to promote adaptive success. But, as we shall see, evolutionary considerations have serious ramifications for the theoretical commitments and research focus of the CR paradigm.

Evolutionary history determines the kinds of interactions an agent can have with the world, thereby affecting how it performs cognitive tasks and constraining the "types of cognitive processes available to it" (Cowart, 2004). The physical structure and abilities that have evolved can also be seen as a reflection of the challenges that the organism had to meet along its evolutionary path. Hence, faced with the same general problem, two organisms may employ different strategies to solve it. Consider the differences in how a cat and an adult human would go about trying to retrieve an object from a shelf. A person can typically reach the object without difficulty, whereas a cat, limited by its size and physical abilities, has to find alternative ways to get to the shelf. This may involve using other objects that are well-suited for the kinds of actions a cat can perform. That, in turn, requires the cat to be better attuned to specific features of the environment that a human might disregard. The upshot is that cognition can be markedly different depending on the form of physical embodiment and exploits the existing resources and capabilities that each organism has in virtue of its evolutionary history (Anderson, 2007., p.65):

> Cognition evolved in *organisms with specific physical attributes,* bodies of a certain type with given structural features, and can therefore be expected to be shaped by and to take advantage of these features for cognitive ends.

Therefore, a CR approach to modelling cognition via computations on symbolic representations may reflect a *particular* cognitive strategy that is only applicable in some

domains and may not be available or even useful to all intelligent creatures.

Cognition has evolved rather than having been engineered from scratch: it takes advantage of the existing capabilities of the organism; and it builds on what has proven to be successful in the past. This leads EC theorists to stress that, from an evolutionary perspective, perception and motor control constitute the foundation of all higher cognitive abilities. As a result, they have to be mastered early on. Rodney Brooks (1991a., p.141) notes that the evolutionary history of biological organisms on earth

> [s]uggests that problem solving behaviour, language, expert knowledge and application, and reason, are all pretty simple once the essence of being and reacting are available. That essence is the ability to move around in a dynamic environment, sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction. This part of intelligence is where evolution has concentrated its time – it is much harder [than high level cognitive abilities like problem-solving and language].

Thus, motor control and perception – sensorimotor skills – constitute the basis on which the rest of high-level intelligence is built. This observation challenges the research focus of CR. Rather than trying to understand how humans play chess, acquire linguistic competence, or deduce logical theorems (tasks that may be difficult to generally intelligent people), EC claims that AI research should focus on low-level intelligence. In order to understand intelligence *per se* we have to first understand it in its most rudimentary form. This insight has prompted a growing number of EC researchers to undertake autonomous agent research, and to seek insights from natural sciences regarding problem-solving strategies employed by natural systems (Chiel & Beer, 1997). In addition, a growing number of cognitive scientists find it fruitful to investigate the influences which shape cognition through simulated evolution and

artificial life research. Artificial life is of particular importance if one holds the view that life and mind share common properties (e.g., they are both self-organizing systems; Wheeler, 1997) or, a stronger thesis, that life and mind are somehow continuous (Thompson, 2004).

An understanding of naturally evolved problem-solving strategies has had substantial implications for our ideas regarding the "design" of cognitive systems. As mentioned earlier, evolution builds on prior success and finds ways to utilize the existing capacities to further new goals. As a result, we find that systemic organization of naturally evolved systems often differs wildly from what human engineers and cognitive scientists hypothesize. Often, unexpected systemic organization found in natural cognitive systems undermines theoretical assumptions made by human scientists about what those systems *do* and *how* they do it.[7] The perceptual system is a case in point.

As several authors have pointed out, the common-sense view of perception suggests that seeing is akin to passively taking in a detailed picture of the world, and representing it in full detail on one's 'mental screen' (Churchland, et al., 1994; Noë, 2004).[8] Two factors conspire to make this view attractive. There is an undeniable phenomenological fact about perception – it surely looks to *me* that the world is presented in rich detail. Moreover, an inverted two-dimensional image of the world can be on the retina leading one to think that the visual system deals with pictoral representations. This suggests, then, quite plausibly, that the visual system is responsible for extracting from pictures a detailed description of the external world and representing it internally on our 'mental screen':

"For if we are capable of knowing what is where in the world, our brains must somehow be

7This idea is succinctly captured by what Francis Crick has allegedly referred to as Orgel's Second Rule – Evolution is clever than you are.

8 This is strongly reminiscent of what Dennett (1991) has dubbed the "Cartesian Theater."

capable of representing this information – in all its profusion of colour and form, beauty, motion, and detail"

(Marr, 1982., p.3). Traditional accounts of vision tend to treat perception as "operating essentially independently of other sensory modalities as well as independently of previous learning, goals, motor planning, and motor execution." (Churchland, et al., 1994., p.23).

Empirical research on the visual system together with evolutionary considerations undermine the plausibility of the traditional view. The primary function of vision, from an evolutionary perspective, is to help guide motor actions in real-time. Detailed representations simply get in the way (Brooks 1991a; Ballard 1991; Churchland et al., 1994). Recent research on change-blindness and inattentional blindness also militates against the idea that we have detailed representations of the world (Rensink et al., 1997; Rensink et al., 2000; Simons & Chabris, 1999). Changing our conception of what the visual system does inadvertently casts doubt on the assumption that it functions independently from motor actions and other sensory modalities.

The role of action in perception has been noted by several authors (Noë 2004; Ballard 1991; Chruchland, et al., 1994). The general consensus is that perception and motor control have coevolved, and are intimately interconnected. An integration, rather than segregation, of motor and perceptual functions opens up a wealth of computationally efficient strategies that the visual system can exploit. For example, Ballard (1991) points out that only 0.01% of the visual field area is capable of high-resolution, so the visual system relies on rapid eye saccades and movement to retrieve detailed information (Churchland, et al., 1994., p.27):

> Some motor decisions, such as eye movements, head movements, and keeping the rest of the body motionless, are often made on the basis of

> minimal analysis precisely in order to achieve an upgraded and more fully elaborated visuomotor representation.

This allows for the visual system to sample the environment in coarse detail and selectively attend to features that are of interest to its immediate goals. This provides an agent with the necessary information, without excessive computational costs. Also note that the amount of information that can possibly be extracted from the environment is staggering, which makes it practically impossible for the visual system to avoid being selective with respect to the data it encodes. As a result, it makes the most evolutionary sense for the visual system to be attuned to properties most relevant to the survival and reproduction of the organism in its particular environmental niche. Thus, it is grossly implausible to conceive of vision as being independent of previous learning and goals.

The role of movement in perception has been given extended treatment by Alva Noë, who defends an *enactive* (Varela et al., 1991) view of perception. According to the enactive view, perception is "a way of acting" (Noë, 2004). Thus, perception and its content become intimately linked to our activity in the world: "*What we perceive* is determined by *what we do* (or what we know how to do); it is determined by what we are *ready* to do… we *enact* our perceptual experience; we act it out" (Ibid., p.1). The way we perceive the world – the way it appears to us – becomes dependent on the particular way we are embodied, and the history of our activity in the world.

These ideas resonate with other views held by EC theorists, namely that we *conceptualize* the world primarily in terms of our sensorimotor relation to it. The most prominent example of this view cited in EC literature is the work of Lakoff and Johnson. These

authors contend that our understanding of the world is largely metaphorical and rooted in spatial concepts, since we had to master those early on in order to cope with our environment (Lakoff & Johnson, 1999., p.4):

> the very structure of reason itself comes from the details of our embodiment. The same neural and cognitive mechanisms that allow us to perceive and move around also creature our conceptual systems and modes of reason. Thus, to understand reason we must understand the details of our visual system, our motor system, and the general mechanisms of neural binding.

In support of this view, its defenders often draw attention to the ubiquity of spatial metaphors in our everyday language. We talk about being "ahead" or "behind" work schedule, being "below" or "above" average, "reaching" a goal etc. Murphy (1996) rightfully notes, though, that linguistic evidence can only take one so far, and in the absence of real data it leads to a problem of circularity. Linguistic evidence appears to be both a motivation and evidence for the metaphorical view of concepts. Thus, the question is: are there any empirical predictions that can be derived from the metaphorical view? One could argue that if our conceptual system is grounded in sensorimotor experiences, then the sensorimotor system ought to be implicated in conceptual processing. In fact, there is a growing body of empirical evidence which supports this idea.

Gallese & Lakoff (2005., p.456) argue that "conceptual knowledge is embodied, that is, it is mapped within our sensory-motor system." Their position is supported by a recent discovery that performing, observing and imagining a particular *action* draws on shared neural resources. Naturally the argument in Gallese & Lakoff (2005) focuses on *action-concepts* (e.g., grasp) but the authors believe that it can be extended to concrete concepts (e.g., animate and inanimate objects) as well as abstract concepts (e.g., justice, time). Martin (2007) reviews neuroimaging

studies concerned with how the brain represents salient features of concrete objects and concludes that "specific sensory and motor-based information stored in their corresponding sensory and motor systems." (Ibid., p.38) There is, however, considerably less empirical literature concerning the processing of abstract concepts. Boroditsky & Casasanto (2008), for instance, assumed that the (abstract) concept of time is understood metaphorically in terms of the (concrete) concept of space and attempted "to test whether the asymmetrical dependence of time on space exists even at a more basic level of the human conceptual system." (Ibid., p.581) The critical question, as Barsalou (2007, p.621) put it, is "whether these metaphors simply reflect linguistic convention or whether they actually represent how people think." Boroditsky & Casasanto (2008) found that spatial information modulated judgements of duration but not vice versa, lending support to idea that spatial representations underlie our representations of time. This result is consistent with the idea that we represent abstract concepts by drawing on sensorimotor knowledge.

The thesis that conceptual knowledge is embodied (i.e., at least partially represented by sensory and motor systems) marks yet another important difference between EC and the classical CR. For symbolicists, knowledge was represented by amodal symbols stored independently of the sensory and motor systems. In contrast EC theorists view knowledge as grounded in sensorimotor experience and represented in modality-specific systems (Barsalou et al., 2003b; Barsalou 2008; Gallese & Lakoff 2005).

An increased appreciation of how the brain represents and stores information has cast doubt on some of the assumptions symbolicists have made about representational format used by natural cognitive systems. In a similar fashion, an increased understanding of how natural

cognitive systems tackle everyday problems has revealed that the relationship between cognition and physical action is much closer than previously suspected by symbolists. Physical actions along with stable environmental properties[9] allow us to "off-load cognitive work onto the environment" (Wilson, 2002., p.7). This phenomenon is not just limited to vision, as discussed earlier, but permeates our everyday cognitive activity. For example, instead of trying to remember the grocery list, we write it down. We do it because, unlike memory in a lot of cases, ink and paper are generally long-lasting and easily available for reference. The world effectively functions as an "external memory store" (O'Regan, 1992). Similarly, instead of doing long division or multiplication in our head we use pen and paper to reduce the strain on short term memory.

Cognitive off-loading is the central theme in the work of David Kirsh and Paul Magilo who proposed a distinction between *pragmatic* and *epistemic* actions. A pragmatic action is one that is taken out of necessity, to bring the agent closer to the goal state. In contrast, an epistemic action modifies some aspect of the world in order to change (and hopefully simplify) the nature of the problem we are confronting. Using the game of Tetris as their case study, Kirsh and Magilo have argued that "certain cognitive and perceptual problems are more quickly, easily, and reliably solved by performing actions in the world than by performing computational actions in the head alone" (Kirsh & Magilo, 1994., p.513). The centrality of this perspective to EC is difficult to overstate. To add to that, the work of Kirsh and Magilo is central to the issue addressed in the last chapter of this thesis – the boundaries of mind – and, thus, merits further elaboration.

9 Stable environmental features are those that can be expected to endure through time allowing us to gain easy access to them in case they become relevant. This is why a grocery list is only useful if you can find it. If you can't then you are better off storing it internally.

The distinction between pragmatic and epistemic action is supported by comparative data of human and machine performance in Tetris. The authors have created an expert Tetris-playing system (RoboTetris) based on the traditional approach to problem-solving and compared its performance to that of human players of varying skill levels. RoboTetris exemplified the, by now hopefully familiar, sequential SMPA-style model of problem-solving consisting of several stages. At first the shape of the incoming Tetris block, or *zoid*, is "perceived" by the system and subsequently encoded into a symbolic representation. At the next stage the system calculates the best fit between the current Tetris block and the existing structure. Lastly, the sequence of appropriate rotations and sideways movements of the block is calculated in order to bring it to the desired position. Processing is sequential with action (i.e. rotation, sideways movement) being the final step which necessitates a pre-determined plan. The function of action in this model is strictly pragmatic. It simply takes the player closer to the goal and serves no identifiable function in the absence of a plan. The data on human performance, however, appears to challenge the accuracy of this model. Kirsh & Magilo found that "rotations sometimes begin extremely early, well before an agent could finish thinking about where to place the zoid" (Kirsh & Magilo, 1994., p.524). The number of rotations performed by human subjects far outstrips the number that is expected on the assumption that rotation only serves a pragmatic function. Kirsh & Magilo (1994., p.526) contend that this discrepancy

> is a direct consequence of the assumption that the point of action is always pragmatic: that the only reason to act is for advancement in the physical world. This creates an undesirable separation between action and cognition... On this view, cognition is logically prior. Cognition is necessary for intelligent action, but action is never necessary for intelligent cognition.

The authors admit that it can often be difficult to tell whether an action is epistemic or pragmatic since the same action can fulfill both purposes. Nonetheless, there are still clear cases where an action takes one further from the goal-state and, therefore, it can only be plausibly interpreted as an erroneous move or an epistemic action. An example of that would be a full rotation of the block so that at the end it returns to its original position. In that case the player is back at the starting point albeit with less time to make the decision where to place the block. An epistemic action interpretation is supported by the fact that it takes a substantially longer time to rotate a piece mentally, rather than on screen. Furthermore, Kirsh and Magilo found that advanced players often rotate an incoming block *before* it is in full view thereby revealing the obscured part faster than it would have been otherwise. Thus, action is not viewed as a merely necessary final step in problem solving preceded by perception and planning but rather as an integral part of the cognitive process. The primary function of epistemic action is to change the input to the agent's cognitive system. On this account motor actions in perception discussed earlier can also be seen as species of epistemic action. On the other hand, Herbert's actions during the obstacle avoidance routine serve pragmatic and epistemic purposes simultaneously. Once an obstacle is detected Herbert halts and then turns in a different direction. This allows the sonar to probe a different direction which, if not obstructed as well, will allow Herbert to navigate around the initial obstacle. Hence, turning serves a dual purpose of altering the received input and being pragmatically necessary for getting around an obstruction.

An evolutionary perspective on cognition reveals a much closer relationship between bodily elements and cognitive abilities than was previously suspected. The form of physical

embodiment, for example, determines how an organism interacts with the world which renders some cognitive abilities/strategies more useful than others. To add to that, physical action appears to be a essential for tackling computationally complex problems by allowing us to off-load cognitive work onto the world.

Furthermore, an evolutionary perspective on cognition suggests that mechanisms underlying sensorimotor capacities are the foundation of higher cognitive functions. Thus, EC theorists suggest that we represent and reason about the world in terms of our sensorimotor relation to it. The upshot is that cognitive scientists should concentrate on low-level intelligent behaviour and to study problem-solving strategies used by naturally-evolved agents in order to fully understand the foundations of high-level cognition.

1.4.4 Physical Embodiment

The third of Ziemke's six notions is *physical* embodiment which focuses on the necessity of a physical body for cognition. EC obviously stresses the role of *body* and *environment* in cognitive performance. Hence, it is natural to assume that *a* physical body, situated in an environment is at least a necessary condition for embodiment *per se*. However, it has been argued that embodiment is not necessarily limited to physical objects (Franklin 1997; Quick et. al., 1999; Etzioni 1993). Consider the following quotation from Quick et. al (1999., p.2)

> There exists a pervasive assumption that only materially endowed entities can be 'embodied'…Unless there is some mysterious quality possessed by material things beyond those that follow from the way that being a material thing in a material environment conditions and shapes interaction, there is no need for embodiment to be ontologically constrained.

Etzioni (1993) has argued that appropriately programmed software robots (softbots) living in a

software environment can be structurally coupled to it. So they can be used to study intelligence just as well as Rodney Brooks' mobots, but at a much lower cost. Softbots are autonomous, goal-driven agents that have to cope with constant changes in their dynamic software environments. For example, SUMPY (Song et al., 1996) programmed using Brooks' subsumption architecture is described as "[A] software agent "living" in and helping to maintain a UNIX file system for better disk space utilization by compressing and backing up" (Song et al., 1996., p.1). SUMPY "wanders" around UNIX directories detecting, compressing and backing up the necessary files.

While software agent research may be a complementary avenue for studying intelligence, it does not militate against research associated with physically embodied systems; the two can exist side by side. However, physical embodiment, at least at this point in time, is the primary focus of EC research. There are at least two reasons for this. First, proponents of EC tend to contrast their approach to that of CR, which allegedly treats the mind as a disembodied computer. Instead EC researchers emphasize the fact that cognition has evolved naturally as organisms learned to cope with their environment. Hence, to understand cognition one has to concentrate on its evolutionary roots. Still, one could insist that software environments provide an opportunity for that as well. There has been a growing interest in genetic algorithms, that simulate evolution, and artificial life research. But these, at best, provide an approximation of "the real thing." Simulated environments obviously lack the complexity and diversity of the real world, and their design and implementation can be laden with potentially distorting assumptions made by their creators. For example, consider an attempt by Zaera et.al (1996) to use a genetic algorithm to evolve schooling behaviour in

software fish. The authors admit that they did not create a realistic simulation leaving hydrodynamics, light, temperature and other environmental variables out of the picture. Software fish were only sensitive to each other's position. The failure to evolve the schooling behaviour, among other things, may have been due to the impoverished simulation of the ecosystem Zaera et al. (1996., p.643):

> It could reasonably be argued that real schooling behaviours in real animals arise because of the complex interaction of a number of factors, and that our approach failed because the simulations lacked sufficient complexity.

To avoid this problem software simulations have to be either extremely realistic from the start or researchers must have extensive knowledge of the factors that are relevant to the behaviour being modelled.

Perhaps the most obvious reasons for treating physical embodiment as central to EC are purely pragmatic. Physically embodied cognitive agents are the paradigm of natural intelligence studied by a number of disciplines. This encourages interdisciplinary collaboration among psychologists, neuroscientists, biologists, ethologists, and so on. The natural, physical, world also provides a plethora of various species spread out across the evolutionary continuum affording a great diversity of examples of the development of cognition. For example, a comparative study of our nearest evolutionary predecessors can help shed light on the mechanisms that separate us from them. An improved understanding of how brains and bodies shape cognition will in turn contribute to, among other things, better medical treatment, educational methods, and artifact design. Consequently, it is only natural to treat physical embodiment as the primary object of research in EC, as it has the most potential for advances in both theoretical and practical aspects of understanding and improving cognition.

It is critical, however, to refine the notion of physical embodiment. As noted by Ziemke (2003), physical embodiment by itself is not a restrictive enough notion for capturing the essence of EC. It again does not demarcate cognitive systems from the non-cognitive ones; tables, chairs and teddy bears have physical bodies, but that does not render them cognitive. What is required is a set of conditions that will delineate the *kinds* of physical bodies that can sustain genuine intelligence from the ones that cannot. Previous discussion revealed several factors that constrain the class of physical objects that can count as embodied. Sensorimotor capacities appear to be necessary for a system to count as embodied. The kind of body an agent has certainly affects available cognitive strategies, but there are no specific requirements as to whether it has to have, say, arms or fins to be embodied. Presumably any body type that is capable of sensorimotor interaction with the environment would be sufficient. The organismic notion of embodiment is perhaps the most restrictive one of all, since only living systems can count as embodied. Precisely how restrictive this notion is remains unclear due to a lack of an agreed upon definition of what counts as a living system. For example, it is debatable whether *reproduction* is necessary for something to count as alive. If that were to be so, then none of the robotics research in EC would count as research on embodied cognition.[10] For this reason I am inclined to think that, at least at this point, limiting embodiment to living systems is too restrictive.

Treating physically embodied agents with sensorimotor capacities as the primary focus of EC research is the most theoretically and pragmatically justified choice. It facilitates a break from the classical view of cognition as disembodied computation and appears quite natural

10 It was pointed out to me that mules are also incapable of reproduction and, therefore, would not count as alive. This constitutes a strong reason to regard the reproduction criterion as overly restrictive.

because naturally evolved, physically embodied agents are the paradigm of natural intelligence.

1.4.5 Social Embodiment

The last of Ziemke's notions, social embodiment, is given a considerably shorter treatment than others. The notion stems from the work of Barsalou et al., (2003a) who use 'embodiment' to mean "that states of the body, such as postures, arm movements, and facial expressions arise during social interactions and play central roles in social information processing" (Ibid., p. 43). The authors review a wide range of studies which demonstrate several types of 'embodiment effects'[11] – ways in which bodily states influence cognitive and/or affective states and vice versa. For example, seeing a facial expression of an affective state (e.g. disgust) activates a brain area involved in experiencing that state first-hand (Wicker et al., 2003).

Social psychology research reviewed by Barsalou (2003a) reveals a number of unintuitive ways in which the bodily, cognitive and affective states modulate each other. Nonetheless the label 'social embodiment' is somewhat unfortunate in the present context. Various notions of embodiment discussed earlier have the common virtue of shedding light on the conditions necessary for cognition. Social embodiment, on the other hand, merely extends, to the social domain, ideas and themes previously discussed in section 1.7. Chief among them is the familiar idea that sensory and motor systems are not peripheral to cognition. Consequently, social embodiment is best seen as a particular application of this more general theoretical outlook rather than a separate notion that can help elucidate the conditions necessary for cognition.

11 Four types of embodiment effects are discussed: "First, perceived social stimuli do not just produce cognitive states, they produce bodily states as well. Second, perceiving bodily states in others produces bodily mimicry in the self. Third, bodily states in the self produce affective states. Fourth, the compatibility of bodily states and cognitive states modulates performance." (Barsalou et al., 2003a., p.43)

1.5 Summary

Since its inception in the 1950s, the field of Cognitive Science has been dominated by the CR approach to studying cognition. Philosophers and cognitive scientists found it attractive to treat cognition as an abstract, knowledge-based computational process, because it allows for the study of cognition in abstraction from implementational details. Success of early AI research in modelling some of the most impressive feats of human intellect encouraged the idea that other cognitive functions will succumb to the same approach. Nonetheless, CR has been extensively criticized by cognitive scientists and philosophers alike. The criticism that cognition is not just disembodied computation localized in the brain has been of particular interest. As Varela et al., (1991) point out, philosophers of the continental tradition have long emphasized that the mind is always situated in and actively engaged with the world. Likewise, albeit for empirical reasons, cognitive scientists have come to appreciate that a proper account of how we perform cognitive tasks has to make reference to the environment, body and the history, individual and evolutionary, of their dynamic, real-time interaction. There is, therefore, a wide agreement among researchers in different fields that cognition is necessarily embodied. However, as Ziemke (2003) has shown, there is an alarming lack of consistency in how the notion of 'embodiment' is used. In this chapter I have discussed the six notions of embodiment outlined by Ziemke and their impact on CR: structural coupling, organismic, organismoid, historical, physical and social. I have argued that each notion, at most highlights an aspect of embodiment rather than being an exhaustive definition of the concept. In the remainder of this section I summarize the most salient points of the preceding discussion in order to make the notion of embodiment and its implications for CR more evident.

I have shown that ideas in EC are greatly informed by an evolutionary perspective on cognition and an enriched understanding of how naturally evolved agents perform cognitive tasks. I have demonstrated that this perspective has several implications for CR: it suggests a rethinking of the role and format of representations in cognitive systems; it challenges the research focus of AI; and it calls into question the segregation of cognition from bodily and worldly activities. However, I have argued that none of these implications are fatal to CR.

EC theorists treat cognition as a set of tools which has evolved because it promoted organisms' survival and reproductive success in hostile environments. This perspective highlights two constraints on cognition that have not been adequately met by CR: (1) cognition is situated; and (2) cognition is time-pressured. In other words, agents are constantly engaged with their environments in goal-relevant ways and have limited time to react. As a result, traditional approaches to modelling behaviour that relied heavily on detailed representations of the world to mediate perception and action were too slow to meet these constraints and were deemed evolutionary implausible. In section 1.4.2 I have claimed that structural coupling allows for more computationally efficient strategies by reducing the need for detailed internal representations and allowing some behaviours to be orchestrated by the environment. I have also argued that the effectiveness of structural coupling in dealing with the aforementioned constraints on cognition does not warrant a radical conclusion that internal representations are not required for intelligent action. Rather, it implies that the extent to which a cognitive system has to rely on internal representations to produce behaviour has been overstated. I have subsequently argued that structural coupling necessitates sensorimotor capacities which would allow an agent to detect and respond to external stimuli, thereby allowing environmental

factors to guide behaviour. Therefore, embodiment as structural coupling highlights the necessity of sensorimotor capacities for cognition, and a substantial role for the environment in shaping behaviour.

An evolutionarily perspective on cognition underlies three of the six notions of embodiment discussed in section 1.4.3: historical; organismoid; and organismic. All of them reflect the tendency among EC theorists to pay much closer attention to problem-solving strategies used by naturally evolved cognitive systems. That, in turn, leads to a re-evaluation of assumptions made by CR theorists regarding the role of bodily elements in cognition, and the relationship between cognition and external environment. Proponents of EC stress that cognition is a reflection of agent-environment interaction, and is simultaneously *constrained by* and *takes advantage of* the particular form of embodiment that has evolved in response to evolutionary pressures. To add to that, the way organisms represent and think about the world is ultimately rooted in the sensorimotor experiences that come with having a particular body. As a result, cognition is closely tied to the form of physical embodiment and actively exploits physical actions and the physical environment to further cognitive goals. This point typifies EC and stands in stark contrast to the widely shared assumption in CR that cognition is a purely computational process that can be studied without recourse to implementational details. It should be noted, though, that EC theorists generally stop short of stating what *kind* of body is necessary/sufficient for an agent to be considered embodied. In section 1.4.4 I looked at the highly counterintuitive idea that embodiment does not necessitate a physical body. Ultimately I argued that treating physically embodied agents with sensorimotor capacities as the primary focus of EC research is the most theoretically and pragmatically justified choice. I have also

briefly considered and argued against the proposal which limits embodiment to *living* organisms. Instead, I suggested that any body capable of sensorimotor interaction with the environment would be sufficient for embodiment.

The discussion in section 1.4.2 revealed a wide agreement among EC researchers that sensorimotor capacities constitute the foundation for all higher cognitive abilities. Thus, sensorimotor capacities and brain mechanisms that realize them are expected to figure in higher cognitive functions. In section 1.4.3 I argued that this perspective challenges the research focus of CR, and its commitment to modality-independent representations. Therefore, instead of trying to understand how humans play chess or deduce logical theorems, cognitive scientists ought to focus on low-level intelligence exhibited by autonomous agents. And that, in turn, might shed light on the role of sensorimotor system in higher cognitive functions, such as abstract thinking.

In conclusion, EC research emphasizes the evolutionary role of cognition and attempts to understand how the environment interacts with the internal capacities of physically embodied agents to support cognitive functions. I have demonstrated that this perspective is compatible with CR, but it suggests a rethinking of the role and format of representations in cognitive systems and challenges the research focus of traditional AI. In the upcoming chapters I pursue the theme of environment/agent interaction and relate it to two issues in Philosophy of Cognitive Science, namely, the nature of explanation and the boundaries of mind.

Chapter 2

Embodied Cognition and Mechanistic Explanation

2.1 Introduction

An evolutionary perspective on cognition has led EC researchers to focus on naturally evolved cognitive systems and the multitude of ways agents exploit their bodies and environment to further cognitive goals. This shift has precipitated a skeptical attitude towards *componential* (Clark, 1997) or *mechanistic explanation* (ME) (Bechtel and Richardson, 1993) that is generally invoked to account for the workings of complex systems. This chapter is structured as follows. In section 2.2 I provide a thorough account of mechanistic explanation and its relation to the CR paradigm. Section 2.3 looks at the skeptical attitude regarding the applicability of mechanistic explanation to embodied and embedded cognitive systems. I consider the view that systemic organization of natural cognitive systems is not conducive to a decompositional strategy, which is definitive of mechanistic explanation. In section 2.4 I describe Dynamic Systems Theory – a mathematical framework proposed as a viable alternative to mechanistic explanation. In section 2.5 I argue that the dynamicist perspective is compatible with mechanistic explanation. Moreover, I contend that our capacity to understand how complex systems work relies almost exclusively on our ability to decompose them into component parts, rendering mechanistic explanation indispensable.

2.2 Mechanistic Explanation

What an explanation appeals to depends on the explanatory target. For example, a folk

psychological explanation of my actions will appeal to my goals, desires, beliefs and so forth; behaviours of elementary particles are best explained by appeal to the laws of physics that govern matter and energy. On the other hand if one wants to explain the functioning of a complex system made up of such particles, then a different mode of explanation may be appropriate. In disciplines such as biology or psychology, among others, *mechanistic explanations* are favoured. Such explanations "...propose to account for the behaviour of a system in terms of the functions performed by its parts and the interactions between those parts" (Bechtel & Richardson, 1993., p.17). Hence, a system is seen as a collection of *mechanisms* that produce behaviours in virtue of how their component parts are put together (Bechtel & Abrahamsen, 2005., p. 423)

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena.

The virtue of mechanistic explanation is that it facilitates intelligibility of and control over the system's behaviour. Reducing a complex system to the organization of functionally simpler constituents renders the overall functionality more transparent. Understanding how components of a system contribute to its functionality allows us to selectively intervene in their operation and alter how the system behaves.

Two heuristics guide the development of a mechanistic explanation: *decomposition* and *localization* (Bechtel & Richardson, 1993). The former refers to the task of determining the functions that a system has to carry out in order to achieve the desired functionality. The latter heuristic involves the identification of mechanisms that perform these functions within the

actual system under consideration. When little or nothing is known about the system, decomposition may amount to nothing more than a hypothesis about how the system might work. The simplest case of decomposition attributes each hypothesized function to an individual component, or *module,* and assigns little to no explanatory role to their organization. Interaction among components is limited to exchanging inputs and outputs without affecting the functions they perform individually. As a result, individual components can be studied in isolation from the rest of the system without distorting the role they play in it. Systems that are in fact organized this way are referred to as *aggregative* systems.

In general, the organization of a system matters in three ways – it can affect the functionality of any given component, the system as a whole, or both. Consider a car as an example of an aggregative system where component parts have distinct and clearly identifiable functions. Component parts that make up a car have to be put together correctly for the car to function the way it does. In this sense the organization of the system affects its overall functionality. At the same time, however, organization has little effect on the intrinsic functionality of the components. The function of an engine is to convert gasoline into mechanical energy regardless of whether it is actually connected to any other auxiliary components. But, "When the relevant systemic properties are at least partially determined by the organization of the system, we no longer have aggregativity" (Bechtel & Richardson, 1993., p.25). Indeed, complex systems, natural and artificial, are rarely aggregative - they owe their functionality, at least in part, to how the components are organized. In the most extreme case we encounter *integrated systems* in which (Bechtel & Richardson, 1993., p. 26)

> systemic organization is significantly involved in determining constituent functions. There may be, for example, mutual correction

> among subsystems, or feedback relations that are integral to constituent functioning.

To highlight the difference between aggregative and integrated systems, think of how a simple neural network differs from a car in its organization. As I have mentioned, the car consists of easily identifiable components, each having a function that contributes distinctly to the system's behaviour. A neural network, on the other hand, consists of functionally homogenous components – model neurons. Thus, if we just look at what individual neurons are doing, an explanation of their *collective* behaviour is likely to elude us; their organization carries a larger explanatory burden. The contrast drawn here between aggregative and integrated systems is somewhat artificial. The difference is a matter of the *degree* to which facts about organization contribute to the explanation of the overall behaviour of a system. All systems fall somewhere on the continuum between aggregative and integrated.

Decomposition and localization play a central role in the development of a mechanistic explanation. However, logically prior to these tasks is the challenge of delineating the system, credited with producing the behaviour to be explained, from the external environment. In other words, prior to decomposing the system into parts one must define what counts as the system under consideration (Bechtel & Richardson, 1993., p.39):

> The scientist must segment the system from its context and identify the relevant functions assigned to it. To substantiate the assignment of a function to a system, the scientist generally must offer theoretical or empirical arguments showing that the physically and functionally independent system identified has substantial internal control over the effect. This is what we describe as treating the system as the locus of control for a phenomenon.

Bechtel and Richardson stress that identifying a system as the *locus of control* does not mean that facts external to it are *irrelevant* for its functioning or that the system can function the way

it does in isolation. As an example, they note that for a car to function there has to be a driver, mechanic, fuel source, etc. Nonetheless, the processes that transform fuel into motion take place inside the car. This, the authors contend, is sufficient to treat the car as the locus of control.

There is a strong relationship between the mechanistic approach to understanding (and building) complex systems and the CR paradigm, which is why, as we shall see, the same theorists often reject both at once. Indeed, Beer (1995) argues that if computationalism is to be taken as an empirical hypothesis, it must be making a claim about *internal organization* of cognitive systems. Beer argues that the CR framework is best suited for systems that exhibit a specific form of organization such that

> [t]hey have reliably identifiably internal configurations of parts that can be usefully interpreted as representing aspects of the domain in which the system operates and reliably identifiable internal components that can be usefully interpreted as algorithmically transforming these representations so as to produce whatever output the system produces from whatever input it receives (Beer, 1995., pp.127-128).

This intuition seems to be widely shared, and in part justified by the accepted practices in engineering and building AI systems. Recall the traditional view of vision discussed earlier. Whatever information the vision system collects has to stand-in for external features such that the planning system could respond appropriately. By the same token, the 'planning module' would pass information to the 'motor system module' and that information is most usefully interpreted as representing the sequence of actions that the motor system has to carry out.

Mechanistic explanation involves three stages: identifying the system which is the locus of control for the phenomenon to be explained; decomposition of the system into component operations which collectively produce the phenomenon; and localization of parts within the

system that perform those operations. An explanation of the phenomenon will appeal to the functionality and organization of the system's component parts. Every system falls somewhere on the continuum between aggregative and integrated depending on the degree to which facts about its organization contribute to the explanation of its workings. The primary virtue of a mechanistic explanation is that they facilitate intelligibility and control over the system's behaviour.

2.3 Embodied Cognitive Systems and Complex Organization

Traditionally, mechanistic explanations of cognitive phenomena treat the agent as the locus of control. An account of internal mechanisms of the agent is regarded as sufficient to explain behaviour. Supporters of the idea that cognition is embodied tend to distance themselves from this view in two ways. The less radical objection is that the brain and/or nervous system is not the correct unit of analysis for the purposes of explaining an agent's behaviour. Instead a more comprehensive system comprising the agent's brain, body, and world should be regarded as the locus of control (Clark & Chalmers 1998; Chiel & Beer 1997; van Gelder & Port 1995a; Haugeland 1998).

> If the significant complexity of intelligent behaviour depends intimately on the concrete details of the agents embodiment and worldly situation, then perhaps intelligence as such should be understood as characteristic, in the first instance, of some more comprehensive structure than an internal, disembodied “mind”, whether artificial or natural (Haugeland 1998., p.211)

The more radical position argues that a proper account of features characteristic of this more comprehensive system requires a new conceptual framework – the Dynamic Systems Theory (van Gelder & Port, 1995a).

Skepticism regarding the applicability of a decompositional strategy is natural given the research focus associated with EC. EC researchers have become increasingly interested in problem-solving strategies employed by evolved (naturally or otherwise) cognitive systems which are not restricted by the aggregative design methodology that pervades traditional AI. Wheeler (1998) notes that while the decompositional strategy works well for well-defined problems, assuming they are representative of problems in the natural environment may be a dangerously distorting abstraction. If we are to try and embed an autonomous agent in the real world, then

> ...the problems confronted by those artificial agents will also be difficult to specify and unavoidably open-ended. Thus any attempt to identify a set of well-defined tasks and sub-tasks for putative inner homunculi to perform may be doomed to failure. (Wheeler, 1998., p.98)

Even if we could initially define some set of problems sufficiently well, and equip an autonomous agent with the capacities to solve them, it would not provide a guarantee of long-term success. Over time, the world that poses these problems will change, likely rendering the agent's inbuilt capacities obsolete. Skepticism regarding the applicability of mechanistic explanation is clearly voiced by Clark (1997)[12], Wheeler (1998) and van Gelder & Port (1995a), and is rooted in the idea that systemic organization of natural cognitive systems exhibits a degree of interactivity and complexity that obliterates useful distinctions among components. Since component parts are tightly coupled and constantly modulate each other's activity in real-time, it becomes practically impossible to discern what their individual

12 Although Clark does not dismiss the utility of a mechanistic approach and does not align himself with the more radical thinkers.

contributions are[13] (van Gelder & Port, 1995b., p.13):

> The cognitive system does not interact with the body by means of periodic symbolic inputs and outputs; rather, inner and outer processes are *coupled,* so that both sets of processes are continually influencing each other. Cognitive processing is not cyclic and sequential, for all aspects of the system are undergoing change all the time.

Roughly the same point is made by Clark (1997) and Wheeler (1998) where they appeal to "reciprocal causation" as one of the features of natural systems that is likely to undermine the mechanistic approach. Reciprocal causation allegedly poses a threat to mechanistic explanation because one cannot abstract any part of the system from the overall functionality. Whatever any given part does, necessarily depends on the workings of the entire system. In the same spirit Beer (1995., p.128) states that

> the conceptual framework of computation seems to work least well for highly distributed and richly interconnected systems whose parts do not admit of any straightforward functional decomposition into representations and modules which algorithmically manipulate them

In effect, as the system increases in complexity, and most of the explanatory burden falls on unintuitive connections and feedback loops, it becomes increasingly difficult, or futile, to undertake a decompositional strategy.

2.4 The Holistic Perspective and Dynamic Systems Theory

The natural reaction to these problems is to shy away from a decompositional approach and treat the system holistically. A Dynamic Systems Theory (DST) framework is proposed as a viable candidate that can deal with features that seem to be problematic for mechanistic

13 However, note that this line of argument does not logically entail that decompositional strategy fails automatically or that it is somehow incoherent to speak of mechanisms as collectively producing some phenomenon. It only implies that decompositional strategy is severely complicated by the high degree of interactivity among components.

explanation (van Gelder & Port, 1995a).

Within a DST framework, a *dynamical system* is a collection of interdependent[14] variables that are assigned numerical values that change over time. At any point, the *state* of a system corresponds to the values of these variables. A collection of all possible states defines the *state space* of a system. Transitions between states is captured by differential equations, and correspond to the behaviours of the system as a whole. When it comes to applying the DST perspective to cognition, cognitive performance is taken to be a behaviour of some dynamical system (an abstract mathematical construct) instantiated by an agent (van Gelder, 1999., p.11)

> there is some set of variables associated with me (and the relevant environment) which constitute a dynamical system of a particular sort, and the cognitive performances are behaviours of that system.

Such cognitive systems operate under a real-time constraint, and owe much of their functionality to a high degree of structural coupling with the external world. The strength of the DST perspective is that it allows for easy integration of two systems, effectively treating them as a unified whole. In addition it necessarily treats cognition as operating in real-time. Indeed, the concepts of situatedness/embeddedness are of central interest to dynamicists. The dynamical perspective also appears appropriate for EC in light of their mutual interest in naturally evolved systems. Timothy van Gelder, a leading proponent of DH, states that "[t]he proper domain of the DH is *natural* cognitive agents—i.e., evolved, biological agents such as people and other animals" (van Gelder, 1998., p.619). In addition to the seemingly natural fit with EC DST also has the advantage of being a well-explored mathematical theory and has established techniques that can be used to analyze behaviours of complex systems (Eliasmith, 1996).

14 A change in any variable within the system is a function of other variables internal to that system.

2.5 Dynamics, Mechanisms and Decomposition

The complexity of interactions and lack of neat modular organization in complex natural systems has led some theorists towards adopting a holistic perspective that employs the tools of DST to explain cognitive performance. *Prima facie* this perspective appears with be incompatible with a mechanistic approach. This, however, is not obviously so. Bechtel (1997) makes the point forcefully by arguing that van Gelder's favourite example of a dynamical system, the Watt governor, can be broken down into distinct components in a way that facilitates understanding how it works. However, Bechtel's analysis has an important shortcoming: the Watt governor was *explicitly designed* by a human engineer, which may be why it is amenable to a decompositional strategy. Therefore, despite seeing great value in Bechtel's analysis, I do not think that the Watt governor presents an ideal case for defending the decompositional strategy from the aforementioned skepticism. In lieu of the Watt governor I propose to look at a *natural* dynamical system to see whether a decompositional strategy can play a significant role for improving our grasp of it.

One of the examples of natural dynamic systems cited by van Gelder & Port (1995b) is the solar system. The planets, the sun, and asteroids all exert mutually altering influences, the extent of which depends on their individual properties, such as mass and position. The system is self-contained and deterministic; given differential equations that describe the solar system, we can deduce a future value of some variable, say Earth's position, based on the initial values of all the variables that make up the system. It is undeniable that DST is of great value for the purposes of describing the motions of heavenly bodies. Yet, it is equally undeniable that logically prior to using the tools of DST one has to know the properties of *individual planets*.

The influence a planet can exert on other bodies, and by extension the effect it has on the behaviour of the *entire dynamical system,* is greatly constrained by its individual properties. Understanding those properties and their principled relation to the behaviour of the whole system is vital for the predictive power DST provides. Therefore, there is nothing about the dynamicist perspective which precludes a decompositional strategy. Moreover, breaking the system down to its component parts appears to be necessary for the application of DST.

Further opposition to a mechanistic approach has been grounded in the idea that evolution is not constrained by human design preferences – hierarchical and modular organization. Even if that is true, it does not follow that such organization is rare, implausible or absent in nature. In fact, quite the opposite may be true. Let us consider theoretical and empirical considerations presented by Simon (1969). Simon (1969) draws attention to the fact that a body of an animal is made up of trillions of cells of various kinds and their differences facilitate the functions they contribute to. Cells ultimately constitute tissues and organs that perform distinct functions that contribute to the overall fitness of an animal. Looking at an animal as a complex system, the organs are clearly identifiable components that to a large extent specialize in function: hearts pump blood; and lungs extract oxygen from air. Simon extends this analysis to non-biological systems as well, and points out that at the microscopic level we find elementary particles, atoms, molecules, and so on. Hence, despite the fact that evolution is not bound by human design preferences, it is clear that hierarchical and componential design is ubiquitous in complex biological and non-biological systems.

Simon argues that hierarchical systems are ubiquitous in nature because this form of organization is more conducive to evolution (Simon, 1969., pp.98-99):

> complex systems will evolve from simple systems much more rapidly if there are intermediate forms than if there are not. The resulting complex forms in the former case will be hierarchic. We have only to turn the argument around to explain the observed predominance of hierarchies among the complex systems nature presents us. Among possible complex forms, *hierarchies are the ones that have the time to evolve.*

Simon introduces a now famous parable of the two watchmakers, Tempus and Hora, that make highly sophisticated watches consisting of 1,000 parts each. Tempus' watch only works once all the pieces have been put together. Hora, on the other hand, first creates ten sub-parts which, when put together, make a functional watch. Any interruption destroys the progress of whatever they are working on. Simon provides a quantitative analysis of the time it would take Tempus and Hora to create one watch given the probability p =0.01 that an interruption will occur. He shows that "it will take Tempus on average about four thousand times as long to assemble a watch as Hora" (Simon, 1969., p.92). Keep in mind that the target system of this analysis consists of 1,000 parts and performs a rather modest function. In contrast, a human body consists of trillions of cells (estimates ranging anywhere between 50 and 100 trillion). Thus, disregarding the complexity of the cells themselves, it would taking an astonishingly long time for an organism to evolve if it did not exploit hierarchical design, and did not consist of stable sub-systems that to some extent specialize in function.

To this point I have defended the virtue and applicability of a mechanistic approach from the allegations that it is incompatible with dynamical and naturally-evolved systems. To wrap up this section I want to briefly discuss what I take to be a well-grounded criticism of mechanistic explanation and provide some positive considerations in its favour. In highly complex systems, organization and unintuitive connections among parts take up most of the

explanatory burden. Therefore, mechanistic explanation becomes increasingly difficult as the complexity of the system grows. Let us consider the dynamicist alternative to see whether it can handle this objection. Recall that dynamicists associate cognitive performance with the temporal behaviour of some dynamical system which is defined as a collection of variables relevant to that specific cognitive function. To account for some aspect of cognition, they first have to specify the relevant variables for the equation(s) that govern the system's behaviour over time. Note that this task alone may be just as, if not more, difficult as disentangling a complex web of causal influences among components. To add to that, Eliasmith (1996) points out the complexity of the brain together with environmental factors that affect cognitive performance conspire to make the number of relevant variables astonishingly large[15]. Even if we could solve systems of equations with that many variables, the utility of such an exercise is highly questionable. It is simply unclear what it means "for the system to move through the trillion dimensional (to be conservative) state space" (Eliasmith, 1996., p.317). Hence, even though systemic complexity certainly poses a significant challenge to mechanistic explanation, the DST alternative does not fare better.

Somewhat ironically, the chief virtue of mechanistic explanation is that it has the most potential for rendering a system analytically transparent (Simon, 1996., p.108).

> The fact then that many complex systems have a nearly decomposable hierarchic structure is a major facilitating factor enabling us to understand, describe, and even "see" such systems and their parts. Or perhaps the proposition should be put the other way around. If there are important systems in the world that are complex without being hierarchic, they may to a considerable extent escape our observation and understanding.

Our ability to grasp how a vastly complicated machine performs its function relies almost

15 Eliasmith (1996) argues that this number could potentially be in the trillions.

exclusively on our ability to break it down into parts.

2.6 Conclusion

The most prominent theme throughout the preceding discussion is the opportunistic nature of cognition. Organisms exploit the environment in a number of ways in order to simplify and further cognitive goals. In this chapter I have examined the impact this has had on the debate as to which explanatory framework is best suited for understanding cognition. I have argued that the allegedly problematic features of evolved embodied cognition, non-aggregative design, and continuous interaction with the world, do not render ME inapplicable. Further, I have argued that an alternative explanatory framework, namely DST, necessitates a decompositional strategy, despite claims to the contrary, and, therefore, cannot replace ME. In addition to defending ME from such skepticism, I have looked at positive considerations in favour of ME provided by Simon (1969). Specifically, I have considered the fact that natural physical systems exhibit hierarchical organization and are, therefore, well-suited to a decompositional approach. The upshot is that breaking a system down into component parts is integral to our ability to make a truly complex system analytically transparent.

But, the opportunistic nature of cognition may carry more severe implications for Philosophy of Mind. An appreciation of how agents interact and exploit their environment during cognitive activity has prompted some thinkers to reconsider the traditionally accepted boundaries of the mind. In the next chapter, I introduce and critically evaluate the thesis that mind and cognition extend beyond the bodily boundary and into the world. Ultimately, I argue that this thesis is unsupported. In addition, I provide empirical considerations that favour of the

conventional view that worldly features are not constitutive of cognition and, therefore, that the mind does not extend into the world.

Chapter 3

(Over)Extended Mind

3.1 Introduction

The first chapter of this thesis was concerned with explicating the foundations of traditional and contemporary directions of research in Cognitive Science. I have looked at the CR paradigm and contrasted its central assumptions and research focus with a relatively new line of research - EC. Having argued that EC does not threaten CR as the dominant paradigm I have subsequently described the impact of EC research on Philosophy of Cognitive Science. The previous chapter examined how the EC focus on evolved, environmentally embedded cognitive systems affected explanation. I argued that, contrary to some radical suggestions, EC research does not seriously challenge the traditionally accepted ME. In this chapter I look at ramifications of EC research on Philosophy of Mind. Specifically the thesis that mind and cognition are extended beyond bodily boundaries and into the world.

This chapter is structured as follows. In Section 3.2.1 I summarize two representative statements of the Extended Mind (EM) thesis – Clark and Chalmers (1998) and Hutchins (1995a,b). In Sections 3.2.2–3.2.4 I critically evaluate their arguments for EM and demonstrate two things: first, that they fail to undermine traditionally accepted boundaries of mind; and second, that conceptual difficulties associated with EM thesis severely weaken its plausibility. Subsequently, in Sections 3.3–3.3.2 I situate the issue within a mechanistic framework discussed in Chapter 2 and show that there are strong empirical reasons for favouring traditional boundaries of mind. Section 3.4 examines how the study of the mind has become

increasingly identified with the study of natural cognitive systems and considers some worries associated with this trend. Section 3.5 concludes this chapter with a summary of the entire project.

3.2 Received View

The received view in Cognitive Science has it that the mind is a natural, observer-independent property that emerges out of appropriately organized physical systems (Lycan, 2002). As a result, attributing a mind or a mental state to some collection of physical features amounts to a substantial ontological claim about the world, not a mere terminological choice. Furthermore, evidence overwhelmingly suggests that currently, as a matter of contingent empirical fact, only natural biological organisms have minds. The fact that biological organisms have bodies that sharply delineate them from their environment has been the strongest *prima facie* reason for treating the physical body as an *epistemologically* significant boundary for the purposes of studying the mind. Stipulating, for research purposes, that the boundary of mind lies at the skin and skull has led to considerable progress in understanding the mechanisms underlying cognition. For this reason, the orthodox position in Cognitive Science takes an organism's physical body to be an *ontologically* significant boundary of mind. In other words, skin and skull demarcate a physical system (viz. an organism) which, as a matter of ontological fact, gives rise to mind and cognition.

The above inference may strike some as contentious because it clearly presupposes a form of scientific realism; we rely on our most successful scientific hypotheses to justify ontological claims about the mind. Whether or not science can ever yield insights about theory-

independent reality is certainly a controversial topic. Nonetheless, the debate between the orthodox position and champions of EM does not carry any significant philosophical weight unless both sides assume a realist position.

The EM thesis is supposed to challenge our ontological commitments about the mind; not just our conventional way of talking about it.

Recently philosophers working in EC have issued a serious challenge to the orthodox position. Impressed by the opportunistic nature of cognition a number of thinkers (Clark & Chalmers, 1998; Hutchins 1995a, 1995b) have suggested that the cognition and mind extend beyond the boundary of skin and skull (Clark, 1997., p.69).

> Individual brains should not take all the credit for the flow of thought or the generation of reasoned responses. Brain and world collaborate in ways that are richer and more clearly driven by computational and informational needs than was previously suspected.

They maintain that there are cases where a cognitive agent coupled to its environment constitutes a single, unified cognitive system. Hence, they maintain that cognition and mind ought to be viewed as spanning the brain, body, and world. If one assumes a realist position and takes this perspective to be a guide to matters of objective fact, then the philosophical upshot is that the body ceases to be an epistemologically and, most importantly, ontologically significant boundary of mind: the mind "leaks out" into the world.

3.2.1 Extending the Mind

The principal exposition of the EM thesis comes from Clark and Chalmers (1998) (henceforth, C&C) where they introduce and defend *active externalism* about cognition: the view that cognitive processes are partially *constituted* and *driven* by factors external to the agent. Thus,

C&C claim that there are instances where a coupled system consisting of an agent and some relevant features of the environment "can be seen as a cognitive system in its own right" (Clark & Chalmers 1998., p.8) From the outset I submit that I do not intend to dispute that cognition is *driven* by external factors. This, as I understand it, means that cognitive development and behaviour are to a great extent causally dependent on the environment. That much is undeniable. What I do intend to dispute, however, is that external processes are *constitutive* of cognition – the claim that challenges the traditionally accepted boundaries of mind.

C&C isolate two conditions that an external process has to meet for it to be constitutive of an agent's cognitive performance. First, it has to be appropriately coupled to a cognitive agent; and second, it has to pass the application of the "parity principle" (Clark & Chalmers 1998, p.8):

> If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. Cognitive processes ain't (all) in the head!

The Kirsh and Magilo (1994) study of epistemic actions is a case in point. Success in Tetris depends on the ability to quickly assess best fit of an incoming piece with the existing structure. Players reliably coupled to game controls (i.e., buttons) are faced with a choice of performing this process *internally* (i.e., using mental rotation and imagery to see how pieces fit) or *externally* (i.e., pressing a button to rotate the piece on screen). C&C rightfully point out that if the process were performed internally it would certainly count as a part of an overarching cognitive performance. They proceed to apply the parity principle and conclude that the external process ought to count as such as well.

This conclusion is pivotal as it leads EM proponents to make further claims about mental states.

Since mental states are instantiated by cognitive processes and cognitive processes can be extended into the world it follows that *mental* states can be extended into the world as well.

> "In particular, we will argue that *beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes. If so, the mind extends into the world" (Clark & Chalmers, 1998, p. 12)

To illustrate this concretely C&C present the case of Inga and Otto. Inga learns that there is an exhibition at the Modern Museum of Art (MMoA) and decides to attend. With little effort she recalls that the museum is located on 53rd Street and sets off to see the exhibit. It is natural to say that Inga wanted to see the exhibit, consulted her memory, and then acted upon her belief about the museum's location. Contrast this (ordinary) case with that of Otto who suffers from Alzheimer's disease. Otto's condition forces him to carry around a notebook everywhere he goes so he can write down and retrieve any information he deems relevant to a task at hand. Being a museum lover himself Otto hears about the exhibit and decided to attend. He checks his notebook, finds the address of MMoA and sets off for the museum. C&C are driving at the intuition that the notebook plays essentially the same role for Otto as biological memory does for Inga. Hence, since Inga's biological memory is constitutive of her cognitive performances in day to day life, so is the notebook for Otto. Therefore, we ought to treat the notebook as holding Otto's *beliefs* and the interaction between Otto and the notebook as an instance of *remembering*.

Let us keep in mind that the application of the parity principle is limited to cases where

external processes are *coupled* to a cognitive agent. As a result, the parity principle does not render Otto's notebook alone cognitive. But once it is appropriately coupled to him the entire Otto-plus-notebook system is deemed cognitive. To be clear on the significance of the coupling criterion it is instructive to attend to a particularly misleading conception of its role in EM arguments. Adams and Aizawa (2001) heavily criticized C&C's appeal to coupling as an instance of a "coupling-constitution fallacy":

> Yet, from the fact that cognitive brain processes are *coupled* to environmental processes we cannot simply infer that the environmental processes *constitute* part of the cognitive process. To think otherwise is to commit the coupling-constitution fallacy.

In reply to this criticism Clark (forthcoming, pp.4-5) clarifies the "precise role of the appeal to coupling" in EM arguments:

> The appeal to coupling is not intended to make any external object 'cognitive' (insofar as this notion is even intelligible). Rather, it is intended to make some object, that in and of itself is not usefully (perhaps even intelligibly) thought of as *either cognitive or non-cognitive*, into a *proper part of some cognitive routine.*

Hence, it is misleading to think of coupling as making an external processes cognitive *per se*. Coupling is meant to be a way of *integrating* an external part/process with the agent, such that the whole can support intelligent behaviour. The real difficulty then, according to Clark (forthcoming), is to spell out what kind of coupling makes an external process a proper part of the cognitive system, not just causally related to it, without arbitrarily treating the skin as a special boundary.

A view similar to that of C&C has also been defended in Hutchins (1995a, 1995b). The author presents an in-depth analysis of how various cognitive tasks are performed "in the wild"; i.e., outside a laboratory setting. He notes that activities such as ship navigation or

pilotage are best understood as being performed by a system comprised of individual(s) together with appropriate tools. Further Hutchins proposes that such a system could be construed as performing a cognitive activity:

> "But I do believe that the computation observed in the activity of the larger system can be described in the way cognition has been traditionally described – that is, as computation realized through the creation, transformation, and propagation of representational states." (Hutchins 1995, p.49)

Thus, various external tools which supply us with data and help perform difficult computations are interpreted as creating, transforming and propagating representational states to be used by other components of the system. Hence, Hutchins' view that cognition is a property of a broader system is justified by appealing to the standard explanatory framework of cognitive science.

3.2.2 Active Externalism

According to C&C for an external feature to be constitutive of an agent's cognitive and mental processes it has to simultaneously fulfill two conditions: first, be appropriately coupled to an agent; and, second, pass the application of the parity principle. In this section I critically evaluate both conditions and show them to be inadequate. I first examine the coupling criterion and argue that it only supports the uncontentious claim of active externalism – cognition is driven by external processes. However, it does not establish that external processes are constitutive of cognition. Subsequently, I focus on application of the parity principle in C&C's thought experiment and argue that it relies on an especially crude conception of functional identity. I conclude this section by showing that EM leads to a gratuitous proliferation of

cognitive systems, thereby making it impossible to determine the ontological boundaries of mind. This, I argue, constitutes a significant obstacle to the search for mechanisms underlying cognition and, therefore, renders EM of limited value to Cognitive Science.

Strictly speaking, coupling between two systems, events, or processes only entails that there is mutual causal interaction between them. Hence, since all physical process are causally connected, then, to varying degrees, everything is coupled to everything else. That, however, does not obliterate epistemological and ontological distinctions between systems that have characteristically distinct properties. This line of reasoning underlies an aforementioned criticism of C&C's coupling criterion advanced by Adams and Aizawa (2001). They argue that it is simply fallacious to conclude from a mere fact that some process *E* is coupled to process *C* that *C* extends into *E*. They provide the following concrete example (Ibid., p.56):

> The kidney filters impurities from the blood. In addition, this filtration is causally influenced by the heart's pumping of the blood, the size of the blood vessels... The fact that these various parts of the circulatory system causally interact with the process of filtration in the kidneys does not make even a prima facie case for the view that filtration occurs throughout the circulatory system, rather than in the kidney alone.

Clark (forthcoming) wrote off the criticism as a case of misunderstanding, and clarified that coupling is not meant to render an external feature cognitive *per se* but only to integrate an external feature with a cognitive agent. It is unclear how this response is meant to dodge the issue raised by Adams and Aizawa (2001). First of all, to say that an external process is 'integrated' with a cognitive agent is, at best, awfully vague; at worst, it is equivalent to the very claim being evaluated, namely that an external process and the agent constitute a unified cognitive system. Thus, Clark seems to merely rephrase the original claim without addressing

the issue. Second of all, in the example above it is evident that mere coupling does not establish that filtration occurs throughout the circulatory system. Likewise coupling alone does not give us any reason to think that the circulatory system is constitutive of filtration. It only shows that kidneys, the system responsible for filtration, are *causally connected* to other physiological processes. Generally, for C&C an appeal to coupling can only justify the uncontentious claim of active externalism: cognition is driven by external processes. Coupling neither establishes that an external process is cognitive nor that it is constitutive of an agent's cognitive performance.

The parity principle is essentially a criterion which allows us to pick out external processes that could be constitutively integrated with a cognitive agent once the coupling condition is met. As long as a process "functions as a process" that would be deemed cognitive[16] if it was performed by the brain, then it ought to be considered as such even if it is external to the agent. In case of Inga and Otto successful application of the parity principles hinges on functional identity of the notebook's content and information stored in biological memory. The question is, then, are the two functionally identical or, at least, similar enough? C&C hold that the role of information in the notebook is on par with that in biological memory insofar as both supply the agent with the information required to inform and guide his actions whenever needed. On this reading, the functional identity consists solely in the capacity to yield and retain required information. But once we consider the details of how information in biological memory functions we can see that "functional identity" of Otto's notebook and Inga's memory is an overly crude abstraction.

The dominant view in psychology treats human memory as comprised of distinct, but

16 Or be considered constitutive of an overarching cognitive process.

interactive, systems (Squire, 2004). Memory is broadly divided into declarative and non-declarative. Declarative memory "refers to the capacity for conscious recollection about facts and events" (Ibid., p.173). Nondeclarative memory on the other hand "is dispositional and is expressed through performance rather than recollection" (Ibid., p.173). Otto uses his notebook to consciously record and retrieve factual information such as a museum address. Hence, the notebook most closely corresponds to the declarative kind of memory albeit disconnected from other memory systems. In contrast, memory systems in the brain interact and support other cognitive functions unconsciously. For instance, Squire (2004, p. 174) states:

> The memory systems of the brain operate in parallel to support behaviour. For example, an aversive childhood event involving being knocked down by a large dog can lead to a stable declarative memory for the event itself as well as a long-lasting nondeclarative fear of dogs (a phobia) that is experienced as a personality trait rather than as a memory.

Since Otto's notebook is not connected to other memory systems, information contained in it cannot systematically and unconsciously lead Otto to display behaviour such as phobic aversion to dogs.[17] Therefore, Otto's notebook and Inga's biological memory are, quite clearly, not functionally identical.

Note that there are, undoubtedly, many other functional differences which I have not considered. Hence, the claim that the notebook and biological memory are not functionally identical does not rest upon the single example I have provided here; many more are readily available. This, I believe, puts C&C in an uncomfortable position: either concede that Otto's

17 C&C may object that Inga's memory can suffer from similar deficiencies (due to trauma perhaps) but we would not be inclined to think that it ceases to be a part of Inga's mind. Thus, they would insist, the fact that Otto's notebook does not exhibit certain characteristics exhibited by normal memory is insufficient to establish that it is not a part of Otto's mind. Nonetheless, since I do not aim to argue against EM *per se* the criticism is misplaced. My sole aim is to demonstrate that C&C have failed to show that Otto's notebook is functionally identical to Inga's memory and, thus, have failed to make a convincing case for EM.

notebook is not functionally on par with Inga's biological memory and, therefore, block the inference to EM; or insist that these differences are irrelevant and embrace a much looser conception of functional identity. Clearly the second option is the only live one for C&C, but it is also highly problematic. It leads to a gratuitous proliferation of cognitive systems and makes it impossible to draw a principled line demarcating systems that have minds from the external environment.

The central problem with the way C&C must construe functional identity is that it is excessively liberal. As a result there is a staggering number of external resources, broadly construed, that can be thought of as cognitive aids and be integrated with a cognitive agent.[18,19] Whether they are, in fact, integrated with an agent depends on the nature of coupling between the two. Precisely what makes some form of coupling appropriate depends on a specific case. For instance, in C&C's example the notebook is meant to be a *long-term substitute* for biological memory. Therefore, the notebook has to be coupled to Otto in a way that makes it as easily and reliably accessible as we would expect biological memory to be viz. practically inseparable from Otto at all times. However, if we are interested in extended cognition in Tetris (usually a short-lived activity), then the coupling requirements are substantially less strict. A player together with the computer constitute an extended cognitive system just for the duration of the game. Once the game is over and the player goes on to, say, organize her daily schedule a new extended cognitive system is formed. At this point it is comprised of the same agent and her preferred organizational tools such as a laptop computer, post-it notes, and so on. More

18 Seeing that notebooks, post it notes, and calculators all pass an application of the parity principle we can imagine a plethora of other devices and ways in which we can modify the environment to assist cognitive tasks.

19 A proponent of EM could try and circumvent this problem by restricting the kinds of resources that would count as cognitive extensions. Thus, cognition would be extended in some cases (e.g., when using external storage media), but not others (e.g., when one is playing Tetris). This, however, would be entirely *ad hoc*.

generally, throughout the day a number of extended cognitive systems are formed, persist and dissipate depending on the agent's needs. Consequently, the boundaries of mind are imprecise and seem to be constantly changing, extending more or less into the world depending on the cognitive task and the nature of external resources utilized. This is highly undesirable from the standpoint of scientific practice. As it was argued in Chapter 2, the strategies of decomposition and localization provide the best method for comprehending complex systems (e.g., cognitive systems). However, these strategies are only applicable when there is a sharp distinction between the system under analysis and its environment. Therefore, shifting and blurring the boundaries of mind impedes our ability to explain it in mechanistic terms.

To summarize, I have argued that coupling between an agent and external resources is only a suitable justification for the uncontentious claim of active externalism. However, when it comes to showing that external features and an agent constitute a cognitive system an appeal to coupling is fallacious. In addition to that, I have shown that C&C's parity principle relies on an untenable conception of functional identity and, for that reason, entails an excessive proliferation of cognitive systems. What's more, this state of affairs makes the boundaries of mind unacceptably *ad hoc*, leaving them free to extend more or less into the world depending on the nature of cognitive task and external resources recruited.

3.2.3 Hutchins' Argument

Hutchins defends EM on the grounds that the explanatory framework of cognitive science is well-suited for understanding cases of allegedly extended cognition. However, the virtues associated with looking at these phenomena from the representational/computational

perspective are insufficient to establish EM. For the argument to work Hutchins requires the following premise:

(P) IF *X* can be studied using CR framework, THEN *X* is cognitive.

In other words mere applicability of CR framework for studying some *X* is supposed to be sufficient to render *X* cognitive. We can already see that P is unacceptably broad because it renders any information-processing system (e.g., desktop computer, digital stereo) cognitive. For this reason, Cognitive Science is only justifiably committed to a much weaker claim that:

(P*) IF *X* is cognitive, THEN *X* can be studied using CR framework.

As a result Hutchins' argument either assumes a false premise P or commits the fallacy of affirming the consequent with respect to P*. In either case the argument does not entail EM.

Hutchins can, of course, bite the bullet and argue that success of CR perspective in Cognitive Science warrants one to endorse P rather than P*. This response, however, is ultimately untenable as it makes Hutchins susceptible to the same criticism I have directed at C&C; it leads to an unacceptable proliferation of cognitive systems and makes it impossible to determine the boundaries of mind. For instance, consider Hutchins (1995a) and the analysis of how a "cockpit remembers its speed." In this case the cockpit together with the pilots, tools and crew is the cognitive unit of analysis and we are interested in how the system remembers its speed and adjusts wing position to accommodate a safe landing. But in the same manner we could concern ourselves with, say, coordination of air traffic to ensure safe landing and departure of multiple planes over the course of the day. In that case, if we follow Hutchins' suggestion, pilots, air traffic controller along with the multitude of staff and instruments would be the proper unit of analysis and a cognitive system in its own right. This system, in turn,

participates in a larger context of activity which is likely to succumb to computational/representational analysis as well. Thus, we have a number of purportedly cognitive systems that start at the level of an individual pilot and grow to include the rest of the crew, air traffic controller and so forth. In other words, one can stipulate what the system boundaries are depending on the target of explanation and relevant knowledge. The difficult question for Hutchins, then, is: Which of these various *epistemological* boundaries most closely correspond to the *ontological* boundaries of mind? At one extreme lies a sole cognitive agent and at other - the most expansive system that yields to a computational/representational analysis. The former extreme represents the orthodox view regarding the boundaries of mind and so is not a live option for EM proponents. But, I submit, once we move *beyond* the bodily boundaries there appear to be no considerations that would allow us to draw a principled boundary demarcating the environment from the physical system that gives rise to mind and cognition.

3.2.4 Assigning Credit

It is definitive of EM to attribute the performance of a task to an agent-world system. In some cases such attributions arouse little suspicion; there seems nothing wrong with saying that Lou and his calculator calculated the right answer to a math problem. To add to that, the claim that a calculator was performing, well, calculations is entirely intuitive and appropriate. In contrast, the case of Inga an Otto has shown that it is highly *counterintuitive* to speak of an inanimate object holding beliefs or to speak of agent-tool interaction as an instance of remembering. C&C admit they may be vulnerable to a "common-sense" criticism that they are abusing or

stretching the conventional meaning 'belief'. But they claim that the criticism is not fatal. It is reasonable to expect that some *mental* concepts may have to be reconsidered in response to a change in our ontological commitments regarding the mind. If the mind really does leak out into the world, then the burden is on us to adjust our concepts to the newfound reality.

I submit, however, that there are a number of instances where crediting an agent-world system with performance of some task is not only counterintuitive but appears to violate grammatical norms of the English language. For instance, let's consider how C&C's reasoning works in the following, now familiar, example. Consider two simple robots: Sven and Herbert. Sven is an old prototype based on the sense-model-plan-act architecture. Sven relies on a highly detailed internal map of the surroundings to guide intelligent behaviour such as obstacle avoidance. Thus, Sven's motor capacities, reliably coupled to an extensive internal database, produce intelligent behaviour. Herbert, on the other hand, is a more recent model based on Brooks' subsumption architecture discussed in Section 1.5. Herbert relies on close and highly reliable coupling with the world to supply him with information about his surroundings and guide movement. Hence, the internal database plays essentially the same functional role for Sven as the external world does for Herbert – both reliably supply information needed to guide goal-oriented, intelligent action. Therefore, we ought to credit Herbert *and the world* with producing intelligent behaviour; Herbert's mind includes the world. But since Herbert's movement occurs *with respect to* this same world it follows that Herbert, *literally*, moves through his mind! While it is perfectly legitimate to speak of moving through one's mind in a figurative sense[20] the literal interpretation does not make sense. Hence, to speak of the world as constituting Herbert's movement is not to just stretch the conventional meaning of 'movement'

20 As in exploring your mental space, searching for ideas or something along those lines.

it is tantamount to violating semantic norms of the English language.

3.2.5 Summary

Thus far I have argued that C&C and Hutchins have not provided sufficient reason for us to believe that the mind extends beyond the bodily boundary. I have critically evaluated EM and argued that C&C only succeed in establishing the uncontentious claim that external processes drive cognitive behaviour. When it comes to showing that external features are constitutive of cognition, C&C's arguments are either weak or suggest an undue proliferation of cognitive systems. I have also looked at Hutchins' argument for EM and argued that it requires an additional premise that we have a strong reason need not grant. However, I have considered the implications of granting this contentious premise and demonstrated that it leaves Hutchins' position vulnerable to the same criticism I directed at C&C: it leads to an undue proliferation of cognitive systems without any principled way of determining the boundaries of mind. In the next section I intend to situate the problem of system boundaries within a mechanistic framework discussed in Chapter 2 and hope to show that it favours traditional boundaries of mind.

3.3 Constitution and the Locus of Control

In Section 3.2.2 I have argued that coupling between an agent and environmental features demonstrates a causal, rather than constitutive, relation between them. However, as a partial concession to C&C, I want to note that the system traditionally credited with producing cognition (i.e., brain) consists of closely coupled sub-systems whose collective behaviour is

responsible for the emergence of mind. Hence, the argument presented in 3.2.2, although technically correct, suggests the need to address the following questions: (1) What does it mean for some *X* to be constitutive of *Y*? and (2) What would it take to show that some *X* is constitutive of *Y*? The most obvious and intuitive answer to the first question is that *X* has to, in some sense, be *necessary* for Y to perform its function(s)[21]. Unfortunately, this won't do because it still fails to distinguish between constitutive and merely causally relevant factors. For instance, proper heart functioning and suitable distance from the Sun are both causally necessary for me to exercise my cognitive abilities here on Earth. Were my heart to suddenly stop or if the Earth's orbit were to change drastically my cognitive competence would undoubtedly suffer. Surely though, these factors are not constitutive of my cognitive abilities. There is a strong intuition that, at least when we are dealing with natural phenomena, constitutive elements belong together independently of our epistemological status. In the present context this means that there is an objective fact about which parts of the world make up a system capable of producing a mind. Seeing that science is our best, but by no means a perfect, guide to matters of objective fact, I propose that we situate the two questions above within the mechanistic framework introduced in Chapter 2. In the next section I briefly review the mechanistic perspective and suggest, as an answer to (1), that we identify elements constitutive of cognition with the locus of control over cognitive activity. Afterwards, as a way of addressing (2), I present empirical reasons for treating nothing finer-grained than an agent itself as the locus of control over cognitive behaviour.

21 In fact, C&C seem to hold this kind of view. They claim that external features which assist us in cognitive tasks are on par with cognitive mechanisms in our head because they have a significant effect on cognitive performance: “If we remove the external component the system's behavioural competence will drop, just as it would if we removed part of its brain.” (Clark & Chalmers, 1998, pp. 8-9)

3.3.1 System Boundaries: A Mechanistic View

In Chapter 2 I introduced mechanistic explanation (ME) and defended its applicability to cases of embodied cognition. Recall that a mechanistic explanation (ME) attempts "...to account for the behaviour of a system in terms of the functions performed by its parts and the interactions between those parts" (Bechtel& Richardson, 1993., p.17). Development of a mechanistic explanation proceeds in three stages: identifying the system which is the locus of control for the phenomenon to be explained; decomposing the system into component operations which collectively produce the phenomenon; and localizing the parts within the system that perform those operations. The first stage, identifying the locus of control, is tantamount to delineating the system credited for producing the phenomenon from the environment (Bechtel & Richardson, 1993, p.39):

> The scientist must segment the system from its context and identify the relevant functions assigned to it. To substantiate the assignment of a function to a system, the scientist generally must offer theoretical or empirical arguments showing that the physically and functionally independent system identified has substantial internal control over the effect. This is what we describe as treating the system as the locus of control for a phenomenon.

On a mechanistic perspective, the distinction between the system (i.e., a collection of constitutive parts) and the environment is based on the degree of internal control over the behaviour under consideration.

Note that there is no requirement that the system has to perform its functions in complete isolation from the rest of the world. As a result, this perspective is consistent with the fact that there are numerous external contributions that drive the behaviour of the system and allow it to perform its functions. I submit that this perspective is well-suited for addressing the two

questions posed in the previous section: (1) What does it mean for some *X* to be constitutive of *Y*? and (2) What would it take to show that some *X* is constitutive of *Y*? Thus, I suggest that, as an answer to (1), we identify elements constitutive of cognition with the locus of control over cognitive activity. The answer to (2) follows naturally; we have to offer arguments showing that the system deemed to be the locus of control "has substantial internal control over the effect" (Ibid., p.39). In the next section I consider the nature of allegedly extended cognitive processes and argue that they have to be *sensorimotor* in nature. Subsequently I present empirical reasons for taking a biological agent to be the locus of control over sensorimotor activity and, therefore, the locus of control over the cognitive activity that depends on these interactions.

3.3.2 A World of Possibilities

The central motivation behind EM is the fact that cognitive agents exploit information storage and processing capacities afforded by the environment. For these capacities to be constitutive of cognition, C&C argue, they have to be appropriately coupled to a cognitive agent. In other words, the manner in which an agent interacts with and detects changes in the environment has to support intelligent behaviour. Since biological agents interface with their surroundings solely by way of sensory receptors and motor actions, possible interactions with the environment are necessarily sensorimotor in nature. Therefore, agent-environment coupling is constituted by the stimulation of an agent's sensory receptors and physical influences an agent exerts upon the environment. I here argue that the nature of these interactions renders an agent the locus of control over cognitive behaviour.

Purposeful interactions with the world that further an agent's cognitive goals necessitate knowledge of what the world is like and various labour-saving opportunities it affords. Biological agents gather information about their surroundings by way of sensory receptors that have evolved to respond to various kinds of stimuli. A human eye, for example, contains millions of light sensitive cells that convert photon energy into electrical signals which are propagated to the brain via the optic nerve. Subsequently, the brain uses these signals to deduce what the world is like, which allows an agent to engage in purposeful interactions with it. This process is complicated by the fact that sensory receptors are continuously bombarded by stimuli that are inherently noisy (i.e., contain disturbances that distort information conveyed by the original signal) and ambiguous (Rao et. al., 2002., Knill & Pouget, 2004., Faisal et. al., 2008). To put it differently, a specific pattern of sensory stimulation contains a lot of irrelevant information and can be consistent with a number of ways the world can be like. How does the brain get around these difficulties?

A promising hypothesis currently explored by a number neuroscientists is that the brain relies on probabilistic inferences in order to deal with uncertainty. In particular, researchers have found the Bayesian statistical framework to be useful for understanding how the brain learns from experience and uses this knowledge to inform sensory and motor functions (Chater, 2006; Eliasmith & Anderson, 2003; Eliasmith, 2007; Knill & Pouget, 2004; Kording & Wolpert, 2006; Rao et. al., 2002). At the core of this framework lies Baye's theorem which provides a mathematically rigorous way of estimating the probability of some hypothesis based on prior knowledge and currently available evidence. In vision, for example, the brain has to estimate the most probable state of the world based on the pattern of sensory stimulation and

prior knowledge concerning the environment and possible causes of stimuli. Hence, in order to construct the most probable representation of the world the brain must data mine a constant stream of noisy and ambiguous data to extract information consistent with previous learning and goals (Friston, 2003). This effectively makes sensory organs and the brain natural filters of information continuously provided by the environment. The most crucial fact about this process is that its outcome – a model of the external world that guides further interactions – is determined by factors *internal to the agent* such as physiology, previous learning experience and current goals[22].

Selecting an appropriate action based on results of perceptual inference is similarly characterized by uncertainty and close dependence on goals and prior knowledge. For this reason Bayesian methods have been successfully applied to modelling motor control (Kording & Wolpert, 2006). These models assume that selecting a specific movement "...can be described as the rational choice of the movement that maximizes utility according to decision theory" (Ibid., p. 319). Thus, each action carries with it a cost (e.g., time, energy, foregoing other actions etc.) and a reward (e.g., escaping from a predator, avoiding cognitive work, getting praise from a peer etc.) and the brain has to infer the action which has the highest probability of maximizing the reward. Utility and disutility associated with specific actions are determined by subjective preferences of individual agents. The upshot is that motor actions are akin to decisions made by individual agents in accordance with internally determined constraints.

Traditional examples of allegedly extended cognition severely downplay the extent of

22. This echoes the point discussed in Section 1.7 that the way organisms represent the world depends on their form of physical embodiment and their history of interactions with the world.

agents' internal control over environmental interactions and overemphasize external contributions. This, in effect, blurs the distinction between the cognitive system (i.e., biological agent) responsible for producing intelligent behaviour and its environment. Thus far I have shown that the outcome of sensory processing and subsequent motor actions are largely determined by factors internal to the agent, making it the locus of control over sensorimotor activity. I believe this perspective drastically changes the way we should look at examples of allegedly extended cognition. Consider, once again, the way Otto uses his notebook as a long-term substitute for memory. What is crucial about this case is that Otto's interactions with the notebook are dictated and made meaningful by factors internal to him, while the notebook plays essentially no active role. Put differently, the decision to use the notebook is driven by Otto's internal goals, and the benefits of so doing accrue only to him. The notebook itself does not play an active role in the allegedly extended cognitive process, and has little to no internal control over how cognitive behaviour will unfold. Any effect that the notebook can have on Otto's cognitive behaviour will be in virtue of information contained within it. However, processes which interpret that information, determine its relevance to a given goal, and ultimately use it to guide motor action all reside within the confines of Otto's body. Hence, despite the fact that Otto's behaviour depends on close cooperation with the notebook its propensity to drive and contribute to cognition depends almost entirely on factors internal to Otto. In light of these considerations, I submit, it is a lot more natural to treat the notebook as a peripheral device causally relevant to cognition rather than a constitutive element thereof.

In this section I have focused on the nature of sensorimotor interactions which underlie allegedly extended cognition. I have discussed a prominent hypothesis within neuroscience that

brains rely on Bayesian statistical inferences in order to deal with uncertainty, which permeates all sensory and motor tasks. The Bayesian perspective, I have argued, implies that both the choice of motor actions and the outcome of sensory processing are determined by factors internal to the agent. Therefore, a biological agent is the locus of control over sensorimotor activity in the world and, ergo, the locus of control over cognitive activity that depends on these interactions.

3.4 Research Directions: Then and Now

Relative to other well-established fields such as chemistry or physics Cognitive Science is still in its infancy. A particularly notable trend in its development is the continuously growing emphasis on understanding natural cognitive systems. In this section I want to explore the origins of this trend and examine the challenges and opportunities it presents. The success and rapid development of Cognitive Science are rooted in a number of intellectual advances of the 20^{th} century. Chief among them is the development of the formal notion of computation and the advent of digital computers. Formal computational methods afforded great power and generality, and provided researchers with a robust framework for modelling cognitive functions. Thinking of cognition as an abstract computation encouraged the idea that mind can be understood in abstraction from implementational details. In fact, this was seen as a distinct virtue of the computational perspective. This outlook represented an extreme functionalist position, where cognitive functions were analyzed independently of their evolutionary and immediate physical context. Initial optimism spurred a torrent of research into building models of advanced cognitive abilities such as formal reasoning, chess playing, and diagnosing illness.

These largely symbolicist models were reasonably successful, which encouraged researchers to believe that other cognitive functions would succumb to the same approach.

However, limitations of the symbolicist approach, and a resurgence of interest in parallel and distributed computation in the late early 80's, paved the way for connectionism. Although artificial neural network models have been around before the 1970's interest in them has diminished after it was shown that their computational powers were limited. In the early 80's learning algorithms were proposed to deal with this shortcoming thereby rekindling an interest in connectionist research (Knight, 1990). Connectionist models, although highly idealized, were explicitly brain-like and were naturally suited for problems that required multiple constraint satisfaction. Thus, connectionist research constituted a move towards greater biological plausibility, and relied on a closer relationship between function and implementation. This trend continued when ideas associated with DST and EC came into prominence. Cognitive scientists started to think about cognition as a naturally evolved, situated activity and began to emphasize practical constraints facing cognitive systems. The most notable one is that cognition operates in a context of limited time and limited resources. Hence, nature places a premium on cognitive strategies that are time and energy efficient – characteristics seldom possessed by traditional cognitive models. But if cognition has to adequately cope with the aforementioned constraints, then the form of implementation will be of paramount importance. The upshot is that the functionalist outlook of the 1950's was clearly overstated; it is difficult to segregate function from implementation when we are dealing with real-world constraints (Eliasmith, 2002). Cognitive Science started off with a view of cognition as an abstract, atemporal process that can be understood independently of implementational

details. This represented a highly idealized position which over the past 50 years has been gradually weakened. With the advent of powerful brain-imaging tools during the period of 1950's-1980's, and a growing emphasis on practical constraints facing cognitive systems, the study of the mind has become increasingly identified with the study of natural cognitive systems. This, in my opinion, represents a positive development. Some, however, would object to this trend as distinctly unphilosophical because it reflects some form of "mental chauvinism," where the mind is identified with the brain. Critics would insist that cognition has to be multiply-realizable and that we should not tie a general theory of mind to one form of implementation. These worries, I suspect, are too premature. If we want to understand what it is to have a mind we ought to begin with the study of paradigm cases. As a matter of contingent empirical fact (setting philosophical worries about other minds aside) evidence overwhelmingly suggests that only natural biological organisms have minds. These systems are enormously complicated and explaining how they work constitutes a formidable challenge that we are only beginning to address. Until we have an extensive appreciation of how natural cognitive systems exhibit robust, adaptive behaviour there is little hope for settling the question of whether other, perhaps non-biological, systems can have minds as well.

3.5 Conclusion

The goal of this project was to distil the foundational ideas of EC research and evaluate their impact on CR and Philosophy of Cognitive Science. In the first chapter I have summarized the core tenets of CR and clarified the meaning and significance of the term 'embodiment' in EC literature. In so doing, I have tried to highlight the most salient discrepancies between EC and

CR and ultimately argued that the two are compatible. Nonetheless, I have shown that EC suggests an alternative research domain and a rethinking of the role and complexity of computations and representations in cognitive activity. In Chapter 2 I have defend the value and applicability of mechanistic explanation to cases of embodied and embedded cognition and demonstrated that a proposed alternative, Dynamic Systems Theory, is not a viable substitute. Lastly, in Chapter 3 I have critically evaluated EM and argued that there are no compelling reasons to accept it. In addition, I have provided empirical considerations that support the received view that the mind does not extend beyond the confines of the body.

References:

Adams, F., & Aizawa, K., (2001) "The Bounds of Cognition" *Philosophical Psychology*, **14**, (pp.43–64)

Anderson, M.L., (2003) "Embodied Cognition: A Field Guide" *Artificial Intelligence,* **149**, (pp. 91–130)

Anderson, M.L., (2007) "How to Study the mind: An Introduction to Embodied Cognition" In Santoianni, F., & Sabatano, C., (Eds.), Brain Development in Learning Environments: Embodied and Perceptual Advancements, (pp. 65–82). Cambridge: Cambridge Scholars Publishing

Ballard, D.H., (1991) "Animate Vision" *Artificial Intelligence*, **48**, (pp.57–86)

Barsalou, L., Niedenthal, P., Barbey, A., & Ruppert, J.,(2003a) "Social embodiment" In B. Ross (Ed.), *The Psychology of Learning and Motivation*, **43**, (pp.43–92)

Barsalou, L., Simmons, W. K., Barbey, A. K., & Wilson, C. D., (2003b). "Grounding conceptual knowledge in modality-specific systems" *Trends in Cognitive Sciences*, **7**, (pp.84–91)

Barsalou, L., (2008) "Grounded Cognition" *Annual Review of Psychology*, **59**, (pp.617–645)

Bechtel, W & Richardson, R.C., (1993) Discovering complexity: Decomposition and localization as strategies in scientific research. New Jersey: Princeton University Press.

Bechtel, W., (1997) "Dynamics and decomposition: Are they compatible?" *Proceedings of the Australasian Cognitive Science Society*. Retrieved from: http://mechanism.ucsd.edu/~bill/research/dynamics.htm

Bechtel, W & Abrahamsen, A., (2005) "Explanation: A Mechanistic Alternative" *Studies in History and Philosophy of Biology and Biomedical Sciences*, **36**, (pp.421–441).

Beer, R.D., (1995) "Computational and Dynamical Languages for Autonomous Agents" In Port, R.F., van Gelder, T., (Eds.),Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge, MA: MIT Press. (pp.121–149)

Boroditsky, L., (2000) "Metaphoric structuring: understanding time through spatial metaphor" *Cognition* **75**, (pp.1–27)

Boroditsky, L & Casasanto, D., (2008) "Time in the mind: Using space to think about time" *Cognition,* **106**, (pp.579–593)

Brooks, R., (1991a) "Intelligence without representation" *Artificial Intelligence*, **47**, (pp.139–159)

Brooks, R., (1991b) "Intelligence without reason" *Proceedings of the 1991 International Joint Conference on Artificial Intelligence*, (pp.569–595)

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006) "Probabilistic models of cognition: Conceptual foundations" *Trends in Cognitive Sciences*, **10**, (pp.287–291)

Chiel, H.J., & Beer, R.D., (1997) "The brain has a body: Adaptive behaviour emerges from interactions of nervous system, body and environment" *Trends in Neurosciences*, **20**, (pp.553–557)

Churchland, P.S., Ramachandran, V.S., & Sejnowski, T.J., (1994) "A Critique of Pure Vision" In C. Koch & J. L. Davis (Eds.), Large-scale neuronal theories of the brain. Cambridge, MA: MIT Press. (pp.23–60)

Clark, A., (1997) Being there: Putting brain, body, and world together again Cambridge, MA: MIT Press

Clark, A & Chalmers, D., (1998) "The Extended Mind" *Analysis*, **56**, (pp.10–23)

Clark, A., (forthcoming) "Coupling, Constitution and the Cognitive Kind: A Reply to Adam and Aizawa" URL: www.era.lib.ed.ac.uk/bitstream/1842/1443/1/coupling2.pdf

Cowart, M., (2004) "Embodied Cognition" *The Internet Encyclopedia of Philosophy*. URL: http://www.iep.utm.edu/e/embodcog.htm

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

Dennett, D., (1991) Consciousness Explained. Boston, MA: Little, Brown and Co

Eliasmith, C., (1996) "The third contender: A critical examination of the dynamicist theory of cognition" *Philosophical Psychology*, **9**, (pp.441–463). Reprinted in P. Thagard (Ed.) (1998) Mind Readings: Introductory Selections in Cognitive Science. MIT Press. (pp.303–333)

Eliasmith, C., (2002) "The myth of the Turing machine: The failings of functionalism and related theses" *Journal of Experimental and Theoretical Artificial Intelligence*, **14**, (pp.1–8)

Eliasmith, C., & Anderson, C., (2003) Neural Engineering: Computation, representation and dynamics in neurobiological systems. Cambridge, MA: MIT Press

Eliasmith, C (2007) "How to build a brain: from function to implementation" *Synthese*, **159**, (pp.373–388)

Eliasmith, C., (forthcoming) "Dynamics, Control, and Cognition" In P. Robbins and M. Aydede (Eds.) Cambridge Handbook of Situated Cognition. Cambridge University Press.

Etzioni, O., (1993) "Intelligence without Robots: A Reply to Brooks" *AI Magazine*, **14**, (pp.7–13)

Faisal A.A., Selen, L.P.J., & Wolpert, D.M. (2008) "Noise in the nervous system" *Nature Reviews Neuroscience*, **9**, (pp.292–303)

Franklin, S (1997) "Autonomous Agents As Embodied AI" *Cybernetics and Systems*, **28**, (pp. 499–520)

Friston, K. (2003) "Learning and inference in the brain" *Neural Networks*, **16**, (pp.1325–1352)

Hutchins, E., (1995a) "How a cockpit remembers its speeds" *Cognitive Science*, **19**, (pp.265–288)

Hutchins, E., (1995b) Cognition in the wild. Cambridge, MA: The MIT Press

Kirsh, D & Magilo, P., (1994) "On Distinguishing Epistemic from Pragmatic Action" *Cognitive Science*, **18**, (pp.513–549)

Knight K., (1990) "Connectionist ideas and algorithms" *Communications of the ACM*, **33**, (pp.59–74)

Knill, D. C. & Pouget, A., (2004) "The Bayesian brain: the role of uncertainty in neural coding and computation" *Trends in Neuroscience*, **27**, (pp.712–719)

Kording, K. P., & Wolpert, D. M., (2006) "Bayesian decision theory in sensorimotor control." *Trends in Cognitive Sciences*, **10**, (pp.319–326)

Kuhn, T.S., (1962) The structure of scientific revolutions. Chicago: Chicago University Press

Lycan, W., (2002) "Materialism" In Nadel, L., (Ed.), Encyclopedia of Cognitive Science, (pp.1019-24)

Fodor, J., (1987) Psychsemantics: The problem of meaning in the philosophy of mind. Cambridge, MA: MIT Press

Franklin, S., (1997) "Autonomous Agents as Embodied AI" *Cybernetics and* Systems, **28**, (pp. 499–520)

Gallese, V. & Lakoff, G., (2005) "The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge" *Cognitive Neuropsychology*, **22**, (pp.455–479)

Haugeland, J., (1981) "Semantic Engines: An Introduction to Mind Design" In Haugeland, J (Ed.), Mind Design. Cambridge, MA: MIT Press. (pp.1–35)

Haugeland, J., (1998). "Mind Embodied and Embedded." In Haugeland, J (Ed.), Having Thought: Essays in the Metaphysics of Mind. Cambridge, MA: Harvard University Press. (pp.207–241)

Hollan, J., Hutchins, E., & Kirsh, D., (2000) "Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research" *ACM Transactions on Human-Computer Interaction*, **7**, (pp.174–196)

Kirsh, D., (1991) "Today the earwig, tomorrow man?" *Artificial Intelligence,* **47***,* (pp.161–184)

Lakoff, G., & Johnson, M. (1999) Philosophy in the flesh: The embodied mind and its challenge to Western thought. New York: Basic Books

Lenat, D. & Feigenbaum, E., (1991) "On the Thresholds of Knowledge" *Artificial Intelligence* **47***,* (pp.185–250)

Marr, D., (1982) Vision. San Francisco: Freeman.

Martin, A. (2007) "The representation of object concepts in the brain" *Annual Review of Psychology*, **58**, (pp.25–45)

Murphy, G. L., (1996) "On metaphoric representation." *Cognition*, **60**, (pp.173–204)

Newell, A., (1980) "Physical Symbol Systems" *Cognitive Science*, **4**, (pp.135–183)

Newell, A., (1982) "The Knowledge Level" *Artificial Intelligence*, **18**, (pp.87–127)

Newell, A., (1990) Unified Theories of Cognition. Cambridge, MA: Harvard University Press.

Noë, A., (2004) Action in Perception. Cambridge, MA: The MIT Press

O'Regan, J.K., (2002) "Solving the real mysteries of visual perception: The world as an outside memory" *Canadian Journal of Psychology*, **46**, (pp.461–488)

Putnam, H., (1975) "The Nature of Mental States" In Putnam, H., (Ed.), Mind, Language and Reality. Cambridge, MA: Cambridge University Press.

Pylyshyn, Z., (1989) "Computing in Cognitive Science" In Posner, M (Ed.), Foundations of Cognitive Science. (pp.49–93). Cambridge MA: MIT Press.

Quick, T., Dautenhahn, K., Nehaniv, C, L., & Roberts, G., (1999) "On Bots and Bacteria: Ontology Independent Embodiment" In Floreano, D., (Ed.), *Proceedings of the Fifth European Conference on Artificial Life,* (pp.339 344). Heidelberg: Springer-Verlag. Extended version cited; retrieved from: www.cs.ucl.ac.uk/staff/t.quick/papers.html

Rao, R. P. N., Olshausen, B. A. & Lewicki, M. S. (Eds.) (2002) Probabilistic Models of the Brain: Perception and Neural Function. Cambridge, MA: MIT Press.

Rensink, R.A., O'Regan, J.K., & Clark, J.J., (1997) "To See or not to see: The Need for attention to perceive changes in scenes" *Psychological Science*, **8**, (pp.368–373)

Rensink, R.A., O'Regan, J.K., and Clark, J.J., (2000) "On the failure to detect changes in scenes across brief interruptions" *Visual Cognition*, **7**, (pp.127145)

Rumelhart, D. E., (1989) "The architecture of mind: A connectionist approach" In Posner, M (Ed.), Foundations of Cognitive Science. Cambridge, MA: MIT Press. (pp.133–161)

Simons, D.J., & Chabris, C.F., (1999) "Gorillas in our midst: sustained inattentional blindness for dynamic events" *Perception*, **28**, (pp.1059–1074)

Simon, H.A., (1969) The Sciences of the Artificial. Cambridge MA: MIT Press.

Smolensky, P.,(1987) "Connectionist AI, Symbolic AI, and The Brain" *Artificial Intelligence Review*, **1**, (pp.95–109)

Song, H., Franklin, & S., Negatu, A., (1996) "SUMPY: A Fuzzy Software Agent" In *Proceedings of the ISCA Conference on Intelligent Systems,* (pp.124–129) Online version cited; retrieved from: http://citeseer.ist.psu.edu/song96sumpy.html

Squire, L. R. (2004) "Memory systems of the brain: A brief history and current perspective" *Neurobiology of Learning & Memory*, **82**, (pp.171–177)

Thagard, P., (2005) Mind: Introduction to Cognitive Science. Cambridge MA: MIT Press.

Thelen, E., (1995) "Time-Scale Dynamics and the Development of an Embodied Cognition" In van Gelder, T., Port, R.F., (Eds.), Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge, MA: MIT Press. (pp.69–101)

Thelen, E., Schoner, G., Scheier, C., & Smith, L.B. (2001) "The Dynamics of Embodiment: A Field Theory of infant perseverative reaching" *Behavioural and Brain Sciences*, **24**, (pp.1–86)

Thompson, E., (2004) "Life and mind: From autopoiesis to neurophenomenology" *Phenomenology and the Cognitive* Sciences, **3**, (pp.381–398)

van Gelder, T. J., & Port, R., (1995a) Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge, MA: MIT Press.

van Gelder, T. J., & Port, R., (1995b) "It's About Time: An Overview of the Dynamical Approach to Cognition" In van Gelder, T., Port, R.F., (Eds.), Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge, MA: MIT Press. (pp.1–43)

van Gelder, T. J., (1995) "What might cognition be if not computation?" *The Journal of Philosophy*, **91**, (pp.345–381)

van Gelder, T. J., (1998) "Dynamical Hypothesis in Cognitive Science" *Behavioural and Brain Sciences,* **21**, (pp.615–628)

van Gelder, T. J., (1999) "Revisiting the Dynamical Hypothesis" Preprint No. 2/99, University of Melbourne, Department of Philosophy. Retrieved from: http://www.philosophy.unimelb.edu.au/tgelder/papers/Brazil.pdf

Varela, F., Thompson, E., & Rosch, E., (1991) The Embodied Mind. Cambridge, MA: MIT Press.

Wheeler, M., (1997) "Cognition's coming home: The reunion of life and mind" In Husbands, P. & Harvey, I., (Eds.) *Fourth European Conference on Artificial Life,* (pp.10–19). Cambridge, MA. MIT Press.

Wheeler, M., (1998) "Explaining the Evolved: Homunculi, Modules and Internal Representation" In Husbands, P., and Meyer, J. A., (Eds.) *Evolutionary Robotics: First European Workshop,EvoRobot98*, (pp.87–107) Berlin: Springer-Verlag

Wicker, B., Keysers, C., Plailly, J., Royet, Jean-Pierre., Gallese, V., & Rizzolatti, G,. (2003) "Both of

Us Disgusted in *My* Insula: The Common Neural Basis of Seeing and Feeling Disgust" *Neuron*, **40**, (pp.655–664)

Wilson, M., (2002) "Six Views of Embodied Cognition" *Psychonomic Bulletin and Review*, **9**, (pp.625–636)

Yuille, A., & Kersten, D. (2006) "Vision as Bayesian inference: analysis by synthesis?" *Trends in Cognitive Sciences*, **10**, (pp.301–308)

Zaera, N., Cliff, D., & Bruten, J. (1996) "(Not) evolving collective behaviours in synthetic fish" In Maes, P., Mataric, M., Meyer, J., Pollack, J., & Wilson, S., (Eds.)From Animals to Animats 4. Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour. Cambridge, MA:MIT Press. (pp.635–644)

Ziemke, T., (2003) "What's This Thing Called Embodiment?" In Alterman, R., Kirsh, D. (Eds.) *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, (pp.1305–1310) Mahwah, NJ: Lawrence Erlbaum Associates