A DEFENCE OF EXTENDED COGNITIVISM

by

MARTIN GODWYN

B.A.(hons.), University of Southampton, 1996
M.Phil., Jesus College, University of Cambridge, 1998

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Philosophy)

THE UNIVERSITY OF BRITISH COLUMBIA

January 2006

ABSTRACT

This dissertation defends extended cognitivism: a recently emerging view in the philosophy of mind and cognitive science that claims that an individual's cognitive processes or states sometimes extend beyond the boundaries of their brain or their skin to include states and processes in the world. I begin the defence of this thesis through a background discussion of several foundational issues in cognitive science: the general character of cognitive behaviour and cognitive processes, as well as the nature and role of representation as it is standardly taken to figure in cognition. I argue in favour of the widely held view that cognition is best characterised as involving information processing, and that carriers of information (i.e., representations) are ineliminable components of the most distinctively human and powerful forms of cognition. Against this background the dissertation argues in stages for successively stronger claims regarding the explanatory role of the external world in cognition. First to be defended is the claim that cognition is often embedded in one's environment. I develop this claim in terms of what I call 'para-information': roughly, information that shapes how we tackle a cognitive task by enabling the extraction of task-relevant information. Proceeding then to the defence of extended cognitivism, I draw most significantly on the work of Andy Clark. In outline, and in general following Clark, it is argued that states and processes occurring beyond the skin of the cognitive agent sometimes play the same explanatory role as internal processes that unquestionably count as cognitive. I develop this claim in two versions of differing strength: firstly, in a general way without commitment to the representational character of extended cognition, and secondly in a specifically representational version with special attention to intentional explanation. Against each of these versions of extended cognitivism are ranged a number of criticisms and objections, many of which stem from the work of Fred Adams and Ken Aizawa. The dissertation examines these objections and rejects each of them in turn.

TABLE OF CONTENTS

Abstract ..... **ii**
Table of Contents ..... **iii**
List of Figures ..... **vi**
Acknowledgements ..... **vii**
Dedication ..... **viii**
Chapter 1 - Introduction **1**
§1.1 – Some History 1
§1.2 – The Strategy 3
§1.3 – Assumptions, Exclusions, and Auxiliary Issues 6
§1.4 – Chapter Summary 10
§1.4.1 Chapter 1 – Introduction *10*
§1.4.2 Chapter 2 – Cognition *11*
§1.4.3 Chapter 3 – Cognitive Representation *11*
§1.4.4 Chapter 4 – Para-information and Embedded Cognition *12*
§1.4.5 Chapter 5 – Extended Cognition *13*
§1.4.6 Chapter 6 – Extended Intentional States *14*
§1.4.7 Chapter 7 – Conclusion *15*
Chapter 2 – Cognition **16**
§2.1 – Getting Clearer on the Questions 16
§2.1.1 The Nature of the Question(s) *18*
§2.1.2 The Relationship Between the Questions *20*
§2.2 – What is Cognitive Behaviour? 23
§2.2.1 Cognitive Behaviour Defined Paradigmatically *23*
§2.2.2 Cognitive Behaviour Defined as 'Intelligent' Behaviour *26*
§2.2.3 Cognitive Behaviour Defined as 'Flexible' and 'Plastic' *28*
§2.2.4 Problems with Flexibility and Plasticity *31*
§2.2.5 Cognitive Behaviour Defined Through Computationalism *37*
§2.3 – Cognition as Involving Information Processing 41
§2.3.1 Information Processing *42*
§2.3.1 Information Processing and its Malcontents *46*
§2.4 – Conclusion 50
Chapter 3 – Cognitive Representation **52**
§3.1 – Representations 53
§3.1.1 Representation and Intentionality *54*
§3.1.2 Truth-Neutral Informational Standing-In-For *55*
§3.1.3 Properties Represent *57*
§3.1.4 Isomorphism and Interpretation *59*
§3.1.5 Misrepresentation or Error *64*
§3.1.6 Preliminary Conclusions *67*
§3.2 – Why We Need Representations 67
§3.2.1 The Dynamicist Critique *68*
§3.2.2 The Critique Deflated *71*
§3.3 – Conclusion 75
Chapter 4 – Para-Information and Embedded Cognition **76**

§4.1 – Para-Information 79
§4.1.1 Explicit vs. Implicit Information *79*
§4.1.2 Internal Para-Information *82*
§4.1.3 External Para-Information *88*
§4.1.4 Para-Information and Representation *90*
§4.1.5 The Explanatory Role of Para-Information *92*
§4.2 – Para-Information and Dispositions 97
§4.2.1 Dispositions In General *100*
§4.2.2 Para-Information and Embedded Dispositions *103*
§4.3 – Conclusion 106
Chapter 5 – Extended Cognition 108
§5.1 – Extended Cognitivism: Clarificatory Remarks 109
§5.1.1 Some More Or Less Closely Related Claims *110*
§5.1.2 Some Possible Misconceptions *114*
§5.1.3 Some Historical Context *117*
§5.1.4 Processes, States, and Systems *118*
§5.2 – Some Examples 123
§5.2.1 Pen and Paper Addition *124*
§5.2.2 Tetris and Scrabble *125*
§5.2.3 Points of Agreement *128*
§5.3 – The Coupling-Constitution Fallacy 130
§5.3.1 A Central Misunderstanding *131*
§5.3.2 Processes and Functional Descriptions *131*
§5.3.3 Two Senses of Cognitive Object and the Object/Process Distinction ... *134*
§5.4 – Intrinsic vs. Derived Intentionality 135
§5.4.1 First Strategy: Denying the Difference *137*
§5.4.2 Second Strategy: Denying that no Intrinsic Content is Involved *138*
§5.4.3 Third Strategy: Denying the Significance of the Distinction *140*
§5.5 – Scientific Kinds and Explanatory Unity 141
§5.5.1 The Diversity of Extended Cognition *143*
§5.5.2 Causal Kinds vs. Functional Kinds *146*
§5.6 – Locus of Control 150
§5.7 – Conclusion 153
Chapter 6 – Extended Intentional States 156
§6.1 – Cognitive Technology 158
§6.2 – In Defence of External Representations 158
§6.3 – Otto's Notebook 163
§6.4 – The Problem of Cognitive Bloat 165
§6.4.1 Some Problems with the Past-Endorsement Criterion *168*
§6.4.2 Ordinary Usage *170*
§6.4.3 Believing as a Causal-Functional Relation *174*
§6.5 – The Otto 2-Step 181
§6.5.1 – The Speeding Ottoman *184*
§6.6 – The Perception Objection 187
§6.7 – Some Revised Conditions for Belief 189
§6.8 – Conclusion 193
Chapter 7 – Conclusion 195
§7.1 – Summary of the Main Arguments 195

§7.1.1 Chapter 1 – Introduction *196*
§7.1.2 Chapter 2 – Cognition *196*
§7.1.3 Chapter 3 – Cognitive Representation *198*
§7.1.4 Chapter 4 – Para-information and Embedded Cognition *199*
§7.1.5 Chapter 5 – Extended Cognition *200*
§7.1.6 Chapter 6 – Extended Intentional States *201*
§7.1.7 Chapter 7 – Conclusion *203*
§7.2 – Some Closing Speculations 203
Bibliography of works cited and consulted 209

LIST OF FIGURES

Figure 3.1: A Simple Stick Figure.. 63

ACKNOWLEDGEMENTS

An undertaking of the scale of a doctoral dissertation is never a solo effort and to thank everyone who has made a contribution to this dissertation (wittingly or not) would probably take a chapter of its own. Obvious thanks must go to the University of British Columbia whose financial support (mostly in the form of University Graduate Fellowships and a Walter C. Koerner Scholarship) was invaluable. Additional financial support was gratefully received from the Department of Philosophy through research and teaching assistantships as well as course teaching opportunities. Additionally, I could not have completed the dissertation without the selfless financial support of both Dania Sheldon and Johanne Bessette.

Intellectual debts are owed first and foremost to my supervisor, Mohan Matthen, whose patient criticism (sweetened with encouragement) has been vital to seeing this project to completion. My deep thanks also extend to the other members of my supervisory committee, Patrick Rysiew and Dominic Lopes, for their ever helpful and exacting criticism. No section is without their influences, although I must, of course, take full responsibility for any remaining errors or inadequacies.

A doctoral dissertation is, or at least can be, an emotionally fraught undertaking. Thanks for mitigating the emotional toll are owed primarily (again) to Dania Sheldon and Johanne Bessette. I cannot hope to repay all the support that they have given me. Thanks also go to a supporting cast of friends including Terry Woolfries, Steve Dark, F.B. and the gang, and many others for helping to keep me (relatively) sane.

I dedicate this to my mother, Amelia Godwyn, and
to the memory of my father, David James Godwyn.
Everything of any value that I have come to know,
I learned from them.

CHAPTER 1 - INTRODUCTION

§1.1 – SOME HISTORY

Let me begin by way of introduction with a little history. A number of years ago, whilst searching around for a dissertation topic, I became struck by the many difficulties that naturalistic accounts of mental representation had encountered. No account seemed quite up to the task of naturalising content. The central difficulty, it seemed to me at the time, was that the very notion of mental representation itself – as an *internal* entity separated from the external world that it was its specific job to be somehow about – was to blame. This led me to wonder whether the whole project of naturalising mental representation might be fundamentally misguided. Perhaps, I even thought at the time, cognition does not involve representations at all. With such thoughts in mind I began examining theories of the mind that are at least agnostic on the existence of mental representations including, especially, dynamic systems theory approaches. The world, dynamic systems theorists suggested, was not at all separated from the mind – the world and the mind were to be seen as a dynamical reciprocally interacting system. The (I soon came to believe, naïve) hope was that the problems that naturalistic approaches to mental representation might be made to disappear through the provision of some powerful arguments that, contrary to what seemed obvious to almost everyone, the mind did not traffic in representations at all.

Then I read Clark and Toribio (1994), Kirsh (1991), and others. I was quickly persuaded that representations really are things that no remotely adequate account of the mind could possibly do without – at least for certain manifestly cognitive behaviours such as hypothetical inference and imagination. Best laid plans and all that. But *something* still struck me as basically right about the dynamicist intuitions. Even if we need representations for many things, perhaps the task of naturalising content might be rendered more amenable if we stop trying to see the mind as some kind of representation-filled box and cognition as a process tucked away from the world somewhere inside our heads. Then I read Andy Clark and David Chalmers (1998). This time I had found something that looked promising, and so, if the arguments presented in this dissertation carry any weight, that promise has borne fruit.

Their central claim was that at least sometimes human cognition[1] literally extends out into the world – that some processes and states of the world beyond our skin are literal parts of our cognitive processing and (thus) literally parts of our mind. The mind, they suggested, sometimes extends out into the world. They called their position *active externalism*, but I preferred to think of it as *extended cognitivism*.[2]

My interest in this idea deepened with a visit to England in 2001 to a conference held at the University of Hertfordshire called 'The Extended Mind: The Very Idea'.[3] The conference focussed on developing themes from Clark and Chalmers (1998) and included a keynote paper by Andy Clark defending his position against some recent criticisms,

[1] In order to avoid needless prolixity, and unless otherwise specified, 'cognition' shall refer to *human* cognition. This is for stylistic purposes only and is without prejudice against those who include non-human cognition within the boundaries of cognitive science.

[2] Terminology varies amongst both its proponents and its critics, and perhaps none seems *entirely* adequate. Dennett does not label the position, but generally uses the descriptive terms such as 'off-loading' our cognitive processing into the environment. As noted, Clark and Chalmers, prefer the term 'active externalism'. But to some extent this may risk drawing a closer association with other famous 'externalisms' (most especially semantic externalism) than is warranted. Robert Wilson's (1994) term, 'wide computationalism', runs similar risks and seems too specific to capture the more general – not necessarily computational – thesis. Neither is it especially clear (at least to this author) why Clark and Chalmers should prefer 'active' as a distinctive feature of externalised cognition. One consideration in favour of preferring the use of 'extended' over 'external' is that it more adequately reflects the *continuity* of external cognitive states with internal cognitive states and processes. Another is that 'externalism' has very much been appropriated to discussions about representational content. Rowlands (1999) prefers the term 'environmentalism' – but that, too, conjures up a wealth of largely extraneous connotations. By the same token, of course, the term 'cognitivism' may perhaps conjure up a number of further associations that have nothing directly to do with the thesis at issue.

most prominently from Fred Adams and Ken Aizawa (2001).[4] The positive arguments for extended cognitivism struck me as already relatively well made, but I found myself not entirely impressed by many of Clark's responses. The basic position still struck me as entirely correct, but more work, I felt, needed to be done to adequately defend the position. And in order to do that, I felt, the cognitive externalism needed to be given a more precise statement. I had found the project around which my dissertation was to take shape: to elucidate a more precise statement of extended cognition, with a view to defending that position against the many objections ranged against it. In rough form then, the central thesis to be defended – the thesis of extended cognitivism – is the following:

> ***Extended Cognitivism***: the view that some human cognitive states or processes are (partly) constituted by states or processes external to the standard organismic boundaries of the cognitive agent.[5]

§1.2 – THE STRATEGY

With the goal being to give a more precise statement of extended cognitivism and then to defend it against criticism, the next problem was to determine a strategy. In broad outline the positive argument for extended cognitivism is as follows:

1) That many *internal* cognitive processes or states are cognitive because of certain of their features.

[3] The subtitle is presumably an oblique reference to Haugeland (1985).
[4] Which eventually turns into Clark (forthcoming).

2) That many *external* processes or states have precisely those features.

3) Therefore, many external things are cognitive. Or, in other words, cognitive states and processes sometimes extend into the environment.

The above is, in essence, the argument strategy of Clark and Chalmers (1998). As already noted, I had also been persuaded that representations play a central role in cognitive explanation, at least for many paradigmatic kinds of cognition such as arithmetical calculation and memory. Secondly, it struck me that the most persuasive examples of extended cognition involved precisely the same kinds of cognitive processes, also involving representations. As a result, a more fine-grained argument for extended cognitivism emerged:

1) Internal cognition (at least often) involves the use of cognitive representations.

2) Cognitive representations become such in virtue of certain of their features.

3) That many external states or structures that directly shape our cognitive processes have precisely those features.

4) Therefore, many external states or structures are cognitive representations used in cognitive processes. Or, in other words, cognition sometimes extends into the environment.

In order to present extended cognitivism in its strongest light and to give it a more precise statement it would, therefore, be necessary to deal with a couple of preliminary

[5] This is a 'headline' statement. More precision and detail will be found in chapter 5 once a number of preliminary issues have been clarified and some additional conceptual framework provided. See especially §5.1.4.

and broader issues. First and foremost, some account would be needed of the general character of *cognition*, which became the central issue of chapter 2. This divided into a pair of issues concerning the general characters of cognitive *behaviour* and of cognitive *processes*, which gave rise to a series of further questions concerning the nature of *cognitive representation*. Given (as I had come to believe) that cognitive representation is an indispensable feature of at least much of human cognitive processing, this required that I detail the general features of cognitive representations such that they might be readily identified, a matter dealt with in chapter 3. Moreover, the role of cognitive representations in cognition, it became clear, was to carry or bear information. Hence, another concept that required deeper elucidation and discussion was that of *information* – a term used in at least two distinct senses in the literature, and to which I added a third – para-information – in chapter 4.

With these broader concepts – cognition, cognitive representation, and information (in several senses) – clarified, the stage would be set for the presentation and defence of extended cognitivism. Rather than launching directly into a defence of extended cognitivism it seemed wise (or at least useful) to 'lever' the reader towards extended cognitivism by first establishing a less controversial and weaker claim: that human cognition is frequently *embedded* (or as it is sometimes put, 'situated' – although I go on to distinguish between these and certain other related terms) in an environment. This position – defended in chapter 4 – amounts to the claim that human cognition is dependent upon and shaped by, environmental states structures. This claim is sometimes treated as an equivalent claim to that of extended cognitivism (often to the detriment of extended cognitivist arguments) but an explanatory dependency of embedded cognition on environmental structures does not entail a literal extension of cognition into the external world.[6]

With the reader now hopefully more attuned to seeing the environment as an important explanatory contributor to our cognitive capacities, the way was set for a defence of extended cognitivism. This divided naturally into two closely related projects. Firstly, in chapter 5, to state in more precise terms the general thesis of extended cognitivism, and thence to argue for that thesis drawing on several of the key examples in the literature. Secondly, in chapter 6, to argue for the stronger claim that some of those external states are cognitive representations. With the central thesis stated more precisely, the central objections raised against extended cognitivism could now be detailed and more adequately addressed. The central project of the dissertation – to defend extended cognitivism against its central challenges – takes place across these two chapters.

§1.3 – ASSUMPTIONS, EXCLUSIONS, AND AUXILIARY ISSUES

The sweep of the dissertation had become dauntingly large, dealing with nature of cognition, cognitive representation, and information before even coming to the issue of extended cognition. It seemed well, therefore, to try to limit its scope in certain ways. In order to have any hope restricting the dissertation to a manageable size certain claims would have to be assumed and certain other issues left partly or completely unaddressed.

A basic undefended framework for the dissertation is a modest metaphysical and scientific realism. The assumption is that cognitive phenomena are generally amenable to scientific study and explanation, and that the findings of such scientific investigations

[6] Clark and Chalmers are sometimes guilty of not adequately distinguishing between embedded and extended cognition as I characterise these terms. For example, they open their paper with a characterisation

reflect aspects of a metaphysically independent reality. Its modesty lies mainly in not supposing that every such cognitive fact is necessarily amenable to empirical determination. Secondly, as will become clear, my sympathies lie with non-reductive over reductive physicalism, although I shall not enter into the quagmires involved in spelling out precisely to what that amounts. Suffice it to say that I suppose there to be nothing essentially 'spooky' about the mental.

Another important restriction, adopted as much for the sake of allowing a more precise statement of the thesis as for constraints of space, concerns the use of the terms 'mental' (and its cognate, 'the mind'). Except where context requires it, I shall avoid referring to the 'mental' in favour of what I take to be a more precise and appropriate focus on the *cognitive*. I shall have a great deal more to say about the precise nature of cognition in chapter 2, but what is essential to note at this point is that it is not to be identified with the mental. We might note, of course, that the title of Clark and Chalmers (1998) is 'The Extended *Mind*' (emphasis added). But presenting the thesis in such terms invites a number of possible confusions and is fraught with dialectical difficulties. The chief difficulty is that 'the mental' is something of an umbrella term that encompasses features and phenomena that are no part of the thesis to be defended. The most prominent such feature is *consciousness*. Fairly clearly, the domain of the conscious falls within that of the mental, but I do not wish to defend (nor do Clark and Chalmers, for that matter, notwithstanding the title of their paper) any suggestion that the *consciousness* of an individual extends into the world beyond their brain.[7] That thesis is considerably less

of 'active externalism' based on 'the active role of the environment in driving cognitive processes'. Clark and Chalmers (1998), p. 7.

[7] Clark (1997a), p. 215, is also quite explicit that he does not wish to defend the suggestion that an individual's consciousness extends beyond their brain or body. At least one extended cognitivist, however, Mark Rowlands (2002), is willing to endorse such an extension.

tenable. Similar confusions relate to the notion of a *self* – a concept also closely bound up with our everyday understanding of the mental – indeed, some might even wish to say that the self *is* the mind. But to defend the claim that the self extends into the external environment would require a substantive (and probably quite substantial) investigation of the general character of 'selfhood'. Clark and Chalmers (1998) end with speculations about the extension of the self (and Clark [2003] seems willing to express his position in terms of there being an 'extended self'), but for the reasons just outlined I shall not to follow them in that direction.[8] Whilst we are forced to touch briefly upon the closely related issue of agency (§5.6), to undertake an examination of selfhood would threaten to turn the dissertation into a wholly unmanageable project. In any case, the thesis of extended cognitivism does not entail any other concerning the parallel extension of self.[9]

None of this, of course, is to say that the notions of the self or the conscious are not legitimate topics of scientific study but simply to point out that they are distinct from the cognitive. For example, it is generally agreed that cognition can be non-conscious and it remains at least questionable whether being conscious is a sufficient condition for being cognitive. There are those – John Searle (1992, 1997) jumps immediately to mind – who have argued that consciousness is at the core of human cognitive processes, and in particular that no account of mental (note the term) representation can be hoped for without appeal to it. Consciousness may well have a place in cognition, but the dissertation will have little or nothing to say on such matters and will follow the

[8] Clark (1997a) is, for example, forced to give an ambivalent 'yes and no' answer to the question of the extension of the self into the environment as a direct result of tackling the issue in such broad terms.

[9] Clark (1997a), p. 218, is more restrained and adopts an approach closer to the one advocated here. He says that he is 'content to let the notions of self and agency fall where they will. In the final analysis, I assert that we have, at a minimum, good explanatory and methodological reasons to (at times) embrace a quite liberal notion of the scope of computation and cognitive processes – one that explicitly allows the spread of such processes across brain, body, world, and artefact.' By Clark (2003) he is more cavalier about the use of 'self'.

mainstream of cognitive science in treating cognitive representation and cognition generally as capable of being studied without requiring an account of phenomenology or consciousness.

I shall not, therefore, give an account of *mental* representation (instead, I shall give an account of *cognitive* representation); I shall not defend the claim that the *mind* sometimes extends beyond the boundaries of brain or skin (but instead that an individual's *cognition* extends beyond their brain and body), and so on. In this respect I largely diverge, in terminology rather than in substance, from many others with whom I strike common cause.[10] This may disappoint some, but I must confess my suspicion that the concepts of 'mind' and 'the mental' do not give expression to a single theoretically unified set of properties or objects, and that switching between the terms 'cognitive' and 'mental' as though they were virtually synonymous only confuses the issues.[11] In any case, to re-emphasise, the thesis to be defended is that an individual's *cognition* extends beyond the boundaries of their brain or body.

A final exclusion of note concerns one of the initial motivations for exploring the thesis of extended cognitivism mentioned above. Although I will have a great deal to say about cognitive representation, I will not be discussing, except in passing, representational *content*. An earlier version of the dissertation included a lengthy discussion of representational content, but it became clear that the central thesis neither required, nor was it particularly aided by a lengthy voyage through that particular quagmire. Although I believe that extended cognitivism has some useful insights to offer (and I offer some nascent speculations to this end in §7.2) such is the scale of such a task that an adequate treatment of these issues will have to wait for another occasion.

§1.4 – CHAPTER SUMMARY

The dissertation can be seen as dividing loosely into two parts. Following this introduction, the first part – chapters 2 and 3 – lays the basic groundwork for an understanding of the key concepts that are essential to assessing the cogency of the claim that an individual's cognition sometimes extends beyond the boundaries of their brain or body: cognition, cognitive representation, and information. The general strategy in these chapters is to move from the outside in, as it were; that is, to begin with the broader issues and then to focus in progressively more narrowly on the issues and concepts that arise out of the broader analysis. Specifically, I begin with an examination of the general character of cognition, which in turn raises questions concerning the nature of cognitive representation and the role(s) of information in cognitive explanation. The second part – chapters 4 through 6 – develops and defends the central substantive claims relating to extended cognitivism. The strategy in the second part is to move from a defence of weaker claims to a defence of stronger claims. Thus, I begin with embedded cognition, move next to a general defence of extended cognitivism, and finish with a defence of a specifically representational version of extended cognitivism. Chapter 7 offers some concluding remarks and some speculations.

§1.4.1 Chapter 1 – Introduction

This chapter provides a brief history (§1.1) of the origins and evolution of the dissertation and introduces in a rough form the thesis of extended cognitivism. The strategy for the defence of the central thesis is explained (§1.2) and the broad structure of

[10] I am thinking here primarily of Clark and Chalmers (1998), Rowlands (1999), and Dennett (1996).
[11] In Clark (forthcoming) he mentions that he has flirted with (though ultimately has rejected) such a claim.

the dissertation is laid out. A number of important assumptions are put in place and a number of issues are identified as orthogonal to the thesis and hence excluded from further consideration. The chapter closes with this summary of each chapter.

§1.4.2 Chapter 2 – Cognition

The central goal of this chapter is to get as clear a handle as possible on precisely what distinguishes *cognitive* processes and *cognitive* behaviours from non-cognitive processes and behaviours. In §2.1, I outline my strategic approach to the issue of the nature of cognition and separate the question into two parts: the nature of cognitive behaviour and the nature of cognitive processing. In §2.2, I discuss several common approaches to the nature of cognitive behaviour. Several inadequacies and concerns arising out of these approaches are canvassed, and, turning then to cognitive processes in §2.3, I focus on, and give a preliminary and qualified endorsement to, the dominant position: the view that cognitive processing involves information processing. This claim, in its turn, makes appeal to a number of concepts – most especially 'representation' and 'information' – the closer examinations of which will be the focus of chapters 3 and 4.

§1.4.3 Chapter 3 – Cognitive Representation

Having arrived at a qualified and preliminary endorsement of a standard line that cognitive processing involves *information* processing – i.e., the processing of *representations* – a number of ambiguities and confusions were noted that beset these notions. The central concern of chapter 3, is to provide a clearer understanding of what it is for something to be the kind of thing that can play a representational role in a cognitive process. I begin (§3.1 – the bulk of the chapter) by identifying and discussing a set of characteristics for something to be such a representation. The resulting characterisation of cognitive representation is intended to be sufficiently broad and ecumenical to cut across partisan divides. Armed with an ecumenical characterisation of cognitive representation and drawing partly on the arguments of Andy Clark, the second goal of this chapter is to defend (§3.2) the claim that we need such representations to explain a significant body of our cognitive behaviour.

§1.4.4 Chapter 4 – Para-information and Embedded Cognition

In this chapter we begin our journey out into trans-cranial and trans-corporeal cognitive world. The primary focus of this chapter is an examination of a pair of closely related concepts: *para-information* and *embedded cognition*. The former is usually (confusingly, I suggest) discussed under the heading of 'implicit' information. Certain cognitive behaviours, I argue, require an explanatory appeal to information-bearing features of the world – para-informational states or structures – that enable a cognitive agent to extract task-relevant information.

After first distinguishing para-information from implicit information (§4.1.1), I begin (§4.1.2), drawing, in part, on Dretske (1981, 1988), with a discussion of some internal cases of para-information. I argue in favour of a central role for para-information in the extraction of truth-dependent information. Next (§4.1.3), I introduce some external cases of para-information where, I argue, it plays precisely the same explanatory role. After a brief comment on the representational status of para-informational states or structures (§4.1.4), these claims are developed through a detailed examination of an example of external para-information (§4.1.5).

The relationship between embedded cognition and para-information is then crystallised in §4.2 through a brief analysis of the nature of dispositions (§4.2.1), and hence (qua dispositions), of embedded cognitive *capacities*. Dispositions, I argue, supervene upon the salient properties of the systems to which they are notionally attributed *together with* those of the systemic elements of their activating conditions. By extension (§4.2.2), embedded cognitive capacities should be seen as supervening upon cognitive agents considered not in isolation from, but *together with*, cognitively significant features of their environments.

§1.4.5 Chapter 5 – Extended Cognition

Over chapters 5 and 6 I discuss and defend the central thesis of extended cognitivism: that certain processes occurring beyond the boundaries of a cognitive agent's body, are as much parts of that agent's cognitive processing as anything going on in their brain. The focus in chapter 5 will be on an elucidation and defence of the *general* thesis of extended cognitivism – one that is silent on the underlying character of extended cognitive processing. I begin (§5.1) by distinguishing extended cognitivism from some of its more or less close relatives and set aside certain misconceptions and confusions concerning extended cognition. This is followed by an attempt to situate extended cognition within a relatively orthodox tradition in cognitive science. After then introducing certain distinctions and defining some important terms relating to cognitive processes, states and systems, I state in more precise terms than hitherto the general thesis of extended cognition. After examining (§5.2) a number of central putative examples of extended cognition, and drawing most prominently on the arguments of Fred Adams and Ken Aizawa, I turn (§§5.3-5.6) to an critical examination of four central objections raised against the thesis in the literature: the coupling-constitution fallacy, the intrinsic intentionality objection, the scientific explanation objection, and the control objection. All of these objections are rejected.

§1.4.6 Chapter 6 – Extended Intentional States

In this chapter I examine and defend a stronger version of extended cognitivism that includes a commitment to specifically *representational* extended cognitive processes and states. I begin (§6.1) with an examination of the central role of *cognitive technology*, broadly understood, in human cognition. This is followed (§6.2) by a defence of a representationalist interpretation of the central examples considered in the preceding chapter. In the following section (§6.3) I characterise extended cognitivism in respect of external belief states and intentional explanation – the central focus of the chapter.[12] The critical example, drawn from Clark and Chalmers (1998), is that of Otto, a sufferer of a mild form of Alzheimer's disease, and his notebook. Following Clark and Chalmers, I argue that the role that Otto's notebook plays in his cognitive economy qualifies it as a repository of his beliefs.

I then consider and ultimately reject three significant objections raised against this claim. The first – the cognitive bloat objection – suggests that extended cognitivism implies the extension of belief states to counter-intuitive cases and is therefore absurd. The central thrust of this objection (that extended cognitivism implies the extension of belief states to counter-intuitive cases) is conceded, but the inference to absurdity is resisted on several grounds. A second objection – the 'Otto 2-Step' (§6.5) – suggests that all the explanatory work is done by in the head processes and states. This is also rejected, primarily through distinguishing between the explanatory roles of occurrent versus non-occurrent beliefs. The final objection (§6.6) argues that the role of perception provides a principled 'barrier' between cognitive processes and states occurring within the head and other (putatively non-cognitive) processes or states occurring outside of the head. Responding to this objection will serve as an occasion to develop an account of belief-desire intentional explanation in the context of cognitive externalism. Problems are identified with Clark and Chalmers original handling of the objection and the chapter culminates with some revised conditions for belief attribution that more adequately avoid the objections raised against the Otto example.

§1.4.7 Chapter 7 – Conclusion

In this chapter the central arguments the preceding chapters are summarised. A number of avenues of future research are explored as springing from the dissertation and some speculations offered.

[12] There is a familiar ambiguity when talking about beliefs: i.e., between the *content* of the belief and the content-*bearer*. The issue of extended cognitivism and its challenge via cognitive bloat concerns the location of such content-bearers and as such, when referring to some cognitive agent's beliefs, I should be understood as referring to the content-bearer.

CHAPTER 2 – COGNITION

In order to defend the extended cognitivist thesis that an individual's cognition sometimes extends beyond their brain and body, it will be well to get as clear a handle as possible on precisely what distinguishes *cognitive* processes and behaviours from *non*-cognitive processes and behaviours. To this end, in §2.1, I adumbrate my approach to the issue of the nature of cognition and separate the question into two parts: the nature of cognitive *behaviour* and the nature of cognitive *processing*. In §2.2, I discuss several common approaches to the first question. Several inadequacies and concerns arising out of these approaches will be canvassed, and I conclude that cognitive behaviour is best identified by certain typical 'markers' rather than in terms of some defining set of necessary and sufficient conditions. Turning then to cognitive processes in §2.3, I give a preliminary and qualified endorsement to the dominant account: the view that cognitive processing involves information processing. I argue that it is important to keep front and centre that this claim is a hypothesis that may admit of exceptions, not an a priori commitment. The information processing hypothesis, in its turn, makes appeal to a number of concepts – most especially 'representation' and 'information' – the close examinations of which will be a focus of chapters 3 and 4.

§2.1 – GETTING CLEARER ON THE QUESTIONS

Before launching into the main substance of the chapter, it will be well to get as clear as possible on the exact question(s) before us. The first important thing of note is that the central thesis of extended cognitivism – that some of an individual's cognitive processing or states sometimes occur outside of his or her skin – does not require the provision of an entirely general or exceptionless characterisation of cognition. More specifically, the thesis to be defended may allow some processes or states that do not fall under the proffered characterisation of 'cognitive' still to count intuitively as cognitive, just so long as the central putative cases of extended cognition are clearly cognitive. For example, should the characterisation of 'cognitive' presented herein exclude, say, perception or emotion, this will be of little concern, for the thesis to be defended is only that *some* of the cognitive processes or behaviours of individuals lie (in part) outside their bodies. What the thesis *does* require, however – and which constitutes the central concern of this chapter – is the provision of some basis on which we can, with confidence, reliably identify case(s) of cognition as such; conditions that (as chapters 5 and 6 will argue) are satisfied in the case of some external processes and states, thereby showing that they too should count as cognitive. Ideally, this basis would be constituted by a set of necessary and sufficient conditions, but as we shall shortly see, such conditions are not to be had and so we will be forced to settle for a more modest basis.

Given that the thesis that some external processes or states are genuinely cognitive in nature is counter-intuitive to many, the case for this thesis will be made more persuasive by adopting a characterisation of cognition as consentient as possible to paradigm cases. Moreover, even though the paradigm cases of cognition may be uncontroversial, the proposed sufficient conditions for cognition may, of course, be controversial, but (if, indeed, it be the case) assuaging such concerns would be the burden of the arguments given below.

§2.1.1 The Nature of the Question(s)

The simplest and most direct way to pose our question is to ask 'what is cognition?'. However, posing it in such terms invites answers that compound two distinct issues. Consider, for example, Hunt and Ellis (1999). Cognition, they tell us, is 'a class of *symbolic* mental activities *such as* thinking, reasoning, problem solving and memory search'.[13] Such answers combine a characterisation of the *domain of behaviour* to be explained with a theoretical commitment to the *kind of process* that explains how such behaviour is possible. Like many, Hunt and Ellis combine a theoretical commitment to cognitive processes being the manipulation of symbols with a characterisation of the domain of cognitive behaviour – in this case, by listing some paradigmatic examples. There are, therefore, at least two respects in which it might be argued that something is cognitive, and hence, two questions hidden within the broader question 'what is cognition?'. The first respect in which something might be said to be cognitive is by belonging to the class of cognitive *behaviours*. This focuses on the *extension* of the behavioural or phenomenal domain to be investigated. The second respect in which something might be counted as cognitive is through being a cognitive *process* – that is, by being a process of the same general kind as that which explains cognitive behaviour. This focuses on the issue of how cognition is *implemented* or *constituted.* These two respects give rise to two closely related yet importantly distinct questions that approach the issue of the nature of cognition from opposite directions: one from the direction of the explananda, the other from the direction of the explanans:

(a) What is cognitive *behaviour*? That is, what is it that distinguishes such behaviour, considered as a unified phenomenal domain, from non-cognitive behaviour?

(b) What is a cognitive *process*? That is, what kinds of processes explain cognitive behaviours?

Answers to the first question seek to delineate the *domain* of behaviour that is appropriately regarded as in need of cognitive explanation rather than to identify the general nature of the processes (e.g., through symbol manipulation) in virtue of which such behaviours occur. In the ideal case, answers to this question will result in the provision of something approaching a definition that captures all and only cognitive behaviours. This is not to say that the class of cognitive behaviours need be determined as a matter of explicit definition – of providing necessary and sufficient conditions – for, prima facie, there may be none. It could in fact be that nothing more than more or less vague family resemblances unite cognitive behaviours, or that cognitive behaviours are best delineated by nothing more than listing its paradigmatic instances.[14]

Answers to the second question typically attempt to provide an *a posteriori* explanatory theory or framework that purports to describe the kinds of processes that underlie cognitive behaviour as such, thereby providing, at a suitably general level, an account of *how* cognitive behaviour takes place. Examples include the claim that cognitive processing consists of computer-like syntactic manipulations of representations, or the claim that it consists of vector transformations in connectionist-style neural hardware.

[13] Hunt and Ellis (1999), p. 335 (emphasis added). A similar running together of these issues can be found in Gardner (1985), p. 6. Not every commentator ignores this distinction. For example, in a refreshingly thoughtful and self-aware discussion, Harnish (2002), pp. 4-5, distinguishes the domain of cognitive behaviour from the characterisation of cognitive processes under the banner of 'broad' and 'narrow' conceptions of cognition, respectively.

§2.1.2 The Relationship Between the Questions

The most likely reason for why these questions (and their respective answers) are frequently compounded is, I suspect, that although distinct, they are intimately connected such that answers to one question will frequently strongly influence answers to the other. For example, if one is committed to the claim that cognitive processes are symbolic in nature, one will likely be strongly inclined to exclude from the domain of cognitive behaviours (or at least push to the periphery) any and all behaviours that resist (or do not necessitate) symbolic analysis or explanation. Conversely, if one insists on the inclusion of certain behaviours that resist (or do not necessitate) symbolic analysis or explanation, one will likely be strongly inclined to reject the claim that all cognitive processes are necessarily symbolic. For example, various researchers have offered non-symbolic accounts of the processes that govern what they take to be cognitive behaviour.[15] Examples include Brooks (1991, 1996) and Beer (1995) who offer analyses of robotic locomotion and ant-walking, respectively. Supporters of such non-symbolic approaches have taken such research to be evidence that not all cognitive processing need be symbolic, or even that perhaps none of it is. By contrast, critics, such as Kirsh (1991), typically taking such things as reasoning and language to be central to cognition, have retorted, in effect, that robotic locomotion and ant-walking are hardly what we should think of as particularly *cognitive* kinds of behaviour.

Given the close connection between the question of process, the question of behavioural domain, and the kinds of disputes mentioned above, the broader question

[14] Similarly, prima facie at least, it may even be that cognition does not constitute either a natural or a functional kind.

[15] Many have made the stronger claim that there is non-representational, not merely non-symbolic, cognitive behaviour – a distinction I address in §2.3, below.

naturally arises as to which question, if either, should assume priority.[16] As we shall see, however, good reasons can be given for denying either priority over the other.

To begin with, a cursory glance over the development of many and perhaps most sciences reveals that it is frequently theoretical considerations that end up determining the domain of behaviour. Often, what was initially or pre-theoretically taken to be a clearly delineated domain of behaviour or phenomena will come to be explained through two or more theories that are sufficiently distinct as to prompt the division of the domain. For example, 'burning' – friction burns, chemical burns, burning wood, the burning of the Sun, etc. – comprises a set of phenomena once considered unified under a single domain, but which is now recognised as resulting from very diverse aetiologies. Despite phenomenological similarities and conceptual associations between its manifestations, the success of quite different *a posteriori* explanatory theories for these phenomena has led to them no longer being considered as constituting a single behavioural or phenomenal domain.[17] On other occasions, by contrast, the scope of a theory may be found to extend well beyond its original domain of behaviour, thereby prompting an enlargement of the domain of the behaviour in question. Game theory, for example, has roots ostensibly in an attempt to model bluffing in parlour games, but was soon to take up prominent positions in the fields of economics, military strategy,[18] and evolutionary theory. Moreover, the justification for the influence of theory on the behavioural domain is not difficult to

[16] Assigning priority to either question will not necessarily, by itself, resolve the kind of disputes just mentioned because, for example, disputes may yet remain over which behaviours are central. Some, such as John Haugeland, a cautious critic of computationalism, assign cognitive centrality to the class of behaviours sometimes classified as 'skillful coping' – everyday activities such as reaching for beer in the fridge, climbing stairs, or dancing [see Haugeland (1995)]. Newell (1990), pp. 15-16, is equally clear that such routine actions are central to cognition – more central than language, he thinks – despite being an arch computationalist.

[17] Note, however, that in the case of burning at least, division amongst separate sciences using quite distinct explanations does nothing to diminish the claim of each phenomena to be considered genuine cases of burning. It is not as though physical chemists, for example, try to tell physiologists that a radiation burn is not *really* a burn.

discern, for to do otherwise would be to bind theories to our pre-theoretic intuitions and to condemn science to operating within a folk taxonomy. Barbara Von Eckardt surely has a point when she suggests that:

> Scientific domains do not come ready-made. Rather, we begin with a working hypothesis as to what set of phenomena will be susceptible to the same theoretical approach; we then gradually refine and modify this hypothesis as the construction of a theory proceeds. (Von Eckardt, 1993, p. 6.)

Equally, however, the construction of a theory that serves to refine and modify the hypothesis concerning the domain of behaviour must presumably be determined by how well it explains some body of behaviour. It may be tempting, therefore, to assign logical priority to the second question: to think that to do otherwise would be to put the explanatory cart before the behavioural horse. After all, the truth or falsity of putative *a posteriori* accounts of the nature of cognitive processing will depend, presumably, on having appropriately delineated the domain of cognitive behaviour from non-cognitive behaviour. In other words, or so it might seem, unless or until we have in place a relatively clear idea of which behaviours are cognitive as opposed to non-cognitive, it will be entirely moot which kinds of processes or states are employed in their performance. By analogy, our willingness to accept or reject a putative theory of *optics*, say, will presumably depend upon the theory's capacity to explain some pre-theoretical class of *optical* behaviour. As Ernest Nagel (1961) has observed, normally we come ready-armed with some pre-theoretic class of behaviour for which we seek explanation.

The exact nature of the relationship between the determination of a domain of behaviour and the explanatory theories and methods that are developed to explain things within that domain raises deep and complex issues within the philosophy of science. A

[18] See Amadae (2003) for a history with a particular emphasis on its intended role in cold-war nuclear strategy.

fuller discussion of this relationship would threaten to take us too far afield. What is suggested by the brief arguments presented above is that both approaches to the nature of cognition – those that focus firstly on the behaviour domain and those that focus firstly on process – can be supported with prima facie good arguments. Without prejudice to either approach, therefore, I will consider each in turn, beginning with various attempts to characterise the nature of cognitive behaviour. Ultimately, as we shall see (§2.3), addressing the issue from the direction of cognitive processing will provide us with the more satisfactory approach to questions concerning the nature of cognition, but without assigning dictatorial priority to questions of process over questions of behaviour.

§2.2 – WHAT IS COGNITIVE BEHAVIOUR?

§2.2.1 Cognitive Behaviour Defined Paradigmatically

A common approach to circumscribing cognitive behaviour characterises the domain *paradigmatically* by simply listing either some set of disciplines said to be involved in its study, or else some subset of the various capacities or behaviours that such disciplines study. The first of these is encouraged by the very idea of a unified programme of research travelling under the banner of 'cognitive science'. Standardly, cognitive science is said to comprise philosophy, linguistics, psychology, anthropology, computer science, and neuroscience, together with the various intersections of these disciplines.[19] Characterising this loose agglomeration of disciplines under the banner of 'cognitive science' presupposes that there is something that each of these disciplines, in its own way, studies and seeks to explain. It might be natural to suppose that the subject matter of these

[19] The classic formulation of the disciplines involved in cognitive science is represented in the 1978 Sloan Report's 'cognitive hexagram', reproduced in Pylyshyn (1983), p. 76.

disciplines is *cognition*, but certain commentators disagree, or at least are not so clear. Some blithely identify cognitive science with the study of *the mind*, thereby implicitly identifying the *cognitive* with the *mental*. This is an especially egregious confusion from the point of view of the present dissertation because it risks indirectly begging the question against its central thesis. Consider Von Eckardt (1993) where it is variously claimed or intimated that cognitive science is 'an approach to the study of *mind*' which theorises about 'what the *mind* is like'; yet, she also says, it is the study of '*cognition*', and (running these elements directly together) 'the *cognitive mind*' (Von Eckardt, 1993, introduction – emphasis added). To identify the cognitive with the mental, or, by extension, to identify cognitive science with the study of the mind risks running roughshod over a potentially important distinction. In the first place, in practice cognitive scientists plainly do not identify the domains of the mental and the cognitive as coextensive. It is, for example, at least controversial whether emotions, which are prima facie mental states, are relevant to cognition. Consider, further, Von Eckardt's criticism of Martin Gardner's definition of cognitive science. Gardner (1985) defines cognitive science as 'a contemporary, empirically based effort to answer long-standing epistemological questions – particularly, those concerned with the nature of knowledge, its components, its sources, its development, and its deployment' (Gardner, 1985, p. 6, quoted in Von Eckardt, 1993, p. 66). She dismisses this characterisation as being 'much too broad':

> Human beings represent and use their 'knowledge' in many ways, only some of which involve the human mind. What we know is represented in books, pictures, computer databases, and so forth. Clearly, cognitive science does not study the representation and the use of knowledge in all these forms. Thus, at a minimum, [Gardner's definition] must be amended to read: The domain of cognitive science is human knowledge representation and use *in the mind*. (Von Eckardt, 1993, p. 67. – original italics.)

So Von Eckardt's grounds for excluding knowledge represented in books, pictures, computer databases, and the like as being outside the scope of cognitive science (and not cognitive) is that they are not represented or used *in the mind.* Clearly, of course, so long as we suppose that the mind is bounded by the skull, or that the mental is characterised by appeal to some feature (such as consciousness or perhaps a unified sense of self) that appears not to extend beyond the skull, such representations and their uses are not 'in the mind'. But nowhere is it made clear *why* we should suppose that such representations as are found in books and computer databases, etc., are not *cognitive.* As indicated in chapter 1, the concepts of the mental and of the cognitive may come apart and we must not lose sight of that fact in trying to characterise cognition.[20]

Refocusing our attention back on the cognitive, however, it would be equally misleading to suggest that cognition is simply *whatever* these disciplines are engaged in studying or explaining, for there are significant portions of many of these disciplines that are engaged in studying or explaining phenomena that are not obviously cognitive in character. Computer science, for example, studies such things as the properties of semi-conductors, and anthropology studies such things as cross-cultural variations in institutions of marriage – in neither of these cases, I take it, do the researchers take themselves to be studying cognition. Hence, in order for this approach to shed any light on the nature of cognition, at least *some* attempt must be made to elucidate what warrants the claim that certain portions of these disciplines study the same thing – i.e., *cognition.* As William Bechtel has pointed out, 'It is a truism to say that the subject matter of cognitive

[20] Perhaps Von Eckardt's goal is merely descriptive – to point out that cognitive science does not study external representations. If so, her claim is at least highly contentious. Anthropology, linguistics (speech productions), and computer science can easily be described in such terms.

science is cognition. But for this to be informative, one must say what one means by *cognition*' (Bechtel, 1999, p. 156.).

Often, though, little more is provided by way of a theoretically unifying notion of cognition than either a) a series of largely uninformative partial answers or near synonyms for cognition, or b) a list of various paradigmatic cognitive behaviours. Of the former we typically find that cognition is said to be 'thinking' or 'intelligence'.[21] Of the latter we typically find a list including (but usually not limited to): perception, language, problem solving, reasoning and inference, decision making, memory, learning, and (less paradigmatically, according to some) emotion, and motor control. There is nothing obviously false about such claims, but the fact remains that such lists fail to illuminate the *nature* of cognitive behaviour – i.e., what distinguishes it from non-cognitive behaviour – in a theoretically useful or interesting way. However convenient or accurate a given list may be, it fails to elucidate what, if anything, unites such behaviours as distinctly cognitive. We are typically left with the feeling that, to parallel a US Supreme Court judge on the topic of obscenity, although one might not be able to say what it is, cognitive behaviour is something that one ought to be able to recognise whenever one encounters it.

§2.2.2 Cognitive Behaviour Defined as 'Intelligent' Behaviour

A more informative attempt at characterising cognitive behaviour comes with the claim that cognitive behaviour is 'intelligent' behaviour. This is amongst the most commonly cited phrases used to characterise the domain of cognitive behaviour. But what makes some behaviours intelligent and other behaviours not? 'Intelligence', and its

[21] For example, Anderson (1980), p. 3, says that cognitive psychology 'attempts to understand the nature of human intelligence and how people think'. The characterisation of cognition as intelligence receives closer examination in the next section.

product, 'intelligent behaviour' has historically been a particularly slippery pair of concepts. In 1921 expert opinion on the meaning of 'intelligence' included the following (Sternberg, 1987, p. 375):

- The power of good responses from the point of view of truth or fact
- The ability to carry on abstract thinking
- Having learned or ability to learn to adjust oneself to the environment
- The ability to adapt oneself adequately to relatively new situations in life
- The capacity to learn or to profit by experience
- A biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behaviour.

Intelligent behaviour would then be the behaviour that issues from this motley assortment of capacities and abilities. Several different features are apparent in the above, including the role of intelligence in comporting the subject appropriately to its environment, learning from experience, and the capacity to abstract or generalise. More recently, however, these have been pared down to two features that stand out especially frequently in the literature on cognition: the first is that cognitive processes give rise to behaviours that are highly *flexible* and *plastic*; the second is that cognitive processes are (or involve) the *processing of information.* In the following two sections I shall examine and criticise the use of flexibility or plasticity as distinguishing or defining features of cognitive behaviour, turning to the claim that cognitive processing is (or involves) the processing of information in §2.3.

§2.2.3 Cognitive Behaviour Defined as 'Flexible' and 'Plastic'

What does it mean to say that a given behaviour is *flexible* or *plastic*? Although these terms are often used interchangeably (and are sometimes used interchangeably with behaviour that is modifiable by experience – see Sterelny, quoted below) for the purposes of the ensuing discussion it will be useful to distinguish between them and to add to them a third, *learnability*. *Flexibility*, as I shall understand it, is the capacity to modify *behaviour* appropriately in relation to circumstances (i.e., behavioural flexibility). *Plasticity* is the capacity to modify an *existing rule* that is guiding behaviour. Plasticity, therefore, can be seen as roughly akin to second-order behavioural flexibility – a flexibility in a rule guiding one's first-order behavioural flexibility.[22] Behaviours can, therefore, be flexible even though the rule guiding that behaviour may not be plastic. Intimately connected (and sometimes compounded) with these two notions is a third: *learnability*. Behaviours are *learnable* to the extent that the rules guiding them are *acquired* or *modified* through experience. Thus, some behaviours may be learnable even though the rules shaping those behaviours are not plastic, because once acquired the rule cannot be *further* modified in the light of novel circumstances.[23]

Some linguistic examples may help to illustrate these terms. Linguistic behaviour is, in general, highly *flexible*. One varies one's linguistic behaviour appropriately in response to differing circumstances, including responding appropriately to entirely novel sentences. These behaviours result from the application of the (typically unconscious) linguistic rules of one's language. Those linguistic rules are also, to a high degree at least, *plastic*. For

[22] The characterisation of plasticity as second-order flexibility may not be exact (hence my use of 'roughly akin'). It suggests that there must be a second-order rule that modifies the first-order rules and this may be overly deterministic.

example, someone who regularly ended sentences with prepositions might learn to change the rule governing that part of their linguistic behaviour, adopting another (stylistically preferable) rule in its place. Linguistic behaviours are also very obviously *learnable*, since it is typically through *experience* of the language spoken by one's immediate linguistic community that one modifies or acquires such linguistic rules. Some linguistic behaviours, however, are shaped by rules that are *not* particularly plastic. For example, parameter-setting in the learning of a language (for example, whether one's phrase structure puts the head first as in English or the head last as in Japanese) results in a rule that shapes flexible linguistic behaviour, but the rule itself it is not very plastic; once it is set it is very difficult to unset or modify.[24] Moreover, given Chomskian orthodoxy, the rules of deep grammar that putatively guide such parameter-setting are not learned through experience.

Kim Sterelny, borrowing from Dennett,[25] attempts to elucidate these notions negatively by way of an example that putatively fails to exhibit the requisite properties.[26] A Sphex wasp builds a burrow for its eggs, finds and paralyses a cricket, and drags it to the burrow. She inspects the burrow, comes out, drags the cricket inside, comes out again, seals the burrow, and departs. But if an experimenter, whilst the wasp is inspecting the burrow, moves the cricket away from the threshold of the burrow a few inches, the wasp, upon emerging from the burrow, always returns the cricket to the threshold and inspects the burrow again. Sterelny comments: 'Sphex has a single, invariable behaviour pattern. It is insensitive to new contingencies and requirements; it is unmodifiable by learning.'

[23] Putting aside a certain undoubted innate contribution, handedness might be a case in point. Once handedness is set through repeated use it can be very difficult to modify handed-behaviours, even for those who, say, might be innately disposed to be left-handed but were conditioned to be right-handed.

[24] True, native Japanese speakers can become fluent English speakers and vice versa, but the evidence suggests that these operate, in large part, through largely different cognitive systems.

[25] Dennett (1978), p. 65. Dennett draws the example from Wooldridge (1963), p. 82.

(Sterelny, 1990, p. 20). The behaviour of human beings, he says, is by contrast highly 'plastic':

> that is, [our behavioural repertoire] is modifiable by experience. Sometimes those modifications are appropriate. This plasticity is a consequence of our sentience: intelligent creatures can learn new tricks, can change their ways. Our behavioural repertoire, and that of some animals, is open ended. (*ibid.*)

Let us note, in passing, that characterising intelligence in terms of flexibility or plasticity renders the concept of intelligence somewhat vague and consequently unable to provide a categorical criterion for intelligence or cognition. This is because it is left unclear quite *how much* flexibility or plasticity in a behaviour might be sufficient for it to be considered intelligent (and, therefore, cognitive). Suffice it to say that, if flexibility or plasticity are taken as the distinguishing features of intelligent or cognitive behaviour, it will lie on a continuum with non-intelligent and non-cognitive behaviour. This is not necessarily a fatal flaw, of course, for many sciences continue to operate quite successfully without having boundaries that unambiguously or uncontentiously delineate their domains from those of their neighbours. The boundaries between physics and chemistry, chemistry and biology, or biology and other life sciences are often far from clear and frequently contentious, but this does not prevent these sciences either from co-existing with each other or from having clear and substantive domains of enquiry of their own. Moreover, a continuity between the cognitive and the non-cognitive may be viewed by many as a virtue more than a vice, since it would more likely render more tractable the

[26] Sterelny (1990), pp. 19-20. Dennett (1978), p. 66, draws the same conclusions as Sterelny concerning the behaviour of the Sphex wasp: 'When we see how simple, rigid and mechanical it is, we realise that [in characterising it as intelligent] we were attributing too much to the wasp'. As we shall see, the claim that the Sphex wasp's behaviour is inflexible is somewhat questionable.

task of situating cognitive science in relation to other sciences, especially biological sciences.[27]

§2.2.4 Problems with Flexibility and Plasticity

Despite the popularity of the emphasis on behavioural flexibility and plasticity (and with them, learnability) as defining features of intelligent or cognitive behaviour, they do not stand well as either necessary or sufficient conditions. There are, in fact, several problems with these notions as criteria for intelligent behaviour, and, thereby, as distinguishing marks of cognitive behaviour. Nevertheless, plasticity in particular serves well as typical 'marker' of cognitive behaviour. To see this, we shall examine each concept in turn.

At first glance, a key feature of the Sphex wasp's behaviour that renders it putatively 'inflexible' is that it is incapable of varying its behaviour appropriately in relation to circumstances outside a very limited range – in particular, of not bothering to inspect its burrow a second time when the paralysed cricket is moved only a few inches from the lip of the burrow. The implication seems to be that if its behaviour varied appropriately with the appearance of novel circumstances its behaviour would stand a better chance of counting as cognitive.

Note, firstly, that such flexibility does not, at least in principle, require that the rule shaping its behaviour be either modifiable or (therefore) plastic, only that it should vary appropriately with circumstance. However improbable it might be, a system might, at least in principle, be equipped with a behaviour-guiding rule sufficiently complex and complete that it covers every actually encountered contingency, always varying behaviour

[27] I take This to be especially true of Dennett, whose gradualist approach to cognition emphasises a

appropriately in relation to the circumstances. Even though such a rule might be entirely fixed, its behaviour would remain flexible in the above sense.

The central problem with behavioural flexibility as a distinguishing mark of cognitive behaviour is that behaviour does not appear to need to be appropriately varied with circumstances in order to be cognitive; many a cognitive behaviour is highly *in*appropriate – many a cognitive behaviour is, indeed, 'stupid'. Those familiar with the Darwin Awards[28] will know of many cases of profoundly inappropriate (and fatal) human behaviour that manifestly do involve such things as perception, (often tragically faulty) inference, and other uncontroversially cognitive behaviours.[29] It will not do to object to this point by appeal to the fact that, notwithstanding that it fails to produce appropriate behaviour, it remains (presumably) the *function* of the system in question to produce appropriate behaviour, for this, too, is (presumably) the function of whatever system controls the Sphex wasp's putatively inflexible behaviour. Neither will it do to drop the insistence that behaviour be appropriate and merely insist that intelligent behaviour must vary with circumstance, for this renders the notion of flexibility so weak as to apply to pretty much anything. The Sphex wasp, for example, also varies its behaviour with circumstance to the extent that, if the burrow is empty it will attempt to place a cricket in it, but if the burrow is not empty, then it will not. Viewing flexibility in this way – as little more than the capacity to respond differently to different circumstances – leaves us with a condition that is surely too weak to substantiate the alleged difference between human behaviour, qua intelligent and cognitive, and Sphex wasp behaviour qua unintelligent and

continuity between the cognitive and the non-cognitive.

[28] A selection can be found at www.DarwinAwards.com, and in Northcutt (2000).

[29] There is, perhaps, a nod to this point in Sterelny (1990). p. 20, where he notes that 'sometimes those [behavioural] modifications are appropriate'. And, presumably, sometimes not – hence the irrelevancy of the *appropriateness* of behaviour as a defining necessary feature of cognitive behaviours.

non-cognitive. Indeed, it is so weak that almost any behaviour could then be said to be 'flexible'. For example, even a light bulb has a 'flexible' response to circumstance in that it behaves differently when the light switch is up from when it is down.

Turning now to plasticity, it was defined above as a capacity to *modify*, in the light of changing circumstances, the rule that guides (perhaps first-order flexible) behaviour. In emphasising plasticity the implication is that intelligent behaviour is behaviour that is governed by a rule that is indefinitely 'modifiable by experience'.

But plasticity is also highly problematic as a distinguishing feature of cognitive behaviour, for there would seem to be several highly rigid and *un*modifiable rule-driven behaviours that are uncontentious manifestations of intelligence and cognition. Take, for example, our ability to engage in arithmetical and logical operations and inferences – cast-iron candidates of intelligent cognitive behaviour if ever there were any. Whenever I find myself in circumstances in which I believe that I have one item in one hand and two items in the other, I *invariably* form the belief that I have three items in my hands. Though the rule shapes highly flexible inference drawing behaviour (always drawing the appropriate conclusion from the premises, say), the rule itself is entirely rigid and unmodifiable by further experience, hence not plastic. This is true even if I learned that rule through experience, for the fact that the rule has been learned through experience does not entail that it is plastic. Like the Sphex wasp, I have a rule that is unmodifiable and entirely insensitive to new contingencies or requirements should they ever arise. Also like the Sphex wasp, that rule shapes my behaviour appropriately within a given range of circumstances and is, within that range, flexible. Whenever I form the beliefs that both p and $p \supset q$ are true (and lack the belief *not-q*) I invariably conclude that q is true as well, no matter what the values of p or q. Put bluntly, I rather doubt that I could change (or learn to change) these flexible inferential behaviours if my very life depended on it. If the kind of flexibility at issue in intelligent or cognitive behaviour is flexibility in the rule itself, i.e., plasticity, we are forced into the uncomfortable situation of having to characterise unmodifiable rule-driven behaviour such as the use of modus ponens as distinctly *unplastic*, and hence, as unintelligent and non-cognitive. That would surely be a mistake. None of this changes the credentials of the use of such things as modus ponens as cognitive processes, only that their status as such does not stem from the plasticity of the rule that guides such behaviour.

Two responses to this criticism might be forthcoming, neither of which is particularly persuasive. A first response might be that the lack of flexibility in such cases is illusory on the grounds that it appeals to too strong a notion of flexibility – that is, to second-order flexibility or plasticity. Even though one's intelligent inferential behaviours might not be flexible in the sense that we fail to exhibit any plasticity concerning such inferential rules as *modus ponens*, and thus cannot help but obey them, yet it manifestly is flexible in that one varies one's conclusion according to the varying premises. This is, I imagine, generally true, and in that sense, we might be said to manifest flexibility in the face of varying circumstances. But as already noted, the problem with this is that it fails to clearly distinguish human intelligent behaviour from the Sphex wasp's allegedly non-intelligent behaviour. The wasp, too, shows flexibility of behaviour in essentially the same respect as humans following *modus ponens*: it, too, is following inflexible (that is unmodifiable and non-plastic) rules wherein its behaviour will vary according to variations in circumstances.

The second response might be to suggest that arithmetic and logic are somehow special cases that are highly unrepresentative inasmuch as arithmetical rules and logical inferences hold necessarily. Modification of behaviour in violation of such rules will always be inappropriate, it might be argued, and so it just never happens that we *need* to change these kinds of behaviours. Such a response is weak on two counts. Firstly, scenarios can be constructed where changes in such inferential behaviour would seem to be highly appropriate. Consider, for example, the gruesome circumstance of Winston Smith in George Orwell's *1984*, who, when undergoing torture in room 101, is faced with the prospect of having his face eaten out by rats if he does not genuinely believe (not merely say) that 2+2=5. Such circumstances would be highly unusual, no doubt, but then hardly more unusual than those of the Sphex wasp – normally Sphex wasps do not have to deal with paralysed crickets that change their location. Secondly, this objection seems to miss the point, for whether we possess the *capacity* to indefinitely change our behaviour in relation to unusual or novel circumstances is entirely orthogonal to whether we ever shall ever *need* to change our behaviour in such circumstances. Thus, if we do not disbar our arithmetical and logical inferences from counting as intelligent and cognitive on grounds of their lack of plasticity, we ought not to do so in the case of the Sphex wasp's lack of plasticity.

The third feature, learnability, is intimately tied to both plasticity and flexibility. It might be suggested that it is in the capacity for behavioural patterns to respond to experience – i.e., to be learned – that marks out a given behaviour as flexible or plastic, and hence as cognitive. Specifically, the claim might be that cognitive behaviour is behaviour that is (in part, at least) established by, and/or modifiable through, experience as opposed to being innate. Note that the appeal to behavioural patterns being a *response to* experience (or learned) is, at least in principle, to be distinguished from the issue of whether the rule guiding that behaviour is *indefinitely modifiable by* experience. There is no reason to suppose that an unlearned and entirely innate behavioural rule might not be modifiable. In the case of humans, at least, it is frequently suggested that innate rules governing human behaviour is heavily modified by cultural pressures and social contingencies. Conversely, as we have already noted, the rules guiding many learned behaviours are entirely unmodifiable and fixed once set. For example, learned responses to past psychological trauma (as, perhaps, in post-traumatic stress disorder) or the desensitisation to phonemes absent from the language heard in one's infancy. Both of these are at least highly resistant to further modification through experience. Moreover, an appeal to the origin of a behavioural pattern in a responsiveness to experience would controversially rule out a significant body of research that suggests that many a cognitive behaviour is innate rather than learned. There seems to be no *prima facie* reason why such research might not be well founded, and certainly no reason to think that such research has simply misunderstood what cognitive behaviour is in the making of its claims.

Why, then, is there a widespread identification of flexibility and plasticity of behaviour with intelligent or cognitive behaviour given that, sometimes at least, inappropriate, inflexible, innate, or unmodifiable behaviours *can* be cognitive? The answer, I believe, is that the usual *means* through which cognition operates (through being flexible, plastic, and modifiable by experience) has been mistaken for a *defining* characteristic that (in part) *makes* such behaviour cognitive. There is no doubt that flexibility, plasticity, and indefinite modifiability through experience are widespread and important features for the larger swathe of behaviours typically classified as intelligent or cognitive. As such, these features may function as very useful markers to identify cognitive behaviours as cognitive. But we ought not mistake such features as conceptually defining necessary and sufficient conditions of intelligent or cognitive behaviour. It is not so much the defining mark of intelligent or cognitive behaviour as simply the usual (perhaps almost ubiquitous) means by which intelligent or cognitive, goal-directed behaviour is rendered possible in a sometimes unpredictable world. The importance of flexibility and plasticity lies not in providing defining features of intelligent or cognitive behaviour, but in the fact that it is normally through it that we maximise our chances of keeping our behaviour appropriate to, and in line with, our functionally specifiable goals. Sometimes cognitive behaviours will be innate (for example, many aspects of perception, certain linguistic behaviours, and perhaps more controversially, certain basic inferential or arithmetical processes), especially with such behaviours as are (almost) always appropriate. In such cases, repeated learning of such behaviour with each new generation incurs a cost-ineffective demand on individual resources. Similarly, sometimes our cognitive behaviour will be or will become fixed and unmodifiable precisely because once set, change becomes (almost) always inappropriate.

§2.2.5 Cognitive Behaviour Defined Through Computationalism

The final approach to characterising cognitive behaviour that we shall consider approaches the issue from the direction of a specifically computationalist commitment to the nature of cognitive processing. With the shift towards a consideration of cognitive processing, this approach serves as bridge to the remainder of the chapter in which we consider the character of cognitive processing in its most general terms.

Following from Von Eckardt's point (§2.1.2) that the boundaries of scientific domains are modified and refined as our theories develop, it might be argued that once we are armed with an explanatorily cogent theory of cognitive processing, we can use this theory to then delimit the domain of cognitive behaviour according to whether the theory in question is required for the explanation of a given behaviour. For example, Zenon Pylyshyn, echoing Von Eckardt, suggests that:

> The set of phenomena that constitute cognition . . . is a long-term empirical question: we have no right to stipulate in advance which phenomena will succumb to the set of principles and mechanism that we develop in studying what appear pre-theoretically to be clear cases of cognition.[30]

In Pylyshyn's case – and his views are representative of many cognitive scientists – the 'principles and mechanisms' that determine the set of cognitive phenomena are emphatically (a critic might say 'dogmatically') computationalist (see Pylyshyn, 1980, §5). He attempts to clarify the question 'what are cognitive phenomena?' by a direct appeal to his theoretical commitment to a computational and representational account of the mind. For instance, in consideration of a number of phenomena that occur during, or vary with respect to, problem-solving tasks (such phenomena include reports of being frustrated, jotting down notes, variations in skin resistance, peripheral blood flow, and skin temperature) he suggests that the phenomena to be investigated as *cognitive* be determined (in essence) by answering:

> which, if any, of these observations could be explained by examining a canonical description of an algorithm, or, taking the position that algorithmic accountability defines the now technical notion "cognitive phenomena", one might ask which of the above reports represents an observation of a cognitive phenomenon? (Pylyshyn, 1980, p. 116)

This commitment to algorithmic (and hence computational) models of cognition is combined with a commitment to representationalism:

> it is clear that a cognitive model will not have values for blood flow, skin resistance, or temperature among its symbolic terms, because the symbols in this [i.e. computationalist] kind of model must designate mental representations (i.e., they must designate mental structures that have representational content – such as thoughts, goals, and beliefs). (Pylyshyn, 1980, p. 117)

[30] Pylyshyn (1991), pp. 189-90, quoted in Dawson (1998), p. 6. Note that Pylyshyn recognises that the principles and mechanism that cognitive scientists develop must address itself, in the first instance, to pre-theoretically characterised cases of cognitive behaviour. I return to this point in the next section.

Cognitive phenomena (or behaviours) are, on such an account, phenomena that have representational content that can be explained algorithmically.

Without disagreeing with the exclusion of the phenomena Pylyshyn mentions (blood flow, skin resistence, and temperature), or with whether Pylyshyn has accurately captured how most cognitive scientists circumscribe the domain of cognitive phenomena, there is a danger lurking in this approach. The danger (already alluded to in §2.1.2) is that the explananda – the phenomena to be considered in need of cognitive explanation – may be prejudicially or pre-emptively restricted to those phenomena having representational content, which is, in turn, a constraint derived from, and imposed by, the explanans – algorithmic (i.e., computationalist) models of cognition in Pylyshyn's case.[31] The worry is that such constraints on the domain of cognitive phenomena, grounded, as they are, in particular theoretical commitments, may, by dint of theoretical bias, illegitimately exclude from the explananda the possibility of there being cognitive phenomena that are not representational. There would surely be something amiss if a cognitive scientist of, say, a computationalist stripe could, by theoretical fiat, exclude from the class of cognitive phenomena putative non-representational counter-examples to their theory on the grounds that their theory requires that all cognitive phenomena be representational. As Stevan Harnad has pointedly (if somewhat rhetorically) commented: 'If it is in fact the case that some representational theory of mind has to be correct for cognitive science to exist, you have a very strange kind of science here. It depends on the truth of a certain class of theories. That seems to be a unique case' (quoted in Von Eckardt, 1993, fn. 3, p. 397). Of

[31] I do not necessarily wish to suggest that Pylyshyn is guilty of allowing his theoretical commitments to dictate the domain of cognitive phenomena. The Pylyshyn quote above suggests an awareness on his part that one's theories must be responsive, at least in the first instance, to pre-theoretic characterisations of the domain of enquiry.

course, particular *theories* within a science clearly do require the truth of certain claims, and are often so central to their respective sciences that it is easy to forget that they are theories and not the science itself. Chemistry, in its modern form, is unthinkable without there being atoms; biology, as it is generally practiced today, is hardly imaginable without the truth of evolutionary theory. Harnad's point, I take it, is that sciences *as such* do not (at least in general) require the truth of particular claims in order to exist. Chemistry *was* practiced long before atomism became universally accepted, and in certain quarters biology may be practiced without supposing the truth of evolutionary theory. The salient point, I suggest, is that whilst we can acknowledge the legitimate *influence* of explanatory theories on circumscribing a behavioural domain (we do, after all, wish to maximise the explanatory payoff of our theories), we ought to be very cautious before we allow particular theoretical commitments – even ones well motivated on empirical grounds – to *dictate* the domain of phenomena to be explained. The methodological danger is that in allowing certain theoretical commitments to rise to the status of 'a paradigm looking for a set of phenomena', as Michael Dawson (1998, p. 6) has put it, one may either artificially render the privileged theory immune from counter-example, or else risk the destruction of the discipline of cognitive science should it ever be forced to admit non-representational cases of cognitive behaviour.

What is important for the present issue, however, is that the danger of theory illegitimately dictating the domain of phenomena is precisely that – a danger, but not an inevitability. It is a danger that can be avoided, in large measure, by never losing sight of the fact that one's theoretical foundations are constituted by hypotheses and associated methodologies, not by *a priori* commitments. Such hypotheses and methodologies are (or at least ought to be) responsive to (though by no means dictated by) a behavioural domain that will likely continue to be, to some extent at least, pre-theoretically constituted. To return to the point of §2.1.2, as Von Eckardt suggests, scientific domains do not come ready-made and are shaped to a large degree by the success or otherwise of competing theories. Equally, however, competing theories address themselves to, and are competitively judged in terms of, their ability to explain a given domain of behaviour – a domain that is, in the first instance, *pre*-theoretically delineated. Thereafter, the shaping of explanatory theory and the domain of behaviour seems better characterised as an interdependent and complex process, with each affecting the other.

§2.3 – COGNITION AS INVOLVING INFORMATION PROCESSING

It seems, therefore, that leading attempts to characterise cognition via the class of cognitive behaviours provides us with, at best, some features that are useful for purposes of identification, but not with defining (necessary and sufficient) characteristics. We turn now to a consideration the nature of cognitive processing. Far and away the most frequently stated claim concerning the nature of cognitive processing is the claim that it involves (or sometimes simply that it *is* – there are various shades of commitment) the *processing of information*, or, in other words, that cognition involves operations ranging over information bearers – i.e., over *representations*.

In proposing information processing as central to cognitive processing three precautionary points need to be made. *Firstly*, the concept of 'information' can be read in at least two importantly distinct ways – what I shall characterise below as the 'truth-neutral' and the 'truth-dependent' senses. It is usually the former that plays the central role in the claim that cognition involves information processing.[32] *Secondly*, the concept of representation is here introduced in relatively impressionistic terms. Putting the finer details on the concept of representation as it features in cognition will be the central topic of chapter 3. *Thirdly*, no commitment is implied as to information processing being *exhaustive* of cognitive processes. The claim on the table is the weaker (and more ecumenical) one that *for at least a large number* of – but, perhaps, not necessarily *all* – cognitive behaviours, cognition is best understood as involving ('truth-neutral') information processing. To reiterate a point made at the beginning of the chapter, the thesis requires only that some strong basis be provided that will allow us to identify certain external processes or states as clearly cognitive, not that that basis be common to *all* cognitive processes or states.

§2.3.1 Information Processing

The most straightforward argument for the view that cognition involves information processing stems from the observation that cognition is sensitive to the *informational content* of a signal rather than its physical form. That is, that a system's cognitive behaviour can be radically and systematically varied by a wide range of conditions that need have nothing more in common than their informational content. Generalisations within cognitive science become possible, therefore, only by characterising cognition in terms of the processing of such content, which in turn requires information-bearers – i.e., *representations*. A classic example is provided by Pylyshyn (1980): seeing that the building one is in is on fire, smelling smoke coming in through the ventilation duct, being

[32] Whilst the distinction between truth-dependent and truth-neutral information is introduced below, a more detailed discussion of the truth-dependent sense of information and its role in cognition will be deferred to chapter 4, where I also distinguish a further dimension of information, what I call 'para-information'.

told by telephone that the building is on fire, or hearing someone shouting 'Fire!' all typically result in the same or similar behaviours (panic, running out of the building, etc.). Each of these environmental events (and many more besides) vary greatly with respect to the others in terms of their physical form, but they all carry the same *information* – in a 'truth-neutral' sense to be clarified below – that there is a fire in the building. In cognitive explanation, the story goes, what explains such similar behaviours (assuming similar background beliefs and desires) is not something that is physically common to the variations in pressure waves in air, the striking of electromagnetic waves on our retina or skin, or suchlike, but rather the common informational (i.e., representational) content that these signals possess.

As just noted, it is important to keep in mind that 'information' is used in at least two very different ways in the literature and with each playing distinct and important explanatory roles. Unfortunately, the two senses are often used without any attempt at disambiguation.[33] This is a result, no doubt, of the fact that states or structures may carry information in both senses of the term. Pylyshyn and most other advocates of the claim that cognition involves information processing are using the term in what I shall often call a *truth-neutral* sense. In the above example, when we seek to explain someone's running and panicking behaviour in the light of someone shouting 'Fire!', we appeal to the fact that the signal means (in Grice's (1957/1989) *non*-natural sense) that there is a fire in the building. This sense of information is truth-neutral inasmuch as the role it plays in explaining our cognitive behaviour – primarily, that of relating physically disparate but cognitively unified states to particular behaviours – is independent of whether the information it carries is correct. Whether or not there really is a fire on some particular occasion, it remains the informational content of such things as shouts of 'Fire!' that (in part) explains why we panic and run when we encounter such events.

The truth-neutral sense of information can be contrasted with that associated with the appearance of thick black smoke gushing from air vents. In such cases we say that the smoke carries the information that there is a fire – i.e., the signal or state means (in Grice's *natural* sense) that there is a fire in the building.[34] This sense of information, which derives its theoretical foundations from Shannon and Weaver (1949), extends well beyond its use in cognitive explanation. In the above example, the smoke can be said to carry the information (or, as some commentators prefer, 'to indicate') that there is a fire quite independently of any role that such information might subsequently play in explaining someone's running or panicking behaviour. In cognitive explanation the role of information in this sense lies primarily in the fact that it provides us with a ready-made account of how such information-bearing states can help shape our behaviour to so *successfully* or *appropriately* fit the world. They do so by reliably reflecting how things stand in the world, i.e., by carrying information in precisely this second sense. This sense of the term is evidently more intimately connected with truth, or at least to some reliable correlation between states, for if there were not (at least) some statistically significant and counterfactual-supporting relation between the appearance of thick black smoke and the presence of fires, then the smoke could not be said to carry the information that there was a fire in the vicinity. For this reason I shall often refer to it as information in the *truth-dependent* sense, although I shall also refer to it as the 'indicator' sense. Within this notion we may distinguish *strongly* from *weakly* truth-dependent varieties. The strongly truth-

[33] A fact noted and bemoaned in Dretske (1981), p. vii.

[34] Grice characterises natural meaning in terms of what I below call a strongly truth-dependent notion of information. That is, he insists that where *x* means (natural) *p*, that *x* entails *p*. Dretske follows Grice in this.

dependent variety requires that, to take the example above, thick black smoke *guarantees* the presence of a fire, whereas the weakly truth-dependent variety requires only that the presence of thick black smoke renders the presence of a fire more or less probable.[35]

As just noted, both of these senses of information play important (but distinct) roles in cognitive processes, and neither need be present at the same time as the other. Truth-dependent information can be carried without being represented, and representations need not carry truth-dependent information. In cognition, however, both kinds of information will *often* be carried in the same state or structure. Notwithstanding the important role of truth-dependent information in cognitive explanation, however, the claim that cognition involves information processing is *usually* intended as a claim concerning *truth-neutral* information processing; that is, concerning the processing of *representations*. This claim comes as close as anything to constituting a universally agreed upon characterisation of the nature of cognition.

It is here that the concerns raised in sections §2.2.1 and §2.2.5 are substantially addressed. In the first place it addresses the issue raised in §2.2.1 by providing a more substantive unity to the member disciplines of cognitive science; some portion of each member discipline within cognitive science may be said to either directly study or to make a contribution to our understanding of the processing of information. It is plausibly suggested that it is some such commitment to information processing that unites each of the disciplines under a single heading and which allows them to communicate with each other (see, for example, Dawson, 1998, p. 5). This view of cognition feeds through into a

[35] Similarly, I shall sometimes talk of their respective cognates as 'strong indication' and 'weak indication'. I.e., some state is a strong indication of some other state when it guarantees that the indicated state obtains; some state is weak indication of some other state when it renders that other state more or less probable.

characterisation of the mind as 'a complex system that receives, stores, retrieves, transforms, and transmits information' (Stillings *et al.*, 1987, p. 1).

Secondly, the near universality of the information processing paradigm goes a long way towards assuaging concerns that cognition is being artificially or parochially delineated by commitment to some particular theory or other. Pylyshyn's computationalist views canvassed in §2.2.5 represent one version of the commitment to information processing, but it is far from being the only one. Connectionist models and neurocomputational approaches, considered by many as rival theories to traditional computationalism, are no less committed to information processing as a basis for cognition. The concern expressed above by Harnad can be mitigated, therefore, to the extent that this claim is common to such a wide variety of differing and (sometimes, at least) competing theoretical frameworks. This claim need not be viewed as *a priori*. Neither need it be viewed as either dictating, or being dictated by, the character of its domain, but might be best expressed, to quote Von Eckardt (1993, p. 4) as 'a general working hypothesis about the character of its domain'. As such, it may bear analogy with the role of DNA or RNA in biological science. Having either DNA or RNA lies at the base of processes and states that are counted as 'biological', but however central, it need not be seen as providing an *a priori* condition for a phenomenon being biological.

§2.3.1 Information Processing and its Malcontents

Despite a very large measure of agreement that cognition involves information processing, there are some who might appear to take exception to this characterisation. The focus of the discontent lies with the putative role of representations. This dispute, however, is often rather more apparent than real and trades largely (though not entirely) on a failure to distinguish information processing in the above broad sense from more specific versions. In it broadest sense the claim entails commitment only to cognition involving appeal to representations and hence, to a '*representational* theory of the mind'. However, many of the chief advocates of this view, Pylyshyn included, incorporate additional elements into this claim including, most commonly, a commitment to computationalism – i.e., to the claim that the processing occurs along the lines of a digital computer; through formal or syntactic operations over discrete ('chunky') symbols. Those who endorse this stronger claim adopt a '*computational* theory of the mind'.[36]

The majority of those who might appear to object to the information processing model of cognition are in fact objecting only (or at least essentially) to the computational theory of the mind. Typical of such apparent critics is John Searle (1984 and 1992) who argues, in characteristically blunt style, that 'the brain does not do information processing'.[37] But a careful reading reveals that Searle does not come out against information processing in the general sense outlined above; his arguments are directed not against the claim that cognition involves information processing nor against the claim that it involves representations, but specifically against the claim that cognition involves processing information in the particular way that traditional computationalists conceive of it – i.e., as analogous to a digital computer. Indeed, Searle (1992, p. 246) admits that his arguments do not count against (and, he suggests, may even be an unwitting defence of) connectionist approaches to cognition. Yet such approaches are no less clear than computationalist ones that cognition involves information processing. Similarly, we have Rodney Brooks (1991a) insisting that we can have intelligent (and hence, cognitive)

[36] Pylyshyn and Fodor, amongst many others, take this view. This view is perhaps encouraged by the parallel characterisation of computers as 'information processors'. To Fodor's credit, he is careful to distinguish the broad claim from its more specific computational cousin.

behaviour 'without representations'. But a careful reading shows that he is specifically arguing against the claim that for intelligent behaviour we need *centralised* or *explicit* representations and *centralised* information processing. Once again, this is not to deny that cognition involves information processing, but only to deny that cognition always involves information processing over representations as computationalists conceive of them.

There are others who cast their criticisms more generally, however. Amongst those leading the apparent charge against *all* representational views of cognition are dynamic system theorists. For example, Thelen and Smith (1994, p. 338) insist that '[w]e are not building representations at all! Mind is activity in time . . . the real time of real physical causes'. Still, however, a close reading reveals that they object to the supposed necessity of certain *kinds* of representations – sometimes called 'objectivist' representations – namely, those that are static, detached, action-independent, highly detailed, and general purpose (see, for example, Clark, 1997b, p. 472). Moreover, dynamic systems theory remains officially neutral on the existence of representations, at least according to one of its chief proponents, Timothy van Gelder:

> [W]hile dynamical models are not *based* on transformations of representational structures, they allow plenty of room for representation. A wide variety of aspects of dynamical modes can be regarded as having a representational status: these include states, attractors, trajectories, bifurcations, and parameter settings.[38]

Once again, those who take a dynamic systems approach to cognition are more centrally concerned with the kinds of representations and processing that are effective instruments for the explanation of cognitive behaviour. Not all of cognitive behaviour, they typically insist, can be explicated through syntactic manipulation of symbols. Perhaps. But as

[37] Searle (1992), p. 222. Dreyfus (1972) makes similar claims.

Dobbyn and Stuart (2003) point out, state trajectories in some high-dimensional state space 'are not barred from being representational by not being symbolic', and they are often correct when they suggest that '[o]pponents of representational theories simply conceal a premise that asserts that the only possible form of representation is symbolic' (Dobbyn and Stuart, 2003, pp. 198-199).

Why, then, the continuing voices of dissent concerning representation and information processing? One possible answer stems from the growing appreciation over the past twenty or so years of the limitations of traditional computationalist models of cognition.[39] This has led many to reject computationalism in favour of different approaches such as connectionism and dynamic systems theory. Connectionists, with few exceptions, have not abandoned talk of representations or of cognition as involving information processing, and have not mistaken perceived inadequacies of computationalism with a necessity to reject all talk of representation or information processing.[40] At the same time, however, connectionists have re-interpreted these concepts in a variety of different ways without achieving a clear consensus on how these concepts should be understood.[41] For those similarly dissatisfied with connectionism, this has resulted in the general concepts of representation and information processing lacking a clear meaning. The computationalist understanding of these terms, meanwhile, has remained relatively clear-cut and intuitive – representations are physical symbols and information processing is the syntactic (formal) manipulation of these symbols. This relative conceptual clarity, when combined with the perceived inadequacies of

[38] Port and Van Gelder (1995), p. 12. For a representational interpretation of dynamical systems theory see, for example, Symons (2001).

[39] Leading the charge have been John Searle and Hubert Dreyfuss. See, especially, Dreyfuss (1972).

[40] Perhaps the most notable exception is Ramsey (1997), who argues that the distributed nature of connectionist 'representations' supports the anti-representationalist.

[41] See Van Gelder (1991) for a summary of these different views.

computationalism, may explain why many, in seeking to carve out new approaches distinct from both connectionism and computationalism, have taken arms under the banner of anti-representationalism but have focussed their attacks, in practice, on specifically computationalist accounts of representation and information processing.

None of the above is intended to show that there are not *some* cognitive processes for which appeal to representations may prove to be superfluous. But the truth of the matter, I would suggest, is obscured by a continuing lack of consensus and a lack of clarity amongst the disputants concerning the concepts of representation and the nature and role of information in cognition.

§2.4 – CONCLUSION

Having distinguished the theoretical question 'how does cognition operate?' from the conceptual question 'what is cognitive behaviour?', it was argued that answers to the first question rightly influence the domain of cognitive behaviour, but ought not to dictate that domain. It was further argued that attempts to characterise cognition by means of paradigmatic cases (or paradigmatic disciplines) fail to throw conceptual or theoretical light on the nature of cognitive behaviour. Attempts to provide a definition of cognitive behaviour in terms of intelligent behaviour prove to be inadequate where this is cashed out, as it usually is, in terms of flexibility and plasticity. However, although these features do not provide a satisfactory conceptually defining mark for cognitive behaviour, they do provide useful markers. Attention was turned to the theoretical question of how cognition operates, and in particular, to the claim that cognition is (or involves) the processing of information (i.e., the processing of representations). This claim was given a cautious endorsement (further defence of the vital role of representations in significant kinds of cognition will be given in the next chapter) but without claiming that *all* cognition must involve representational processing. It was also argued that confusions amongst cognitive scientists about precisely what is involved in this claim, together with related confusions concerning its sister concept of representation, have resulted in frequently artificial disputes and divisions. This led to the conclusion that greater clarity is needed both to round out the claim that cognition involves information processing, and to help resolve these (I have suggested) largely superficial disputes. Hence, in the following chapter, I shall detail the nature of representation as it figures in cognition.

CHAPTER 3 – COGNITIVE REPRESENTATION

The previous chapter concluded with a qualified and preliminary endorsement of the standard line that cognitive processing involves information processing – i.e., the processing of representations – but noted (hence the caution) a number of ambiguities and confusions that beset these notions. A significant task of the dissertation, therefore, and the central concern of the present chapter, is to lay out in as clear a manner as possible just exactly how 'representation' is to be understood in the context of its generally supposed role in cognition. I shall, therefore, begin (§3.1 – the bulk of the chapter) by identifying and discussing the central characteristics of such representations. It is hoped that the resulting characterisation will be sufficiently broad and ecumenical to cut across partisan divides, such as those mentioned towards the end of the previous chapter. Armed with this characterisation of cognitive representation and drawing partly on the arguments of Andy Clark, the second goal of this chapter (§3.2) will be to defend the claim that we need such representations to explain a significant body of our cognitive behaviour.

Three immediate points of clarification are in order. The first is that the central issue to be dealt with in this chapter concerns what it is for some *X* to be a cognitive representation, that is, with the *status* of something *as* a cognitive representation. I shall not be concerned with attempting to determine what it is for some *X* to be a representation *of Y*. The latter is the problem of representational *content* and continues to be the subject of intense debate in the literature without, it must be said, the emergence of anything approaching a clear consensus.

The second point of clarification is that my intention is to capture the general features of representations *only* in the context of their supposed role in cognition. Notwithstanding that some of the features I shall identify apply to representations outside of the context of cognition, I am not attempting to characterise the nature of representation in its full generality or even of *all* things that are often or typically called 'representations'. The goal here is to identify the characteristic features of cognitive representations that enable them to play the kind of roles that most cognitive scientists suppose them to play. With that firmly in mind and in order to avoid unnecessary prolixity, I shall not be overly concerned to repeatedly qualify the term 'representation' with 'cognitive' in the remainder of this chapter or, indeed, throughout the rest of the dissertation.

A third point of note is that the ensuing analysis is intended as an elucidation of the central features of cognitive representations, not as a reductive analysis or as a set of necessary and (jointly) sufficient conditions. I am not seeking to *explain* cognitive representation in non-intentional terms, but instead to *describe* cognitive representation in its broadest terms, and in a way that will help us, ultimately, to identify putative examples of such in the external world.

§3.1 – REPRESENTATIONS

There is a considerable variety in the kinds of things that go by the tag 'representation': a road map, a graph, a sentence, a gesture, a diamond ring, a drawing, a photograph, a musical score, a totem pole, a petrol gauge, a constellation of stars, a footprint – the variety seems almost without limit. Such representational states or structures may be distributed over space (such as written sentences) or time (such as spoken sentences or musical performances), constructed by deliberate fashioning (such as totem poles) or merely chanced upon ready formed (such as constellations of stars representing mythical characters or stories), they may be dynamic or static, material or temporal, rooted in biology, culture or individual idiosyncrasy, and so on. What unites such disparate examples under a common and coherent conceptual umbrella?

§3.1.1 Representation and Intentionality

The most general characteristic of representations (whether specifically cognitive or otherwise) is that they each exhibit *intentionality* or *aboutness*. Indeed, exhibiting intentionality is synonymous with representing: *X* is a representation if and only if *X* is *about* (or *stands for*) some *Y*.[42] There are, of course, other uses of the verb 'to represent' that lack this feature. Anne Jaap Jacobson (2003, p. 190), for example, gives us a putative example of something that represents without being about anything: an elected official who, she suggests, 'may represent a district without being about the district. Here clearly the marks of the intentional do not apply.' Be this as it may, such examples are not in any obvious way representations in any sense that concerns us here. Indeed, it is more natural to speak of elected officials and the like as 'representat*ives*' rather than 'represent*ations*'. It may be, of course, that some of the representative's *behaviour* – such as their voting behaviour – is used as a representation of (and hence, in the relevant intentional sense, is *about*) the voting of those whom the representative represents.[43] Outside of such behaviours the marks of the cognitive (see chapter 2) do not apply, for no one is using the

[42] Note that 'stand for' is not the same as 'stand *in* for', as the soccer substitute example in the next section demonstrates. For sake of simplicity, I leave the range and nature of X and Y deliberately vague for the moment. As will become clearer below (§3.1.3), I take it that it is certain *properties* of objects not the object *simpliciter*, that carry the cognitive representational load, so to speak.

representative to engage in information processing and therefore such examples fail to demonstrate representation without the marks of the intentional.

More interesting putative candidates for non-intentional representation that Jacobson offers include the use of fabric swatches and other such *exemplars* – what she calls 'token-realizations' (Jacobson, 2003, p. 191). For instance, when asked what kind of cat one likes, one might point to an instance of that kind and say 'these'. Jacobson insists that such token-realizations are not intentional in that they are not *about* what they represent. Rather, she insists, 'they are *of* their type in the sense of being instances of their type, but not in a semantic sense of "of." That is, token-realizations are not of their type in the sense of *meaning* their type' (*ibid.*). But, so far as I can discern, she provides no good reason to think that a token of a given type cannot be both 'of' its type in the sense of being a member of its type, *and* a representation of its type, in the sense of being about that type.

§3.1.2 Truth-Neutral Informational Standing-In-For

Closely tied to the claim that representations are intentional is the claim that they 'stand in for' something else. So central is the idea of 'standing-in-for' to our understanding of representation that when speaking loosely we might almost use it synonymously with representation and it is a universally cited characteristic feature of representations that they can be said to stand in for something that they are about. But we need to be clear exactly in what manner representations stand in for what they represent. In its most general terms, standing-in-for is a broad functional notion: something stands in for something else when it fulfils a functional role of whatever it replaces. Consider, for

[43] A similar example may be when A welcomes B *on behalf of* C. It is possible, of course, for B to use A's behaviour as a representation of C's regard for B, but it would be precisely for this reason that we might then say that A's behaviour is about C's regard for B.

example, a substitute soccer player *A* substituting for an injured player *B*. It seems entirely appropriate to say that *A* 'stands in for' *B*, at least to the extent that *A* fulfils a functional role of *B* within the team. But it would be very odd to say that *A* 'represents' *B*. Here, 'standing-in-for' amounts only to playing a functional role within the team that is (in most respects or at least typically) adequately similar to the player who was replaced. Contrast this with, say, some person *R* standing in for a group of people *S* at a public meeting. Here, by contrast, it seems very natural to say that *R* represents *S*. In what lies the difference between *A* and *R*? The salient difference, I suggest, is that unlike the properties or behaviours of soccer substitutes and the like, representations (including, where appropriate, the behaviour of representatives at a meeting) function as truth-neutral *informational* stand-ins. Another way to put this point is to say that representations stand in for other things by *saying* something about them. There is little, in general, that we would be led to infer about the replaced soccer player from an examination of the behaviour or properties of the substitute player. By contrast, the behaviour or properties of the representative at the meeting is used by those present at the meeting as an informational stand-in for the group that the representative represents; the representative's actions (such as their expressed views or their voting behaviour) say something about the group of individuals for whom he or she stands-in.[44]

This requirement is of a piece with the view of the preceding chapter that cognition involves information processing in the truth-neutral sense of information. It is also a feature that helps to distinguish such representations from entities that merely *refer*. On

[44] This is not to say that a soccer substitute *cannot* stand in for another player in an informational role. For example, that the substitutes plays as a winger, say, may allow us to infer (at least with some confidence) that the player substituted was a winger, insofar as it is generally the case that substitutes replace the positional role of the player substituted. Neither is it to say a representation necessarily carries information in the truth-dependent sense of the term.

standard accounts, names, for example, merely refer. They are, therefore, about things (hence, they satisfy the intentionality condition above) but they carry no truth-neutral information – they do not *say* anything. Names (or any such cognitive equivalent to names) do not, therefore, count as cognitive representations in the sense that concerns us here, although, of course, they may be constitutive *parts* of such representations and make a significant contribution to things that are said or to information conveyed.[45]

§3.1.3 Properties Represent

It is entirely commonplace to think or say that objects, *per se*, represent other objects. Few things might seem more innocuous than the claim that a statue, say, is a representation of a particular individual. But the foregoing emphasis on the importance of the truth-neutral informational role of a representation should alert us to the fact that this hides a potential confusion, for it is always some *property* of an object that carries the representational load (so to speak) in cognition and not the object *per se*. Another way to put this point is that objects do not represent other than in virtue of some property or structure, be that property or structure internal to the object in question or in its relations with other objects. To count as a representation in the sense that matters in cognition – that is, to carry truth-neutral information – an object must represent always in some respect or other – through its shape, its size, its colour, through an ordering of its parts, through its position in relation to other objects, through it being present or absent, etc.

For example, consider the sentence 'Fido likes Sue'. This sentence, considered as the representational object in question, has an internal structure that makes a direct

[45] It is quite commonplace to refer to a name as a 'representation' of whatever it is to which it refers. I remind the reader of the qualification made at the beginning of the chapter concerning my use of 'representation'. See also §3.1.4.

contribution to it carrying truth-neutral information. Change the structure (to, say, 'Sue likes Fido') and different truth-neutral information is carried. Remove that structure entirely – or remove the structure-providing elements (spaces or punctuation in the case of English sentences, as in 'FidolikesSue') and no truth-neutral information is carried. If, by contrast, we remove 'Fido' from a structured sentence it is clear that it lacks a representational structure of its own to carry information; although, of course, it clearly has a structure (consisting of four letters ordered 'F', 'i', 'd', and 'o'), that structure is entirely arbitrary with respect to the properties of Fido and contributes nothing to its representational capacity. Nothing is said by 'Fido' and no truth-neutral information is carried by 'Fido'.

Similarly, neither does merely pointing to a coin and stipulating that it represents Napoleon's forces at Waterloo suffice to render the coin, in the sense that concerns us here, a representation of Napoleon's force's at Waterloo. We must always ask: which of the many features of this coin represents which of the many features of Napoleon's forces at Waterloo?[46] Only thereby will it be able to *say* something *about* Napoleon's forces (as was required of cognitive representations in the previous section). Without the specification of some property of the coin that will serve the representational function – for example, its spatial relation to the pepper pot and saltcellar – the coin is incapable of playing a representational role. A more general way to express this is to say that some object *A* will represent some object *B* only in virtue of some property of *A* representing some property of *B*. In many (perhaps most) cases the relevant property is left entirely implicit. In the case of statues, for example, the relevant property is one of approximate

[46] This is not, of course, to say that representation requires resemblance.

spatial congruence; in the case of words, it is a case of the position of the word in relation to other words, the context of use, and so on.[47]

§3.1.4 Isomorphism and Interpretation

It is a frequently heard claim that representations must be *isomorphic* with what they represent. As we shall see, what is important is that there be some *determinate* (as in *set* or *specified*) isomorphism. Expressed somewhat technically, an isomorphism is a one-to-one mapping relation between two sets, in each of which an operation is defined such that the mapping relation preserves the operation. Less technically, an isomorphism occurs when there is a mapping relation that maps the structure of the one onto the other.

An obvious looking exception to the claim that cognitive representations must be isomorphic with what they represent might be 'unstructured' representations, quintessentially names. For example, it might be suggested that 'Fido' represents Fido, but without being in any way isomorphic to Fido, since 'Fido' is, qua representation of Fido, unstructured. But as already noted, whilst names may play an often important role *in* cognitive representations they do not (in the sense of that term that concerns us here) themselves *constitute* such representations. The name 'Fido' only acquires a capacity to play a representational role in cognition when it appears in a structured context, such as in sentences and in language use more generally.[48] Note: this is not to say that 'Fido' is not a representation in a perfectly legitimate and natural sense of the term. 'Fido' very clearly does represent Fido – in the specific sense that it refers to Fido. But as stated at the outset

[47] This might be seen as a rough parallel of Frege's (1884/1959) context principle: never seek the meaning of a word outside of the context of a proposition. I might say: never seek the cognitive representational character of an object outside the context of a property.

[48] Nothing is being claimed here concerning whether all such structure is propositional. It is simply that names (such as 'Fido') typically figure in propositionally structured objects. As suggested earlier, spatial relations, colour relations, etc., are also capable of providing the necessary structured context.

of this chapter, I am not attempting to capture how the word is used in all its uses. The point is that nothing is *said about* Fido (or about anything else, for that matter) by 'Fido', and it is the capacity of cognitive representations to *say* something, to be *about* something, that enables them to play a significant role in cognition.

Recall the point made earlier that no truth-neutral information about Fido is (or can be) imparted merely through the stipulation that 'Fido' shall stand for Fido; no truth-neutral information about Fido can be gleaned merely from the examination or manipulation of 'Fido'. Sans its appearance in a more structured context, 'Fido' can carry no truth-neutral information, and so cannot be the subject of information processing. By contrast, 'Fido is hungry' can represent something, in part because, under appropriate linguistic conventions, it is isomorphic to what it represents. It may not escape notice, of course, that sometimes names can be used outside sentences to carry truth-neutral information. For example, one might draw up a list of objects in a room and write 'Fido' on the list. Here again, however, the appearance of 'Fido' *on the list* constitutes the structured context wherein 'Fido' acquires its representational capacity (in the sense of representation that concerns us). The appearance of 'Fido' on the list is, in such a context, isomorphic to Fido being in the room.

That there will be *an* isomorphism between representation and represented is an *extremely* weak criterion – far too weak to help distinguish representations from non-representations – for it is trivially satisfied for any two concrete objects. At least in principle, pretty much *anything* can be said to be isomorphic with pretty much *anything* else, just so long as we carve up the systems into an appropriate number of elements and choose an appropriate transformation or mapping relation between their elements. To take an extreme example, a grain of sand may be said to be isomorphic to the universe, given a certain assignment of regions within the grain to elementary particles (or what have you) under an appropriately chosen mapping relation.[49] The only general exception to such isomorphic 'freedom' concerns abstract objects, most notably mathematical structures. This is because the number of elements and the structure of abstract objects is either implicitly or explicitly built into them, constraining their ability to stand in isomorphic relations with other structures. For example, it is impossible to use the rational numbers to represent the real numbers because their respective structures ensure that there are not enough elements in the former to map onto the latter. Similarly, for example, a two-dimensional vector space cannot fully represent a three-dimensional phenomenal colour space consisting of hue, saturation, and luminosity. But for concrete structures, there remains no *in principle* cardinality constraint (since any portion of an object can be notionally interpreted, for example, as a segment of the real number line). Whether there is a constraint on the richness of structure of a concrete object is harder to say, but it hangs on whether there is a constraint on the number of properties that the elements of a concrete object can be said to possess. In *practice*, of course, there are numerous further constraints on the effective utility of using various structures as representations of other domains. At a minimum, the isomorphism between representation and represented must be of a kind usable by the cognitive agent in order for an object to function as a representation.

It should be clear, therefore, that for a system to function as a representation, some *particular* isomorphism (from amongst the innumerable that are available) must be 'set', or 'selected' or (as I shall usually put it) rendered *determinate*. This is often expressed by

[49] Perhaps it may be objected that even if the regions of the grain of sand were isomorphic at some instant under a certain interpretation, since the universe is in a continual state of flux, and since the grain of sand (let us suppose) is static, it is not isomorphic with the universe over time. But this is easily side-stepped by simply stipulating that there be a second-order function determining a different transformation or mapping relation for each instant.

saying that isomorphism for a representation is always '*under an interpretation*'. Any process whereby an isomorphic mapping relation between distinguished elements of two systems is rendered determinate I shall call an *interpretational assignment*, or simply, an *interpretation.*[50]

Commentators vary considerably on precisely how this is done. Information-theoretic accounts (using 'information' in the truth-dependent sense) such as those given by Fred Dretske (1981) and Jerry Fodor (1990a), emphasise a nomic or counterfactual-supporting isomorphism between physical systems, whilst teleological accounts, such as those given by Ruth Millikan (1984) and David Papineau (1987), emphasise a functional isomorphism – that is, they emphasise the function of a representation to be isomorphic with what it represents.

Stepping above such differences, the basic character of interpretation involves, in the first instance, a referential assignment of elements of a representation to elements of the represented. Secondly (and this is generally less well appreciated) it is in interpretation that the representation is divided into a determinate set of elements and *assigned* a determinate structure, qua representation. As John Heil (1981, p. 335), drawing on Wittgenstein, points out, '[t]he structure of states of affairs is given, natural. The structure of signs, however, *qua* signs, is something that we must determine [i.e., fix or bring into being].' This is not to deny that there are innumerable structures present in any concrete object, of course. The point Heil is making, I take it, is that whatever natural states of affairs there may be, and whatever natural structure there may be, those structures are not determined, in a metaphysical sense, *by us*. For example, there are innumerable natural structures present in, say, a cat being on a mat, but all of those structures are presumably

[50] I make no claim that my usage of 'interpretation' follows that of others.

objective. It is presumably a natural and given fact that there is a cat on a mat and not a mat on a cat. Leaving aside who or what does the interpretation, the central point is that representation requires interpretation not just to assign elements to elements via a mapping relation, but also to distinguish the elements and structure of the representation. The structure of a representation is something that we assign to it, privileging one structure from amongst its innumerable available structures.

We can characterise an interpretation, therefore, as involving three distinct steps: 1) the division of the representation into distinct elements, 2) the determination (i.e., fixing or selection) of some property of the representation to be used in mapping the elements of the representation to elements of the represented, and 3) the mapping of those elements to elements of the target domain.

Consider, for example, a simple stick figure:

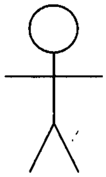

Figure 3.1: A Simple Stick Figure

Firstly, it is a matter of interpretation that the figure, qua representation, is divided into such elements as a circle, a vertical line, two horizontal lines, and two angled lines; the figure is no more objectively or logically structured into such spatially related line elements than into, say, point elements related according to the sequence in which they were laid down. *Secondly*, having determined the set of elements thus, it is a matter of interpretation that they represent in virtue of their spatial relations, rather than, say, in virtue of the sequential order or colour in which they were drawn.[51] Note that the first aspect – the distinction of the representation into denoting elements – does not suffice to determine the *structure* or 'syntax' of the representation. For it remains to be determined which of the many relations in which such distinguished elements stand is to be privileged as determining or constituting its representational structure. The structure might, for example, be determined by the circle's and lines' approximate spatial relations within a Euclidean plane (as is conventionally the case), but equally it might be the sequence in which the elements were laid upon the page, the thickness of the lines, or any other structural relation in which the elements happen to stand. *Lastly*, we have the aspect most commonly associated with the idea of an interpretation – the determination of which elements of the figure are assigned to which elements of the represented. For example, the left vertical line might be assigned to the right arm or to the head, etc.

§3.1.5 Misrepresentation or Error

Another feature often taken to be central to cognitive representation is a capacity of representational states or structures to *misrepresent* – the power to say or represent the world as *P* whether or not *P* is the case. Dretske, for example, suggests that:

> The ability to correctly represent how things stand elsewhere in the world *is* the ability of primary value, of course, but this value adheres to representations only insofar as the representation in question is the sort of thing that *can* get things wrong. In the game of representation, the game of 'saying', telling the truth isn't a virtue if you *cannot* lie. (Dretske, 1988, p. 65)

One clear motivation for believing that representational states or structures will be capable of misrepresentation comes from the central role that they play in much of our more

[51] Searle (1992, p. 207ff) makes a similar point when he argues that syntax is not intrinsic to physical structures, but is instead a matter of assigned interpretation. In its way, this is a parallel of Frege's insight (1884/1959, §22ff) that the cardinality of something is indeterminate until it is placed under a concept.

sophisticated cognitive behaviour. Such a misrepresentational capacity is central in imaginative cognition, including hypothetical or counterfactual reasoning. As Dretske points out, without such a capacity to misrepresent how the world stands a system's cognitive powers will be considerably impoverished, having, at best, the ability to say or 'represent' what *is*, but not what is not. The capacity for such cognitive behaviour distinguishes human beings as being amongst what Dennett describes as 'Popperian systems'; i.e., as having evolved the capability to pre-select from amongst possible behaviours or actions through the use of representations.[52]

This, however, only gives us grounds to say that many representations – especially those found in such cognitively sophisticated and powerful systems as ourselves – will be capable of misrepresentation. It may be somewhat less clear exactly why (or indeed whether) the manifest cognitive power that the capacity to misrepresent clearly generates should provide us with an entirely general feature of cognitive representations. In other words, it is one thing to point out, as Dretske does, that all cognitively *powerful* systems will make significant use of representational states or structures with the capacity to misrepresent, but quite another to suggest that *all* such states or structures (even those of the least powerful cognitive systems) need exhibit the same capacity. On the face of it, a state or structure nomically incapable of representational error under a certain interpretation might be no less a representation for all that.

Likewise, the structure of something is indeterminate until placed under an interpretation via the assignment of a structuring property.

[52] Dennett (1996), ch.4. The term arises from Popper's claim such a capacity allows our hypotheses to die in our stead. John Haugeland presents the flip side of this point: 'If the relevant features are reliably present and manifest to the system (via some signal) whenever the adjustments [within or by the system] must be made, then they need not be represented. Thus, plants that track the sun with their leaves needn't represent it or its position, because the tracking can be guided directly by the sun itself'. ('Representational Genera', in Haugeland (1998), p. 172).

A more central (and more general) ground for the claim that all representations will be capable of misrepresenting falls directly out of the indetermination of isomorphism between representation and represented prior to interpretation. Given that there must be an interpretational assignment to fix the isomorphism relation, then it will always be possible for the representation to misrepresent how the world stands. Until placed under an interpretation, the isomorphism relation between representation and represented remains an open issue, hence the possibility of misrepresentation will always exist; there are always possible interpretational assignments that result in the representation *mis*representing the represented.[53]

It is important to emphasise that the capacity of a representational state or structure to misrepresent – to 'get things wrong' – arises at the point of interpretation and directly out of the indeterminacy of representational structure prior to interpretational assignment. *After* interpretation a state or structure might be incapable (for either nomic or logical reasons) of misrepresenting. For instance, *if* one makes an interpretation of a state or structure such that the resulting representation expresses a *necessary* truth, then, for obvious reasons, that representation cannot misrepresent how the world stands.[54] The point remains, however, that under a *different* available interpretation, the very same state or structure might go on to misrepresent how the world stands. It is not that every representation, *given* an interpretation, will be capable of getting it wrong, but that every representational state or structure accommodates an interpretation that renders it capable of misrepresenting how the world stands.

[53] Misrepresentation has beleaguered causal and informational accounts of representation, for if something represents in virtue of co-varying with the represented as a result of nomic-causal relations, the capacity for the 'representation' to misrepresent appears to be eliminated. Dretske (especially 1986 and 1988) has attempted to work his way around thing by incorporating teleology into a nomic-causal framework.

[54] My thanks to Alex Byrne for raising this point.

§3.1.6 Preliminary Conclusions

We have seen that for our purposes a representational state or structure, as it features in cognition, is something that:

a) Is *about* something, in something like the standard intentional sense.

b) *Stands in for* something, but in a way that it plays some truth-neutral *informational* role.

c) Represents through and in virtue of its *properties*.

d) Is, under an interpretation, *isomorphic* in some determinate way or other, with what it represents or whatever it has the function to represent.

e) Requires an *interpretation* to determine the isomorphism from amongst an indeterminate number of possible candidates.

f) Is (under some interpretation) capable is *misrepresenting* how the world stands.

Thus far, the role that representations may play in cognition or, indeed, whether they play a significant role at all, has not been addressed. Armed now with an outline characterisation of cognitive representation, in the next section I shall put that claim to the test and examine the case for such representations in cognitive explanation.

§3.2 – WHY WE NEED REPRESENTATIONS

The idea that cognition involves representing the world in some way is intuitive, powerful, and (almost) all-pervasive amongst cognitive scientists. Towards the end of the previous chapter apparent dissent against this view was characterised largely as a misdirected attack on a particular conception of representation – i.e., the computationalist idea of representation as syntactic manipulation of symbols analogous to the information processing of a digital computer – rather than an attack on the idea that the cognition involves representation *per se*. Thus, it was argued, the dissent was at least largely the result of mistaking computationalism – a particular, if venerable, version of representationalism – for representationalism *simpliciter*. In this section, armed with the ecumenical characterisation of representation of the previous section, I shall defend an equally ecumenical version of representationalism – that is, a representationalism that remains neutral on questions concerning, for example, whether cognitive representations are 'chunky' (as characterised by traditional computationalists) or distributed (as for the clear majority of connectionists who see no reason to dispense with representations). Moreover, it is a weak version of representationalism that will be defended – i.e., that representations are needed to explain at least *some* cognitive processes – in contrast with strong representationalism that insists that *all* cognition involves representations. This is in line with the characterisation of cognition given in the previous chapter, that *for at least a large number* of – but, perhaps, not necessarily *all* – cognitive behaviours, cognition is best understood as ('truth-neutral') information processing. The view to be defended is, therefore, entirely consistent with arguments from several quarters that there are some cognitive processes that do not require representational explanation.

§3.2.1 The Dynamicist Critique

Dynamical systems theory has been at the forefront of the anti-representation movement in recent years. In essence, a dynamical system is simply any system for which we can give a rigorous analysis of how it evolves over time. This requires nothing more than a finite number of variables or magnitudes that capture the state of the system, plus a set of equations that describe the evolution of those variables over time through a 'state-space'. Thus, trivially, all actual computational systems fall within the class of dynamical systems (though not all dynamical systems will be computational). There are, however, three variously emphasised but closely allied features that help to distinguish between dynamical systems models of cognition and, especially, traditional computational models.

The first of these is that dynamicist models emphasise the 'real-time' temporal evolution of cognitive processes. They directly draw on and emphasise the fact that '[c]ognitive processes and their context unfold continuously and simultaneously in real time'.[55] Our cognitive models, they argue, should directly reflect this fact by drawing on the tools and concepts of dynamic systems theory; tools such as the modelling of behaviour as a temporally evolving system via state-variables and differential equations), and concepts such as 'coupling', 'attractors', and 'repellors'. Traditional computational processes, by contrast, are essentially *a*temporal: whilst all realised computational processes obviously take place *in* time, they are described and modelled algorithmically as a sequence of discrete events or states. Turing machines, for example, are characterised by rules that specify state transitions, but such transitions are modelled without any reference to them having temporal extension. By characterising cognitive processes in such terms, they abstract away from the temporal features that accompany all such realisations – features that may (dynamicists argue, do) play a role in real world cognitive processes.[56]

The second feature of dynamicist models is that the elements of the model are often 'tightly coupled'. That is, the behaviour of various elements in the models often exhibit a dynamically complex and reciprocal evolution over time. For example, perceptual input to the system may be modelled as directly and continuously influencing behavioural output

[55] Port and Van Gelder (1995), p. 3. The quote is taken from the introductory chapter appositely titled 'It's About Time: and Overview of the Dynamical Approach to Cognition'.

which in turn directly and continuously influences perceptual input (and so on). This, too, contrasts with computational models wherein processes are defined via computationally discrete cycles of input-process-output. This is not to say that tightly coupled processes *cannot* be modelled computationally, but rather that the tools and methodology of dynamicist models capture such tightly coupled relationships directly and perspicuously.

The third feature is closely related to the preceding one. Unlike traditional computationalism, tight coupling can occur between elements that transcend traditional cognitive boundaries of brain or skin. Thus, no fundamental or theoretically important boundary is recognised in the ultimate input-output boundaries of traditional computationalism: the sensory periphery (input) and physical behaviour (output). As a result, dynamicist models of cognition frequently take the cognitive system to comprise not only the brain (together with sensory and motor systems) but also elements of the 'external' environment.

To see these features more clearly, consider the following analogy to dynamicist models of cognition: that of a 'Watt governor' (Van Gelder [1998]) – a mechanical device for smoothing (i.e., regulating) the power output of a steam engine through a direct and reciprocal connection between the behaviour of its sub-systemic elements: a fly-wheel, a valve controlling steam pressure, and a rotating centrifugal spindle (the 'governor'). The Watt governor's sub-systemic elements occupy separate spatial regions, but their tightly-coupled causal influences on each other result in a dynamic and reciprocal evolution of each over time that defies analysis in terms of iterated cycles of input-process-output of each sub-systemic element. Instead, the only (or at least, the best) way to describe the

[56] Perhaps the best philosophical defence of this position is Van Gelder (1998), to which empirical flesh has been added through the likes of Beer (1995), and the collection of papers in Port and Van Gelder (1995).

behaviour of the Watt governor is as a systemic whole – as a complex dynamically evolving system; in other words, as a dynamical system.

The key feature of analogous models of cognition is that the dynamic real-time reciprocal relations between the sub-systemic elements (traditionally divided by computationalists into, at its broadest, sensory input, in-the-head computational process, and behavioural output) necessitates that the description of behaviour be given in terms of the system as a whole, including (often) relevant features of the environment, and not in terms of discrete operations carried out computationally by sub-systemic elements. This results in the rejection of both causally and cognitively significant boundaries between notionally 'internal' and 'external' sub-systems, which in turn has been taken by some to eliminate the need for representation-bound computational explanations of cognition. Drawing these threads together, Port and Van Gelder argue that:

> 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 not a discrete sequential manipulator of static representational structures; rather, it is a structure of mutually and simultaneously influencing *change*. (Port and Van Gelder, 1995, p. 4)

§3.2.2 The Critique Deflated

The merits or otherwise of the dynamic systems approach to cognition are not our direct concern here. What is of concern is the use made of such models to criticise representationalism. Such criticisms can be deflated both negatively, by arguing that neither their arguments nor their research establish the conclusions that they (often, at least) take them to establish regarding the putative role of representations in cognition, and positively, by arguing (independently of any weaknesses in dynamicist anti-representationalist arguments) that we need representations to explain significant portions of cognitive behaviour. I take each in turn.

Dynamicists seeking to further the cause of anti-representationalism often cite a growing body of research involving dynamicist analyses of such cognitive behaviour as finger-wiggling, ant walking, and infant tread-mill walking to name but a few.[57] We have already had cause to note (§2.1.2) that such behaviours as finger-wiggling and other motor behaviours are, at best, controversial candidates for being cognitive behaviour. Moreover, leading dynamicists note the official neutrality of dynamicist approaches on issues of representationalism. Even putting such issues aside, the putative success of dynamicist analyses of such limited and relatively simple portions of cognitive behaviour as are standardly cited would show, at most, that we do not need representations for *those* portions of cognitive behaviour. Such research would therefore show, at most, that the strong claim that *all* cognition is representational is mistaken. It remains difficult, however, to see why such research should lead one to reject the more modest claim that representations are absolutely central to a significant portion of cognition – portions such as inference and deduction – for which dynamicist research remains notably thin on the ground.

The second and third strands of dynamicist models – that the unfolding of a cognitive process over time is a feature from which we (sometimes, at least) ought not to abstract, and that traditional cognitive boundaries of brain or skin constitute no theoretically significant barriers – can be conceded without concern. Let us allow the dynamicists their due, for it may well be that a persuasive case can be made that temporal relations play a hitherto under-appreciated role in certain cognitive processes. Similar dues can be paid regarding the importance of embodiment in many cognitive processes (especially, and perhaps unsurprisingly, with respect to motor behaviour involving reciprocal feedback), and, more generally, to their claim that cognitive processes typically involve elements that extend beyond traditional cognitive boundaries of brain or skin. But none of this demonstrates that representation can or should be dispensed with, either quite generally, or even in particular cases where the processes have significant temporal components or that transgress traditional cognitive boundaries. For instance, it has been long argued that an incubation period (in which the individual cognitive agent stops thinking consciously about a problem) has a significant positive effect on problem-solving performance, yet this has not impeded representational models that incorporate such temporal components amongst their elements.[58] Equally, representational accounts have been offered for embodied cognitive processes[59] and the central examples that will occupy us in the later chapters of this dissertation will likewise be characterised representationally but cut across traditional boundaries of brain or skin.

Turning now to some positive grounds for (weak) representationalism there are several arguments for keeping representations amongst our cognitive explanatory armoury. One such argument points out that the distinctively human cognitive processes – ones that mark us out as probably the most cognitively accomplished of species – are precisely those that cry out for representational explanation. David Kirsh (1991) persuasively argues that claims that 97% of behaviour is non-representational, whether accurate or not, ignore the fact that it is the behaviour that remains that makes all the difference to human intelligence. It is the remaining 3% (or whatever figure we pick), and

[57] Kelso (1995), Beer (1995), and Thelen and Smith (1994) respectively. Additional examples can be found in Port and Van Gelder (1995).
[58] See, for example, Silveira (1971) and Yaniv, et al. (1995).

our consequent abilities to deal in representations and engage in conceptual thought, that sets us cognitively apart from other species.[60] Similarly, Clark and Toribio (1994), have argued that there remain crucially important domains of cognitive behaviour that show little prospect of the explanatory elimination of representations.[61] The most plausible or successful anti-representational accounts of cognition have been for 'online' behaviours where the cognitive agent is dynamically interacting with its environment rather than engaging in a process suitably characterised as 'offline' reasoning or thought. Such 'representation-hungry' cognitive processes (to borrow a phrase of Clark's) involve reasoning about absent, non-existent, or counterfactual states of affairs. It seems highly implausible, to say that least, that we might dynamically interact, as dynamicists might suggest, with objects of features that aren't there. The second kind of representation-hungry cognitive process involves what Clark and Toribio (1994, p. 419) describe as the selective sensitivity 'to parameters whose ambient physical manifestations are complex and unruly' leading to a lack of perceptual salience (see also Clark, 1997a, pp. 167-8). Typical cognitive processes of this kind may involve properties at a high level of abstraction or some form of conceptual thought – a feature of representational cognition also emphasised by Kirsh (1991) – such as identifying those objects that belong to the Pope. Both of these kinds of cognitive processes show little prospect of usefully eliminating representation from their explanation in favour of representation-free dynamic interaction. Insect-like mobots making their way through an obstacle strewn environment (Beers [1994]), or retrieving and disposing of soft drink cans in a 'messy' and dynamically

[59] For a connectionist style example, see Bechtel (1997).

[60] With reference to linguistic representation Tomasello (1999), pp. 8-9 makes much the same point.

[61] See also Clark (1997a) chapter 8 and Clark (2001b). Slightly greater agnosticism over representation is expressed in Wheeler and Clark (1999). For (in the opinion of this reader, unpersuasive) rejoinder to these views, most specifically of Clark and Toribio (1994), see Dartnall (1996).

changing environment (Brooks [1991]) may be fine examples of online cognitive tasks that can be achieved without appeal to representations (at least of the centrally located symbolic sort) precisely because the relevant features of the world are causally present and perceptually salient. However, the kinds of cognitive operations most distinctive of humans and which appear to have given us a significant cognitive edge over other species, necessitate using representational stand-ins for features of the world.

§3.3 – CONCLUSION

In this chapter I have elucidated and defended an ecumenical conception of representation as it appears in cognitive explanation, identifying the characteristic features of such. I have argued that cognitive representations: 1) are intentional (in something like the standard sense), 2) stand in for what they represent by carrying truth-neutral information about them, 3) represent in virtue of their properties, 4) are isomorphic with what they represent under an interpretation, and 5) are capable of misrepresentation. Moreover, with such a conception of cognitive representation in place, I defended the need for such representations in accounting for significant swathes of cognitive processing, and in the process have deflated arguments for a strong anti-representationalist position.

CHAPTER 4 – PARA-INFORMATION AND EMBEDDED COGNITION

With the two central concepts of cognition and cognitive representation more clearly in place, we are now better prepared to begin our journey out into the trans-cranial and trans-corporeal cognitive world. The broad strategy of the next three chapters will be to defend increasingly strong claims, with each chapter building, to some extent, on the arguments of the previous chapter.

The primary focus of this chapter is an examination of a pair of closely related concepts: 1) what I shall call *para-information* (sometimes treated in the literature – misleadingly, I shall suggest – under the heading of 'implicit' information) and 2) *embedded* cognition. Para-information is information (whether carried by a state or structure within a cognitive agent's brain or in its environment, and whether truth-dependent or truth-neutral) that makes a special explanatory contribution to *how* the cognitive agent tackles a cognitive task. In particular, para-information shapes cognition through enabling the extraction of task-relevant information from a signal. Detailed discussion will follow shortly (§4.1), but we can immediately set out a definition:

> ***Para-information***: A state or structure *S* carries the para-information (if *P* then *Q*) if and only if *S* enables the extraction of task-relevant information that *Q* from a signal that carries the information that *P*.

Embedded cognition is a rather more familiar term and comes to the fore in §4.2. As I shall use it, it refers to cognition of an individual cognitive agent that is shaped by *externally* located para-information. Putting these two together we have the following as a definition:

> ***Embedded Cognition***: a cognitive process or capacity is *embedded* in an environment to the extent that an agent's cognitive behaviour is shaped by para-information *external* to the standard organismic boundaries of the cognitive agent.[62]

A central thesis to be defended in this chapter is that cognition is often deeply embedded in a cognitive agent's environment, i.e., it is often shaped by, and dependent upon, such externally located para-information. As a consequence, I argue, cognition cannot be effectively explained or studied simply by examining processes, states, or structures that occur within the confines of the body of the cognitive agent or between sensory input and motor output. In other words, the environment of a cognitive agent is often as central to the explanation of their cognitive capacities as anything going on inside their skin or brain. As will become clear, this is a weaker claim than that defended in the following two chapters; that the environment is often central to cognitive explanation does not entail that the environment is itself cognitive nor that it is partly constitutive of cognitive processes or states. I shall, however, suggest that the central role of the environment in the explanation of embedded cognitive capacities implies that such capacities supervene upon the combination of the cognitive agent *together with* their environment. A defence of this claim will serve to prepare the way for the stronger and more contentious claims made in the chapters 5 and 6.

Whilst most of the specific examples of para-information that are considered below are drawn from Dretske (1981, 1988), Dennett (1983), and Cummins (1986, 1989), I vary from their respective analyses in various ways. In the first place (§4.1.1), I distinguish para-information in terms of its special role in cognitive explanation – enabling the extraction of information. Consequently, I argue that para-information is *not* to be distinguished in terms of being 'implicit' (Dretske and Cummins) or 'tacit' (Dennett), i.e., in supposed contrast with 'explicit' information. Such a contrast fails to identify the distinctive feature of para-information. Secondly, I present both internal and external cases of para-information on an explanatory par.[63]

Moving then to a discussion of *internal* para-information (§4.1.2), and drawing in part on Dretske's (1981) seminal account, I argue that standard accounts of the role of truth-dependent information in cognition make (often tacit) appeal to para-information. This point is then expanded through an examination of Dretske's (1988) discussion of 'implicit belief'. Moving to external para-information (§4.1.3), I examine examples from Dennett's (1983) and Cummins' (1986, 1989) discussions of, respectively, 'tacit representation' and 'inexplicit information'. This is followed (§4.1.4) by a brief discussion of the representational status of para-informational states or structures and a defence (§4.1.5) of the character and central role of para-information (both internal and external) in cognitive explanation. Lastly, the relationship between cognitive capacities and para-information is crystallised (§4.2) through a brief analysis of the nature of dispositions, and hence (qua dispositions), of embedded cognitive capacities. Dispositions, I argue (§4.2.1),

[62] The significance of the traditional organismic boundaries in biological explanation more generally is itself a matter of some debate. Richard Dawkins (1982), although not entirely rejecting the continuing importance of the organism as traditionally conceived, has done much to demonstrate that biological explanation need not respect the traditional organismic boundaries such as those constituted by skin.
[63] Dennett and I are in agreement on this point; it is less clear to me the extent to which Cummins or Dretske would be so inclined.

are best characterised as supervening on the salient properties of the systems to which they are notionally attributed *together with* those of the systemic elements of their activating conditions. For example, traditional accounts suggest that solubility is a dispositional property of salt and that such a property supervenes on the base properties of salt. By contrast, I suggest that the solubility of salt supervenes only on the base properties of salt together with those of water, not (as many traditional accounts of dispositions appear to suppose) on the base properties of salt by itself.[64] By extension, I suggest (§4.2.2), since cognitive capacities are kinds of dispositions, embedded cognitive capacities should be seen as supervening upon cognitive agents considered not in isolation from, but *together with* cognitively significant features of their environments.[65]

§4.1 – PARA-INFORMATION

What I term para-information is touched upon in the literature under various headings and with certain variations. In order to detail para-information I shall draw, in part, on three such discussions: those of Dretske, Dennett, and Cummins. Dretske discusses internal para-information in the guise of what he terms 'implicit belief', whilst Dennett and Cummins discuss external para-information under the banner of 'tacit representation' and 'inexplicit information' respectively.

§4.1.1 Explicit vs. Implicit Information

One immediate point of clarification concerns the relation of para-information to implicit information. In discussing para-information it is all too easy to compound the two

[64] The attribution of dispositional properties to base properties of the object to which the disposition is notionally ascribed is encouraged by the appearance of dispositional predicates being one-place. To say that glass is fragile, for example, appears not to involve anything other than glass.

and indeed it is sometimes discussed under the heading of the latter.[66] However, this way of delimiting para-information, i.e., in terms of some supposed contrast with explicit information, is at best confused.

The intuitive idea of explicit information is that the information is carried 'on the surface' such that it can be immediately 'read off' from a state or structure without further ado. Conversely, information will typically be said to be carried implicitly if it is not 'on the surface' and cannot simply be 'read off', but instead must be retrieved through further cognitive processing of information that is carried explicitly. Implicit information is often said to be 'potentially explicit' or 'potentially recoverable' from information explicitly represented. For example, information concerning Joe's age would typically be said to be explicitly represented in the sentence 'Joe is 5 years old', but implicitly represented in 'Joe is the 5th root of 3125 years old'.

One attempt to add some precision to this distinction can be found in Dennett (1983, p. 216). He suggests that information is represented in a system *explicitly* if and only if

> there actually exists in the functionally relevant place in the system a physically structured object, a *formula* or *string* or *tokening* of some members of a system . . . of elements for which there is a semantics or interpretation, and a provision (a mechanism of some sort) for reading or parsing the formula.[67]

By contrast, suggests Dennett, a representation can be said to carry information *implicitly* if and only if 'it is implied logically by something that is stored explicitly'. There is, however, an immediate problem with Dennett's way of drawing the distinction. As he recognises, his definition of 'implicit' is not equivalent to the intuitive idea that

[65] I am indebted to Vincent Bergeron and Dustin Stokes for helpful comments on several of the ideas presented in the earlier sections of this chapter.

[66] Cummins (1986), for example, refers both to inexplicit information and to such information as being carried implicitly. See §4.1.3.

[67] As we shall see immediately below, this runs together the mere presence of information encoded within a system with the availability of that information to the system via the provision of a mechanism for interpreting or parsing the formula.

information is implicit if it is potentially explicit or potentially recoverable from information carried explicitly. To take Dennett's example, a mechanical 'Euclid machine' that had the axioms and definitions of Euclidean geometry explicitly stored within it together with a means of churning out theorems would carry every theorem of Euclidean geometry implicitly. Equally, however, there will be many theorems that it would *not* be capable of making explicit because of such limitations as the Einsteinian speed limit during relatively brief life spans. This suggests a problem for Dennett since a great deal of information carried implicitly (by such a criterion) is entirely unrecoverable in practice and hence causally inert. Dennett is clearly aware of this for he suggests that

> [I]t is an interesting question whether the concept of *potentially explicit* representations is of more use to cognitive science than the concept of (merely) implicit representations. Put another way, can some item of information that is *merely* implicit in some system ever be cited (to any explanatory effect) in a cognitive or intentional explanation of any event?

Dennett attacks those offering a negative answer to this question by attacking the conviction that he supposes to motivate such an answer – that only by being explicit can a representation 'throw its weight around' and have any causal effect. Notwithstanding whether Dennett is correct in rejecting such a supposed conviction, this does nothing to show that implicit information as Dennett characterises it constitutes an explanatorily useful kind. Unless the cognitive agent is *actually* capable of extracting the implicit information it can do no explanatory work in respect of that agent's cognitive behaviour.

For such a reason it might be tempting to revert to the narrower definition of implicit information along the lines mentioned above: that implicit information is information that is potentially recoverable by a cognitive agent from information carried by an explicit representation. But as Kirsh (1990) has argued, on such a view it is not at all clear that we can then sustain a categorical distinction between implicitly and explicitly carried information. The distinction, he argues, reflects a matter of degree rather than of kind in which '[e]xplicitness really concerns how quickly information can be accessed, retrieved, or otherwise put to use'.[68] The fact that even putatively explicit representations require interpretive processing (which of necessity will take a non-trivial time-increment to perform) can be seen from a consideration of Dennett's definition of 'explicit representation'. In his definition he insists that for a string or token to count as an explicit representation (in addition to the mere existence of the token) there must also be a mechanism of some sort for reading or parsing it. This compounds 1) there being a state or structure carrying information with 2) the cognitive agent having the wherewithal to access that information or otherwise put it to use. In other words, just as with information that is potentially explicit there must be a way for the cognitive agent to get the information out of the token such that the information it carries can go on to play an explanatory role in that agent's cognition.

Wherever we draw the line between information carried explicitly and information carried implicitly (or even if we reject the distinction wholesale), neither will be equivalent to para-information. Para-information is distinguished not in terms of how quickly one can retrieve it or the ease with which it is retrieved, but in terms of the particular kind of explanatory role it plays in cognition: i.e., in shaping *how* a cognitive agent tackles a cognitive task through enabling the extraction of task-relevant information.

§4.1.2 Internal Para-Information

To begin to see both the character and central explanatory role of para-information in cognition it is perhaps simplest to begin by examining its role in the internal extraction

[68] Kirsh (1990), p. 361. Kirsh's arguments are multi-stranded and I shall not pause to do justice to them since

of truth-dependent information. As we noted in chapter 2, there are two broad senses of information prevalent in the literature, each with a distinct explanatory role. The central role of the truth-dependent sense of information (the 'indicator' sense) lies, it was suggested, in helping to explain how our behaviour gets to be so intelligent. Within folk-psychological explanation, for example, it is the extent to which our beliefs are reliably true (or vary counterfactually with respect to the world) that ultimately underpins their role in explaining the intelligence of our behaviour. A central role of para-information, I suggest, lies in facilitating the extraction of such information from signals or states. In cases of internally situated para-information this results in the cognitive agent being said to have the relevant cognitive *know-how*.

Consider Dretske's (1981) seminal account of the role of truth-dependent information in cognition. Central to Dretske's account is the informational content of a signal (or state). Where *s* is some system, *F* is a property of that system, *r* is a signal (or, equivalently, some state), and *k* is background knowledge concerning possibilities existing at the source, the informational content of a signal can be defined in the following way:

> A signal *r* carries the information that *s* is *F* = the conditional probability of *s*'s being *F*, given *r* (and *k*), is 1 (but, given *k* alone, less than 1). (Dretske, 1981, p. 65)

At the heart of this concept of informational content is a counterfactual-supporting relation between states. In the above definition, for example, were *s* not *F* then either *r* or *k* would not have obtained. This is common ground with a number of others who, like Dretske,

the implicit-explicit distinction fails to capture what is distinctive of para-information.

look to counterfactual relations as a basis of semantic content, including, for example, Barwise and Perry (1983), Stalnaker (1984), and Fodor (1990a).[69] [70]

Clearly, Dretske's account of informational content is crucially dependent on the role of background knowledge (*k*) concerning possibilities existing at the source. Such knowledge is encoded in the agent as a result of earlier information carrying events. It is, on Dretske's account, caused (or causally sustained) by information. In elucidating the role of background knowledge Dretske asks us to consider the following case:

> Suppose there are four shells and a peanut is located under one of them. In attempting to find under which shell the peanut is located, I turn over shells 1 and 2 and discover them to be empty. At this point you arrive on the scene and join the investigation. You are *not* told about my previous discoveries. We turn over shell 3 and find it empty. (Dretske, 1981, p. 78)

Dretske suggests that '[s]ince I was able to learn something from the observation that you were not, the observation must have been more pregnant with information for me than for you. I must have received *more information* from this single observation than you.' This is clearly because he (Dretske) has more background knowledge (*k*) concerning the possible locations of the peanut: he, unlike the late arrival, already knows that the peanut is not under either of the first two shells.[71]

[69] Notice that given that information is dependent on a nomic regularity between event *types*, whether tokens carry information will depend on how such tokens are typed. In one standard example, pheasant tracks laid down in an area containing otherwise type-identical quail tracks does not carry information that a pheasant has passed by because otherwise type-identical tracks can occur when no pheasant has been around – i.e., when a quail has passed by. But if we type identify a token track as *bird* tracks then it does carry information that a bird has passed by (assuming no non-bird makes type-identical tracks).

[70] It is important to distinguish such *counterfactual-supporting* co-variation from *mere* co-variation. Fodor (1986), citing Dretske (1981) as an advocate, identifies the 'Standard' view of information as co-variation, without distinguishing from this its counterfactual-supporting subclass that I am here identifying as the truth-dependent or indicator view. This seems unfair to Dretske who is a great pains to insist that information is not *mere* co-variation but must be nomic (hence counterfactual-supporting). Fodor is quite critical of the Standard notion of information arguing that it is an inadequate basis for an account of cognition. His primary focus, however, is against associationist accounts of cognition to which the Standard view might be seen to be a natural partner. Later, however – for example Fodor (1990a) – he endorses a counterfactual supporting co-variation view of information, and his own semantic theory comes to rely heavily upon it.

[71] Dretske candidly admits that intuitions may differ as to how best to characterise the situation in intuitional terms. He points out that one might equally argue that 'the single observation (the one we made together of shell 3) carries the same information to both of us. The explanation for why I learned more from it from you (viz., that the peanut is under shell 4) is that *I knew more to begin with.* The information that the peanut is under shell 4 is a piece of information that is carried by the composite set of *three* signals – not by any single

But as Dretske acknowledges (see below), there is much more required than background knowledge of the possibilities existing at the source (i.e., about where the peanut might be) before the information in signals or states can be put to explanatory use in cognition. In the above example, in order to know that the peanut is under shell 4 Dretske must also know *how* to extract the information from the turning of the shells and *how* to combine that information with information carried in prior cognitive states. There are several hypothetical routes by which Dretske might go about this. The one Dretske supposes is indicated by his claim that '[h]aving already examined shells 1 and 2, I know that they are empty. The peanut is under either shell 3 or 4' (Dretske, 1981, p. 79). Dretske seems to suppose (plausibly enough) that he reasons in stages, generating intermediate cognitive states as each shell is turned. In more detail, the story might go something like this: The first peanut is turned and Dretske extracts from the signal the information (a) that there is no peanut under the first shell. This is combined with the background knowledge (*k*) that there is one peanut and it is under one of the four shells to generate (b) the peanut is under shell 2, 3, or 4. The second peanut is turned and Dretske repeats the extraction of information from the signal then combines it with (b) to generate (c) the peanut is under shell 3 or 4. And so on. The central point is that in order for the information carried in the turning of the shells can play an explanatory role in his cognition, Dretske must know *how* to extract the information from the signals and must know *how* to combine that information with information carried by pre-existing and intermediate knowledge states. The states or structures that enable the extraction of such information and, hence, give rise to such know-how, carry what I refer to as para-information.

observation. Hence, it is not true that I learned more *from the third observation* than you did.' Dretske (1981), p. 79.

Consider another example from Dretske (1981) concerning a child who is looking at a daffodil. The child (unlike the teacher, say) may not know that it is a daffodil despite seeing the daffodil perfectly clearly. Regarding the child Dretske suggests that

> The requisite information (requisite to identifying the flower *as* a daffodil) is getting in. What is lacking is an ability to extract this information, an ability to decode or interpret the sensory messages. What the child needs is not more information about the daffodil, but a change in the way she codes the information that she has been getting all along. (Dretske, 1981, p. 144)

As Dretske points out, what the child lacks is not information about the daffodil but something more like a cognitive *ability*, or a cognitive *capacity*, i.e., a particular kind of cognitive *know-how* pertaining to daffodils. As Dretske suggests, all the information about the flower that is required to determine that it is a daffodil is already in the perceptual signals. But the *capacity* to take that sensory information and process it correctly – i.e., the capacity to identify a flower *as* a daffodil – is not carried in her perceptions at all, either explicitly or implicitly. For the child to be able to identify the daffodil as such she must come to know something over and above that the object before her is a flower with four large yellow petals, etc.; she must come to know that having those features makes it a daffodil. Similarly, returning to the peanut and shells example, in order to be able to determine that the peanut is under the fourth shell Dretske must know something more than that there are no peanuts under the first three shells; he must also have the know-how to apply a functional equivalent of the disjunctive syllogism rule of propositional logic.

Such cognitive know-how is equally well characterised in propositional terms. For example, the teacher's cognitive ability to distinguish daffodils from other flowers can be expressed by saying that she knows *that* flowers with such-and-such properties are daffodils. Similarly, Dretske must know *that* if the peanut is not under any of the first three shells, then it is under the fourth. Such knowledge exhibits a kind of 'dual aspect', being equally characterisable as either a cognitive ability (a cognitive capacity or a kind of cognitive know-how) or in propositional terms as a kind of knowing-that. In explaining some cognitive agent's behaviour (say, an ability to identify daffodils), we can dress up our explanation in terms of either and to much the same explanatory effect.

Dretske (1988), drawing, in part, on Cummins (1986), develops this dual aspect in more detail by distinguishing between *explicit* and *implicit* beliefs. Whilst he does not define 'explicit', it would seem to include, for example, Dretske's belief that there is no peanut under the first shell or the child's belief that the flower in front of her has four large, yellow petals. Implicit belief, by contrast, 'is a disposition or rule that describes the relationship among entities that are already intentionally characterised . . . or among such intentionally characterised entities and movements' (Dretske, 1988, p. 118). For example, the disposition to believe that *Q* when one believes that *P*, allows us to speak of the implicit belief that *if P* then *Q*. Dretske's disposition to believe that the peanut is under the fourth shell if he believes that it is under none of the first three shells (together with the belief that there is a peanut under one of the four shells) are of this kind. As Dretske makes clear, the explanatory role of such beliefs is different from that of the intentional entities over which they operate; they are, he suggests, 'perhaps better thought of as ways a system has of manipulating information than as part of the information they manipulate. They are, as it were, part of the program, not part of the data on which this program operates.'[72]

It is also important to note that such knowledge is not *merely* a matter of know-how but fully cognitive in character. It is, Dretske point out:

[72] *Ibid.* A Dretskean inspired account but which, like Cummins (see below), allows for 'false information' can be found in Matthen (2004). What Matthen characterises as a 'background theory' is akin to Dretske's 'implicit beliefs'.

> a unique mixture of the practical and the theoretical. It isn't *just* a piece of know-how, like knowing how to swim or to wiggle one's ears. . . . [It is] a genuine cognitive skill, something more like knowing-*that*, a piece of factual knowledge rather than just a piece of knowing-*how*. (Dretske, 1988, pp. 116-7.)

It is for this reason that we can equally well speak of the teacher as knowing *that* if a flower has four large yellow petals (etc.) then it is a daffodil as we can of her having the know-*how* to identify daffodils. Such 'implicit' knowledge directly shapes cognition by enabling the extraction of task-relevant information from information carried by sensory signals, and is para-informational in virtue of such. Such knowledge reflects conditional relations between states – in the above case, between something having four large yellow petals, etc., and it being a daffodil. It reflects conditional relations by giving rise to a move between an informational state (the flower having four large yellow petals, etc.) and a cognitive action (the coming to believe that the flower is a daffodil).

§4.1.3 External Para-Information

The examples considered by Dretske are all cases of internal para-information. As mentioned earlier, however, para-information can be carried *externally* as well. Dennett (1983) identifies precisely such under the banner of 'tacit representation'.[73] He does not give it an explicit definition but instead elucidates the notion through examples. A pocket calculator, for example, gives us cognitive access to a virtual infinity of arithmetical facts. Yet, as Dennett points out:

> If one looks closely at the hardware one finds no numerical propositions written in code in its interior. The only obvious explicit representations of numbers is either printed on the input buttons or, during output, displayed in liquid crystal letters in the little window. (Dennett, 1983, p. 221.)

[73] As will be clear, Dennett's use of 'tacit' is not to be confused with the not uncommon usage by many (e.g., Fodor (1968) and Chomsky (1965)) of 'tacit' to refer to unconscious intentional states such as one's 'tacit knowledge' of the grammar of one's language. Such tacit knowledge is unconscious and is standardly taken to be explicit.

With due allowance for Dennett's infelicitous confounding of numerical propositions with sentences representing numerical propositions (given that propositions are abstract entities we would presumably not expect to find them inside calculators), his central point is that there remain no explicit representations of arithmetical propositions anywhere in the calculator and hence neither explicit nor implicit representation (in his usage of those terms) of arithmetical facts. This is true even if one were to insist that there are temporary explicit representations of numbers in such things as internal electronic buffers. It is simply that

> [i]ts inner machinery is so arranged that it has the fancy dispositional property of *answering arithmetical questions correctly*. . . . [I]t does this without relying on any explicit representations within it – except the representations of the questions and answers that occur at its input and output edges and a variety of interim results.[74]

In a very similar vein Cummins (1986 and 1989), expanding, in part, on similar views expressed in Dennett (1978),[75] details several ways in which 'information'[76] or 'content' as he characterises it, can be carried *inexplicitly* (or, as he sometimes puts it, 'implicitly') by a system – i.e., 'without benefit of any symbolic structure having the content in question'. Like Dennett, Cummins focuses on the cognitively significant role of states or structures *external* to the traditional organismic boundaries of the cognitive agent.

[74] Dennett (1983), p. 221. More precisely, we might say that the pocket calculator exhibits a dispositional capacity to take inputs systematically interpretable as arithmetic operations over numbers and converts those inputs into outputs systematically interpretable as the results of such arithmetic operations.

[75] Specifically, Cummins cites Dennett (1978), p. 107. On the attribution by a chess-program designer to a chess-playing program that 'it thinks it should get its queen out early', Dennett comments that 'for all the many levels of explicit representation to be found in that program, nowhere is "I should get my queen out early" explicitly tokened'.

[76] Cummins is less than clear precisely how 'information' is here to be understood. Cummins (1986) consistently talks about such states or structures as carrying inexplicit 'information', but notes that he intends the term 'information' to be used is a manner that allows (contra Dretske) for 'false information'. This might suggest that Cummins has in mind truth-neutral information and hence considers such states or structures to be representational. This is perhaps further suggested by Cummins (1989, p. 157, fn. 9) where he states that he should have called Cummins (1986) 'Inexplicit Content' rather than 'Inexplicit Information'. But any suggestion that he views such states or structures in such terms is immediately scotched by his insistence (Cummins, 1989, p. 158, fn. 10) that 'inexplicit representation' is a contradiction in terms.

One important kind of inexplicit information Cummins identifies is *domain*-implicit information; i.e., information lodged in or carried by the environment of the cognitive agent. For example, the only way one could use a set of directions (a program) to locate Cummins' house 'would be to execute it, either in real space or using a sufficiently detailed map. The information in question is as much in the map or the geography as it is in the program' (Cummins, 1986, p. 119). A second kind of inexplicit information that Cummins considers – *medium*-implicit information – is carried in some respect by the structural properties of the medium in which the information is carried. For example, whether two complex but differently oriented figures are congruent may be possible to determine only if the figures are represented on a medium, such as transparencies, that allows transparent overlay and rotation of the two representations. Without representing them on a suitable medium – one that facilitates the extraction of the relevant information – one might be incapable of determining whether they are congruent. But by overlaying and rotating the transparencies one can quite easily determine whether in fact they are congruent.[77]

§4.1.4 Para-Information and Representation

[77] There are two further kinds of inexplicit information (or content) that Cummins discusses that are less central to our present purposes. *Control*-implicit information is carried in the logic or structure of the flow of control in a system. For example, a circuit fault diagnosis system that is so organised that it checks capacitors only after it has verified the power supply and is currently checking the capacitors carries the information that the power supply is okay. All without necessarily explicitly representing to itself that the power supply is on. Another kind – *form*-implicit information – is carried in the 'form' of a representational system. For example, the partial product algorithm for calculating products of two numbers relies on information implicit in the form of the denary numerical system – moving a numeral one column to the left is equivalent to multiplying it by ten. The importance of such formal features can be seen from the fact that one cannot determine the product of two numbers using such an algorithm where the same sum is expressed in, say, Roman numerals. Moving a Roman numeral to the left one place does not amount to multiplying it by ten, as it does in the denary system. The general point is that such algorithms might be said to carry information only in relation to the structures through which they operate and that such structures may, therefore, be said to carry part of the information-bearing load.

Before examining these examples in more detail, it is important to note that the absence of *explicit* representation (Dennett would also add, under a different usage of the term from Cummins, the absence of implicit representation) does nothing by itself to suggest that no task-relevant information – para-information in these cases – is being represented. This is relatively clear in Dretske's internal examples – the teacher (but not the child) internally represents para-information regarding the relation between being a daffodil and having four large yellow petals, etc., and Dretske internally represents a functional equivalent to the disjunctive syllogism rule of propositional logic. Neither of these need be explicitly represented. Equally, in Dennett's external example of a pocket calculator, a strong case can be made that the causal relations in which the parts of a pocket calculator stand (via its circuitry) represent various arithmetical relations, but without those relations being represented explicitly. Nevertheless, it seems quite appropriate to say that the para-information regarding those arithmetical relations is represented 'tacitly', as Dennett puts it, in the design of its circuitry.[78]

In other cases of para-information, for instance in Cummins' examples of the local geography and the transparencies, one would be ill advised to characterise them as representations. In the first place, it defies the general description of a cognitive representation as an *intentional* entity to characterise them in such terms. There seems to be nothing, for example, that a transparency or the local geography is *about* in the standard intentional sense of that term, least of all (respectively) the congruence of certain figures or the location of Cummins' house. A second feature of cognitive representation that is not satisfied by such examples is that it be capable of *misrepresenting*. Both the set

[78] A glance back at the characteristic features of cognitive representations (§3.1.6) should satisfy us of this point. We will return to such examples in chapter 6 where we consider a representational version of extended cognition.

of directions and the map can sensibly be said to misrepresent matters, but this makes no sense in the case of the local geography. Were someone to fail to determine the location of Cummins' house by walking the route specified in the directions, it would makes no sense at all to say that the structure of the local geography was 'wrong', or 'mistaken'.

§4.1.5 The Explanatory Role of Para-Information

As we have noted, however, not being representational does nothing to undermine the credentials of states or structures as carriers of truth-dependent information generally, so neither does it bar such states or structures from carrying para-information in particular. This is, perhaps, best seen by recalling that truth-dependent information is carried in virtue of a counterfactual-supporting relation between states. For example, a bid of 'four no trump' under the Blackwood convention in bridge is counterfactually dependent on the bidder having either four or no aces. Because of that counterfactual-supporting relation, the signal or state (i.e., the bid of 'four no trump') eliminates possibilities and reduces uncertainty about the conditions at the source (the bidder having neither 1, 2, nor 3 aces). At first glance, it might seem that no truth-dependent information is carried by such states or structures as the local geography or the transparencies. The transparency and rotational rigidity of transparencies does not vary in a counterfactual-supporting way with respect to the congruence or otherwise of figures printed on them. Similarly, the structure of the local geography does not vary in a counterfactual-supporting way with respect to the location of Cummins' house. As such, it might seem, no possibilities concerning the congruence or otherwise of the figures, nor concerning the location of Cummins' house are eliminated by such states, and no (task-relevant) truth-dependent information is carried.

Recall, however, the teacher's ability to identify a flower as a daffodil. She has acquired a disposition to believe that a flower is a daffodil if it has four large yellow leaves, etc., and has thereby stored some information that *is* (we may suppose) counterfactually dependent on the relationship *between* the property of have four large yellow petals (etc.) and the property of being a daffodil. Certain possibilities are eliminated by such information: namely, the possibility of the flower before her being both not a daffodil and having four large yellow petals (etc.). Moreover, without this information she is entirely incapable of tackling the cognitive task (i.e., of determining whether a flower is a daffodil). Such information – para-information as I term it – does not carry information about (say) whether this or that flower is yellow-leaved (etc.), but instead, carries information about what it is to be a daffodil or (equivalently) *how* to identify daffodils. It is this information that enables the teacher to extract the information that the flower before her is a daffodil from the information delivered by sensory signals that it has four large yellow petals, etc.

It is in essentially the same respect that the transparencies and the local geography can be said to carry para-information, i.e., by enabling the extraction of task-relevant information from the figures and the directions, respectively. Given that the figures are congruent, the salient properties of the transparencies (its transparency and rotational rigidity in the plane of the figures) ensure that it carries task-relevant information: namely, that if one places one transparency over the other and rotates one of them in the plane of the figures, then they can be made to visibly overlap. This is para-information because if the figures are overlaid, rotated, and checked for visual overlap, they enable the extraction of information from the figures. Similarly, the structure of the local geography carries the para-information that if one follows the directions, one will arrive at Cummins' house.

This likewise shapes the cognitive process through enabling the extraction of information (here carried in the directions).

To flesh all of this out, let us take the example of determining where Cummins lives from a set of directions and consider it in more detail. Suppose that someone (call her 'Alice') is given a set of directions (a 'program') for getting to Cummins' house from (say) *The Pig and Whistle* pub. As Cummins (1986, p. 119) points out, 'in some sense, the program represents me as living in a certain place, perhaps correctly, perhaps not'. But as he also points out, nowhere is the proposition 'Cummins lives at location L' explicitly represented in the directions. The directions, therefore, provide no more than a *partial* explanation of Alice's cognitive capacity to determine where Cummins lives. This is demonstrated by the fact that Alice is entirely incapable of determining the location of Cummins' house given only the directions. As Cummins suggests, in order for Alice to use the directions to determine where he lives she must execute the program either out in the world or on a map. To do either of these is to *add* something to the directions; it is only the combination of the directions *together with* either the structure of the local geography or the structure of the map that enables Alice to determine where Cummins lives. In the case of using the map, what is added is the para-information carried by (and retrieved from) the map. Given that Alice can read the map – i.e., that she can extract the task-relevant para-information from the structural features of the map – then, by (say) following the directions with her finger on the map (and assuming that both the directions and the map are accurate), she can determine where Cummins lives. The map, in virtue of accurately representing the structure of the local geography, provides for Alice something utterly essential to the explanation of her capacity to find Cummins' house.

By parallel, when Alice follows the directions out in the world without the aid of a map, it is *the structure of local geography* that provides something utterly essential to explaining her capacity to locate Cummins' house. As with the structure of the map, the para-informational structure of the local geography makes the same essential contribution to the explanation of Alice's capacity to determine where Cummins lives. It determines a set of conditional relations between origins, sets of directions, and locations (end points). The particular para-informational conditional extracted from that structure will depend on, and be relative to, the directions. In Alice's case, it may be something like 'if one turns left out of the Pig and Whistle pub, follows the road down the hill to the post office, turns right, goes to the end of the road, then Cummins house is directly in front'.[79]

The explanatory role of para-information can now be drawn more precisely. Either the map or the local geography itself, together with the directions, provides Alice with the capacity to determine where Cummins lives. Either provides this capacity precisely because the directions are, so to speak, reliably 'keyed-in' to a common structure – the structure shared by both the map and the local geography. In other words, the set of directions is counterfactually dependent on the *combination* of the location of Cummins' house together with the structure of the local geography. *Given* the structure of the local geography, if Cummins had lived elsewhere, then the directions would have been different. Similarly, *given* where Cummins lives, had the structure of the local geography been different, then the directions would likewise have been different. In being counterfactually dependent on the combination of the structure of the local geography and the location of Cummins' house, the directions 'triangulate', so to speak, a counterfactual-supporting relation *between* the location of Cummins' house and the structure of the local

[79] A similar point regarding the connection between certain sensory states and epistemic action is made in

geography; the directions 'tie' the location of Cummins' house and the structure of the local geography together in a counterfactual-supporting way. Thus, given the directions, had the structure of the local geography been different, so would have the location of Cummins' house.

Moreover, the explanatory appeal to a common structure shared by both the map and the local geography has much the same explanatory virtue as Pylyshyn finds in truth-neutral information (see §2.3.1). Pylyshyn points out that a common representational content allows us to locate in a disparate array of physical events a common explanatory feature. Similarly, we can locate a common explanatory feature between information-bearing representations such as maps and para-informational states or structures such as local geographies. Both possess a common structure, and it is the sensitivity of a cognitive agent to that structure that enables both the map and the local geography to make essentially the same explanatory contribution to their behaviour.

It is possible that a critic might try to downplay the role of the structure of the local geography by suggesting that it functions only as what Dretske calls a 'channel condition'. Channel conditions, he tells us, are 'the set of existing conditions (on which the signal depends) that either (1) generate no (relevant) information or (2) generate only redundant information (from the point of view of the receiver)'.[80] The role of such background conditions is to underwrite the transmission of information by ensuring reliability but, as Dretske makes clear, the receiver need have no knowledge or information about them. To use an example of Dretske's, in order for a voltmeter to carry information about some voltage drop across a resistor, the various internal workings of the voltmeter (such as

Matthen (2005, pp. 230-1).

electrical contacts, springs or what have you) that mediate the flow of information must be in good working order. The fact that some particular cognitive agent may be completely ignorant both of the inner workings of the voltmeter and whether they are in good working order in no way undermines that fact that, just so long as the inner workings *are* in good working order, the voltmeter – specifically, its read-out – carries information about the potential difference across a resistor (see Dretske, 1981, pp. 111-114).

But it seems clear that, given only the directions, the para-information carried by the local geography is far from redundant. Without access to it (or a representation of it in a map) Alice is entirely incapable of completing the task. Far from being a condition that can, from the point of view of cognitive explanation, remain entirely in the background, it is difficult if not impossible to see how Alice could succeed in determining where Cummins' lives without her being cognitively sensitive, whether directly or indirectly, to the structure of the local geography. The structure of the local geography plays a role in the explanation of Alice's cognitive behaviour not merely in helping to determine (in combination with the directions) the location of Cummins' house, but by influencing her behaviour in a way that is every bit as significant in explaining her success as that played by the directions themselves.

§4.2 – PARA-INFORMATION AND DISPOSITIONS

Cognitive capacities are, qua capacities, plausibly construed in terms of being a certain kind of dispositional power of a cognitive agent. Precisely how, though,, are we to understand such dispositional properties? Analyses of the general nature of dispositions

[80] Dretske (1981), p. 115. Channel conditions equate, with minor modifications, to what Stampe (1977) calls 'fidelity conditions', Barwise and Perry (1983) call 'constraints', and Stalnaker (1984) calls 'relevant normal conditions'.

have generated some debate in recent years and the correct analysis appears to remain a matter of dispute.[81] In the light of this and given constraints of space, I shall restrict myself merely to outlining an allegiance to a naturalistic and realistic account of dispositions. I shall, however, add my own slant that brings out more clearly the connection between para-information and embedded cognition.[82] Embedded cognitive capacities, I shall suggest, are realised in properties of cognitive agents *together with* their environment – they are not fully realised in properties of the cognitive agent alone.

Concern might be expressed that to focus on cognitive capacities shaped by one's environment constitutes a radical and unpromising departure from traditional cognitive science. The general goal of cognitive science has been to determine the mechanisms through which we have our cognitive capacities. Moreover, this has taken the form of a nearly universal focus on environmentally robust capacities – capacities that a normal cognitive agent possesses regardless of the specifics of their environment. There is much merit in such a focus, for an important desideratum of any science is explanatory generality. Our shared biological inheritance, abstracted away from the varied contingencies of our environments, provides a natural and obvious realm for the seeking of such generalities. By contrast, it might be said, the variety of our physical, cultural, and technological environments, were we to include them within our explanatory domain, place a significant (and perhaps insuperable) barrier to the formulation and discovery of such generalities.

[81] See for example, Johnston (1992), Martin (1994), and Lewis (1997) for difficulties associated with the 'conditional analysis' of dispositions. For a defence of conditional analysis see Gundersen (2002). I have gained an appreciation for the difficulties and complexities associated with dispositions from reading Mumford (1998).

[82] Thus, I make no presumption as to whether Dennett, Dretske, or Cummins would agree with my characterisation of dispositions. In fact, Dretske explicitly states that 'an implicit belief or representation is something like what Ryle (1949) called a single-track disposition'. Dretske (1988), p. 117. Ryle adopted a conditional view of dispositions, which I reject.

There are, however, several features of human cognition that mitigate this concern. The first is a matter that I take up in greater detail in the following two chapters, namely, the extent to which externalisation of representations and processes over those representations constitute *the* most distinctively human mode of cognition. Cognitively speaking, from cave paintings to digital computers, we are first and foremost a species who utilises the environment for symbolic purposes in order to significantly enhance our cognitive capacities. To focus exclusively on cognitive capacities independently of an environment that is very often specifically shaped to suit our cognitive purposes simply for the sake of explanatory generality is to ignore what is perhaps the single most striking feature of human cognition. A second mitigating feature concerns the extent to which features of the environment may be said to play a highly ubiquitous role in the cognitive lives of a great proportion of humanity. From tattoos and knots in strings to Palm Pilots, information technology is, for many of us at least, a ubiquitous and expanding feature of our cognitive lives. Lastly, notwithstanding how pervasive the role of the environment might or might not be in our cognitive lives, explanatory generality is not the sole goal of any science. Generality can always be achieved by ignoring diversity and by restricting one's explanatory domain to suit, and can always be bought at the expense of completeness. To draw a parallel, the science of human kinetics may be based in an essentially biological domain. It is presumably within that domain that greatest explanatory generality may be found. Nevertheless, many humans utilise features of their environment (from a fallen tree-branch used as a walking staff to state-of-the-art prosthetic limbs) to aid their movement. Notwithstanding the extent to which the inclusion of such environmental features may hinder explanatory generality, the science of human kinetics cannot (and does not) shrink from studying the use of such environmental features. Likewise, I suggest, for cognitive science.

§4.2.1 Dispositions In General

If we consider the general character of dispositions a number of general features emerge. Firstly, dispositions – paradigmatic examples include the fragility of glass and the solubility of salt in water – may be roughly characterised as the propensity of something to do certain things under certain conditions. This involves reference to at least three elements: 1) a *system* to which the disposition is attributed, 2) an *activating condition*, and 3) a *manifestation* of the disposition. For example, in attributing to salt the disposition to dissolve in water we have a system (salt), an activating condition (being placed in water), and a manifestation of the disposition (dissolving behaviour, in this case of the system with the disposition). The activating condition (being placed in water) will refer to a system or 'systemic element' (in this case *water*) that plays a causal role in the manifestation of the disposition.

Secondly, dispositions are attributed in accordance with relevant conditionals.[83] For instance, glass does not have to be struck by a sufficiently hard and momentous object in order to be fragile, and salt does not have to be placed in water to be soluble in water. What matters is that glass *would* shatter *if* it were involved in a suitable collision, and salt *would* dissolve *if* it were placed in water, etc. These conditionals will contain a (typically implicit) ceteris paribus clause that is required in order to allow for various background conditions that, even given the existence of the nominal activating condition, may

[83] This is sometimes expressed in terms of *counterfactual* conditionals, but this is not universally the case. Elastic may retain a disposition to stretch even though it is currently being stretched.

interfere with the manifestation of the disposition. For instance, salt will not dissolve when placed in water if the water is completely saturated or frozen.

Thirdly, I make the naturalistic assumption that there will be some property or properties in virtue of which the system has its disposition – a causally efficacious *base* or *ground* for the disposition.[84] Many commentators suppose that the grounds for dispositions must be amongst the intrinsic or categorical properties of the systems to which dispositions are attributed. This is perhaps encouraged by dispositions being frequently characterised through the use of one-place predicates, such as 'glass is fragile'. In any case, I disagree. Consider again the standard example of salt dissolving in water. What is certainly clear is that if we are to explain why being placed in water results in the dissolution of salt but not, by contrast, of carbon, we must presumably appeal to some property of salt not shared with carbon. Quite clearly, some property of the salt makes a causal and explanatory difference to it having the disposition to dissolve in water. Equally, however, our explanation of the disposition will need also to appeal to some property of the systemic element of the activating condition, namely, *water*, if we are to account for the fact that salt dissolves in water but not in, say, oil. Clearly, the properties of water in contrast with those of oil, just as much as the properties of salt in contrast with carbon, make a causal and explanatory contribution to salt having a disposition to dissolve when it is placed in water. It seems, therefore, that the intrinsic or categorical properties of salt do not *fully* explain its disposition to dissolve in water independently of the intrinsic or categorical properties of water. Dispositions are to be explained by, and grounded in, an appeal to the intrinsic or categorical properties of *both* the system to which the disposition is notionally attributed *and* the systemic elements of the activating condition, with each making its own causal and explanatory contribution.

Our ordinary talk regarding dispositions, however, tends to locate them within one system or another and in doing so obscures the distributed grounding of dispositions. We attribute the disposition to dissolve when placed in water *to the salt* as though the disposition, per se, were to be found somewhere in the salt along with (or supervening upon) its intrinsic or categorical properties. This tends to obscure the key explanatory contribution of the systemic elements of the activating condition. As the above considerations suggest, the disposition to dissolve in water is no more literally in the salt than it is in the systemic elements of the activating condition. All that we can properly locate within the salt is whatever intrinsic or categorical properties constitute the salt's causal-explanatory contribution to it having the disposition in question.

Moreover, the distributed grounding of dispositions in both system and systemic elements of the activating condition is of a piece with the fact that for each disposition attributable to a system, an equivalent disposition – equivalent in that its attribution is predicated upon precisely the same set of conditional facts – will be attributable to the systemic elements of what is *notionally* considered the activating condition. In other words, one can, mutatis mutandis, swap the system to which one notionally attributes the disposition with the systemic elements of the activating condition, treating the erstwhile system as constituting the systemic elements of the activating condition and the former system as constituting the systemic elements of the activating condition. For example, we typically say that salt has the disposition to dissolve when placed in water. Equally,

[84] This distinguishes the view of dispositions advocated here from conditional accounts such as advocated by Ryle (1949). On such accounts, dispositions are not properly speaking causally efficacious properties of

however, we can circumscribe precisely the same set of conditional facts by saying that *water* has the disposition to dissolve salt that is placed in it. Certainly, we distinguish transitive from intransitive dispositional relations – a disposition *to dissolve* something is a complementary disposition to the disposition *to be dissolved by* something. But this distinction amounts to little more than approaching the same set of conditional facts from different directions, much as we can say, to take a non-dispositional example, either that *x* is to the left of *y* or that *y* is to the right of *x*. In either case the same relational fact is being described.[85]

§4.2.2 Para-Information and Embedded Dispositions

With these general observations concerning dispositions in mind, the connection between para-information and embedded cognitive capacities should be plainer to see. Particular cognitive processes can be seen as manifestations of cognitive dispositions (capacities). For example, a particular in-the-head arithmetical calculation can be seen as a manifestation of a cognitive capacity for doing in-the-head arithmetical calculations. Where those processes are embedded – i.e., where they are shaped by para-information external to the traditional organismic boundaries of the cognitive agent – the cognitive capacities of which the process is a manifestation will likewise be embedded. This conclusion maps directly onto the above general analysis of dispositions. Dispositions (I have argued) are best analysed in terms of, and grounded in, the intrinsic properties of the system to which the disposition is notionally attributed (in the case of cognitive capacities, the cognitive agent) *together with* those of the systemic elements of the activating

objects, but are reduced to 'if ... then' statements ranging over actual or possible events. See Mumford (1998), pp. 15-17, for a brief discussion of the conditional account.

condition. Para-information lying beyond the traditional organismic boundaries of the cognitive agent constitutes the activating conditions of such dispositions.

Consider, again, the role of what Cummins calls 'medium implicit' information carried by transparencies in the determination of whether two figures are congruent. As he puts it:

> if you had each [figure] on a transparency, you could simply put one over the other and rotate them relative to each other to see if they would match. But this works only because of two properties of the *medium* (i.e., the transparencies): They are transparent, and they are rigid in the plane of the figures. (Cummins, 1989, p. 17.)

The central point in this example is that, under certain conditions, certain properties of the medium of representation (at the most immediate level, its transparency and rigidity in the plane of the figures) play an enabling role in the extraction of information about the figures, namely, whether they are congruent. For sufficiently complex figures, normal human cognitive agents are unable to determine whether the two figures are congruent if they are printed on materials that were either opaque or not rigid under rotation.[86] The properties of the medium also shape the cognitive process inasmuch as extracting the task-relevant information requires that we overlay and rotate the figures. The result is that we humans have a cognitive capacity to extract such information from transparencies, and that this capacity is, in crucial respects, dependent upon and shaped by certain structural properties of transparencies. Such a cognitive capacity, therefore, is embedded and (as argued in general terms above) the explanation of the cognitive capacity supervenes on

[85] It may be, of course, that the base properties of salt that contribute to it dissolving in water may contribute to it dissolving in a broader class of liquids. The disposition of salt to dissolve in water may, therefore, be a specific case of a broader disposition to dissolve in liquids of a kind of which water is a particular member.

[86] We might, of course, find some alternative ways to determine whether the figures are congruent. We might, for example, scan the images into a computer and then use some software to rotate and overlay the scanned figures. The general point remains – some property of the medium plays a central role in explaining such a cognitive capacity.

certain properties of the cognitive agent taken together with certain properties of the transparencies.[87]

One might draw an analogy here with locks and keys.[88] We commonly attribute to keys a dispositional capacity to open certain locked doors. But it is not the properties of the key *alone* that explain the unlocking of doors. If we wish to explain the fact that a certain locked door was (or could be) opened we must appeal to the salient structural properties of the key *together with* those of the lock. It is only when locks and keys are causally united as a single key-plus-lock system that the door is (or can be) opened. It is not as though the key, all by itself, gets the door partly opened and the lock then takes over and does the rest (or vice versa). Keys, we may say, have an embedded capacity to unlock doors, given an environment with the right kind of locks.

None of this *dissolves* the traditional ontological boundaries between cognitive agents and their environments, any more than it dissolves the ontological boundaries between keys and locks. The properties of keys and the properties of locks can, and at least for some of their properties, should be investigated independently. But door-unlocking behaviour cannot be properly investigated by looking merely at the properties of keys, any more than it can by looking merely at the properties of locks. To explain such behaviour we must look at the properties of keys together with those of the locks with which they are paired. Similarly, cognitive agents and their environments can and sometimes should be investigated in complete isolation from each other to the extent that they each retain interesting behaviours independently of the other. But when we appeal to

[87] Similar points to those being made in this section are made in Wheeler and Clark (1999). They draw their inspiration from the embedded character of 'genic representation'. That is, from the fact that DNA 'codes' for phenotypic traits whilst at the same time the expression of those traits is 'causally distributed'. That is, the expression of traits depends not just on the DNA but also upon a variety of parametric conditions of the zygotes environment.

[88] An example drawn from Cummins (1996, p. 102).

dispositional capacities to explain the various kinds of cognitive behaviours examined above, we cannot do explanatory justice to the phenomena without treating the cognitive agent and the salient features of their environment as a single explanatory unit. Door-unlocking capacities putatively attributed to keys, are properly to be attributed to both keys and locks taken together. Similarly, such embedded cognitive capacities, though we might notionally attribute them to the cognitive agent, are properly to be attributed to both cognitive agents and their environments taken together.

§4.3 – CONCLUSION

In this chapter I have outlined a position on a number of related topics including para-information, embedded cognition, and cognitive dispositions more generally. Para-information, I have argued, makes a significant explanatory contribution to cognitive processing by shaping how an agent tackles a particular task through enabling the extraction of information. Embedded cognition, by extension, I characterised as cognition that is shaped in significant ways by para-information external to the standard organismic boundaries of the cognitive agent.

Drawing, in part, on Dretske, Dennett and Cummins, I firstly distinguished para-information from implicit information (as this term is usually understood) and discussed some internal and external examples, drawing out and detailing its central explanatory role in shaping how a cognitive agent tackles a cognitive task. The concept of para-information was then linked to embedded cognition via a general account of dispositions of which cognitive capacities are plausibly a species. Dispositions, I have argued, supervene upon the salient properties of the systems to which they are notionally attributed *together with* those of the systemic elements of their activating conditions. In the next chapter I begin the examination of *extended* cognition with a defence of the general claim that our cognitive processes and states sometimes extend beyond our bodies into external processes and states.

CHAPTER 5 – EXTENDED COGNITION

In the previous chapter we explored the cognitive role of para-information and the consequent extent in which cognition is often embedded in the environment of the cognitive agent. Over this and the following chapter I shall discuss and defend the central thesis of extended cognitivism: that certain processes occurring beyond the boundaries or a cognitive agent's body, are as much parts of that agent's cognitive processing as anything going on in their brain. In this chapter the focus will be on a defence of the *general* thesis of extended cognitivism – one that is silent on the underlying character of extended cognitive processing. In the next chapter, I shall focus on a stronger version of the thesis that includes an appeal to the *representational* character of certain processes and states external to the body of the cognitive agent.

After first providing a rough definition of extended cognition (§5.1), I distinguish it from some of its more or less close relatives (§5.1.1) and set aside certain potential misconceptions concerning extended cognitivism (§5.1.2). This will be followed by an attempt to situate extended cognition within a relatively orthodox tradition in cognitive science (§5.1.3). After then introducing certain distinctions and defining some important terms relating to cognitive processes, states and systems, I state in more precise terms the general thesis of extended cognitivism (§5.1.4). After examining (§5.2) a number of central putative examples of extended cognition, and drawing most prominently on the arguments of Fred Adams and Ken Aizawa, I turn (§§5.3-5.6) to an examination (and rejection) of four central objections raised against the thesis in the literature.

§5.1 – EXTENDED COGNITIVISM: CLARIFICATORY REMARKS

Before beginning the defence of extended cognitivism, it will be well to distinguish it from several more or less closely related positions, and in doing so, to identify certain claims that are *not* part (or at least, not an *essential* part) of the extended cognitivist position. An extended cognitivist might well be sympathetic to some of these related claims – for example, that cognition is *situated*, *embedded*, *embodied*, or *distributed* – some of which may even offer, with extended cognitivism, varying degrees of mutual support. But it remains vitally important that we be as clear as possible as to what is essential to the position to be defended here and what is inessential or invites confusion. Let us proceed, therefore, with a rough definition (a more precise definition will be given in §5.1.4) of our central term:

> ***Extended Cognition***: a cognitive process/capacity is *extended* into the environment to the extent that an agent's cognitive behaviour is shaped by an *informational process* that is external to the standard organismic boundaries of the cognitive agent.[89]

Extended cognitivism – the position to be defended over this and the following chapter – is, by extension, the claim that there are such cases of extended cognition. Authors who have defended this view include, most prominently, Clark and Chalmers (1998). That paper will serve as a useful focal point for the issues to be considered in this and the following chapter. Clark has remained particularly prominent in defending and developing

[89] In considering this definition it needs to be borne in mind that 'informational' is here used entirely neutrally between the truth-neutral and truth-dependent senses of information.

this view through a number of publications[90] whilst other proponents include Mark Rowlands (1999), Daniel Dennett (1996), and David Houghton (1997). Slightly more cautious endorsement for extended cognitivism is given by Sterelny (forthcoming). In addition, many further commentators including Ruth Millikan (1993), Merlin Donald (1991) and Robert Wilson (1994, 2004) have, in the course of defending a variety of different theses, been drawn in similar directions.

§5.1.1 Some More Or Less Closely Related Claims

Regrettably, extended cognitivism has sometimes been presented alongside a number of more or less closely related positions with which it is sometimes compounded, including those mentioned above: that (at least some) cognition is embedded, situated, embodied, or distributed. Properly evaluating the tenability of extended cognitivism is greatly complicated by a general lack of uniformity in the literature concerning the usage of such terminology. To see this, consider the concept of 'embedded' cognition that was examined in the previous chapter. A number of commentators have employed this concept but without clearly distinguishing it from related concepts. For example, in a paper entitled 'Mind Embodied and Embedded' John Haugeland (1995) does little if anything to clearly distinguish between the two. Similarly, Clark (1998) flits relatively freely between terms such as 'situated', 'embodied', 'distributed' and 'embedded', but leaves the demarcation lines between these concepts rather unclear.

Moreover, for many of these terms a number of significantly different uses can be found in the literature, some of which overlap with extended cognition as I have defined it above. To take, for example, the claim that cognition is 'embodied', Margaret Wilson

[90] See, for example, Clark (1997a), Clark (2001a), and Clark (forthcoming).

(2002) has pointed out no less than *six* distinct senses in which this claim has been advanced. They are: (1) cognition is situated; (2) cognition is time-pressured; (3) we off-load cognitive work onto the environment; (4) the environment is part of the cognitive system; (5) cognition is for action; (6) off-line cognition is body based.[91] Whilst there may well be something to the suggestion that these uses constitute an interrelated cluster of claims and concepts, their differences are perhaps more striking than their similarities. Indeed, in some cases it is unclear why the term 'embodied' is used at all. Why, for example, should time-pressured cognition be considered as 'embodied'? Similar variations in usage surround the use of 'situated'. For example, in the first of the senses identified by Wilson it is treated as being synonymous with 'embodied'. By contrast, others, such as Robert Wilson (2004), treat situated cognition as synonymous with embedded cognition and include under that banner at least most of the senses that Wilson (2002) identifies as falling under 'embodied' cognition.

In the light of such variations in usage and frequent lack of clarity, being precise and univocal will inevitably do some violence to how some commentators use the above terms. Notwithstanding this, the following general distinctions can be put in place.

It is perhaps 'situated' cognition that comes closest to constituting a catch-all term for the cluster concepts mentioned above. Generally speaking, it is used to emphasise the cognitive significance of the 'situatedness' of the cognitive agent a) within a body, b) within a physical environment, and c) within a social, cultural, and technological

[91] Wilson identifies these respective views with (amongst others) the following: (1) Clark (1997a), Beer (2000), Port and Van Gelder, (1995), Thelen and Smith (1994); (2) Brooks (1991b); Pfeifer and Scheier (1999), Port and Van Gelder (1995); (3) Kirsh and Maglio (1994); (4) Beer (1995), Thelen and Smith (1994); (5) Churchland, Ramachandran, and Sejnowski (1994); (6) Lakoff and Johnson (1999). As will become clear, my use of the term 'embodied' comes closest to (6) but without requiring that it be 'off-line'. Senses (3) and (4) are treated in the following two chapters as extended cognition, and embedded cognition comes closest to sense (1), although I treat 'situated' as a very broad term encompassing several of the others

environment. The broad position that emphasises the cognitive significance of such situatedness is summarised as a pair of claims by Andy Clark (1998, p. 506) as follows:

> (1) That attention to the roles of body and world can often transform our image of both the problems and the solution spaces for biological cognition.
>
> (2) That understanding the complex and temporally rich interplay of body, brain, and world requires some new concepts, tools, and methods – ones suited to the study of emergent, decentralised, self-organising phenomena.

The central claim is that a science of cognition cannot afford to treat the cognitive system as comprising merely a central nervous system in isolation from its real-world, real-time material, social, and physical situation; body and world (not merely brain and nervous system) play a central role in cognition. This, however, is a *very* broad and sweeping claim, and so it is within this broad position that the other more narrowly focussed (but, to some extent, overlapping) concepts and claims can be more effectively and usefully located.

Extended cognition fairly clearly falls within the general scope of situated cognition inasmuch as the world plays a significant role in it. But there are kinds of situated cognition that clearly are not extended cognition as I have defined it. Chief amongst these is 'embodied' cognition, which can be most clearly delineated as cognition in which the *physical body* of a cognitive agent plays a cognitively significant role, over and above those bodily features traditionally associated with cognition – i.e., the brain and nervous system. To take an example from Haugeland (1995), his capacity to type or to tie his shoes depends on his specific bodily contingencies over and above the issuing of a certain pattern of electronic signals from his brain:

> In the first place it depends on the lengths of my fingers, the strengths and quicknesses of my muscles, the shapes of my joints, and the like. . . . [T]here need be *no* way – even in principle, and with God's own microsurgery – to reconnect my

> neurons to anyone else's fingers, such that I could reliably type or tie my shoes with them. (Haugeland, 1995, p. 225.)

Such views sometimes draw on Heidegger (1926/1962), especially on his notion of 'skilful coping' as essential or basic to cognition, and Merleau-Ponty (1942/1963).[92] Extended cognition contrasts with embodied cognition in that the former, unlike the latter, concerns cognitive processes that occur *beyond* the traditional organismic boundaries of the cognitive agent.

A more closely related concept is that of 'distributed' cognition, which is typically characterised as cognition that supervenes upon multiple cognitive agents or upon collections of cognitive agents together with their technological or cultural environment. The view that certain cognitive processes are distributed (sometimes called the 'group-mind hypothesis') has received most recent attention through the work of Ed Hutchins (1995) and typically characterises such cognition as an essentially or significantly *social* or *cultural* activity occurring *amongst* cognitive agents rather than simply *within* cognitive agents.[93] As Hutchins puts it when discussing ship navigation considered as a cognitive system:

> not all the representations that are processed to produce the computational properties of this system are inside the heads of the quartermasters. Many of them are in the culturally constituted material environment that the quartermasters share with and produce for each other. (Hutchins, 1995, p. 360.)

Like extended cognitivism, this view emphasises the significant cognitive role of informational processes occurring within one's technological environment. But unlike extended cognition, distributed cognition does not focus on the *individual* cognitive agent,

[92] Heidegger's influence is especially strong in, for example, Dreyfus (1972) and Wheeler (1996). Other proponents of specifically embodied views – although not explicitly Heidegerian – are Lakoff and Johnson (1999). In their view metaphors are the life-blood of cognition, and those metaphors are themselves structured by the contingencies of our bodily situation in the world. The details of the Heideggerian view remain, I regret to say, largely beyond the ken of this writer, despite several attempts to come to grips with it. Merleau-Ponty is especially influential in Varela, et al. (1991).

[93] Another proponent of this view is Wilson (2001, 2004).

but instead recasts cognition in non-individualistic terms. Distributed cognition, its proponents argue, supervenes on *collective* entities over and above the cognition of individuals – collections of cognitive agents, together with their technological environment and cultural practices.

Embedded cognition has already been discussed at some length in the previous chapter and it clearly shares with extended cognition some of the same distinguishing features: it is likewise focussed on individual cognitive processes (contra distributed cognition) and on the world beyond the traditional organismic boundaries of the cognitive agent (contra embodied cognition). But whilst embedded cognition may be shaped by certain environmental states or structures beyond the traditional organismic boundaries of the cognitive agent, this in no way entails the inclusion of such states or structures as constitutive elements of an individual's cognition. To draw an analogy, the path that a mountaineer follows in the process of climbing a mountain is directly shaped by the physical structure of the mountain, but this does not imply that the *mountain* is either climbing or mountaineering. It is largely in this respect that embedded cognition is distinguished from extended cognition. Extended cognitivism claims that there are external processes that are quite literally part of the cognitive processes of individual cognitive agents.

§5.1.2 Some Possible Misconceptions

In addition to the above conceptual confusions, there are a number of occasional misconceptions about extended cognitivism that deserve to be scotched. There is, for example, a very commonplace view with which it might, at first glance, be confused. This view is sufficiently uncontroversial as to lack a received name; in lieu of such I shall refer to it simply as the 'Commonplace View'. The Commonplace View is that human cognitive processes often operate with the aid of external resources, most prominently with content-bearing representations external to our skin. There is a veritable panoply of such representational devices, including maps, pictures, written sentences, computer printouts, graphs, and so on, that serve as paradigmatic examples of outside-of-the-body content-bearing representations. Moreover, I take it to be equally uncontroversial that such representations frequently play *a* role in guiding or influencing considerable portions of our cognitive behaviour. The central difference between the Commonplace View and extended cognitivism lies in the status of both the external representations and the operations or process that we engage in with respect to them. Extended cognitivism, but not the Commonplace View, treats such external representations and the processes that operate over them as *fully and genuinely cognitive*. Such representations are not mere 'crutches' to cognition, as it is sometimes put, nor as mere informational resources, but are cognitive states of the agents who use them in as full and legitimate a sense as certain of their neuronal states are standardly taken to be. Similarly, certain operations or processes involving these external states or structures, such as rotating a map so as to orient it in line with the world, are viewed by the extended cognitivist as proper parts of cognitive processing.

Another misconception might be that the extended cognitivist is somehow denying or otherwise significantly downplaying the centrality of the role of the brain in cognition. Nothing could be further from the truth. The extended cognitivist does not in any way deny either the vital importance or the centrality of in-the-head brain-bound cognitive processing. Neither is the extended cognitivist suggesting that there are ever cognitive states or processes that acquire that status independently of in-the-head cognitive processes or states. What distinguishes extended cognitivism from more standard approaches is a rejection of the generally held assumption that cognition is *completely exhausted* by states or processes that exist entirely within the brain or body. It suggests, by contrast, that there are external states or structures that serve as potential or actual constituents of an individual's cognitive states, and, moreover, that manipulations of those states or structures are sometimes constitutive components of an agent's cognitive processes. Cognitive processes and states, suggest Clark and Chalmers (echoing Putnam), ain't *all* in the head.

On that no doubt familiar note, there is a well known relative to extended cognitivism in semantic externalism. Semantic externalism suggests that the contents of representational states are individuated, in part, by reference to external features of the physical or social environment.[94] Both semantic externalism and extended cognitivism make an important appeal to states entirely external to the brain or skin of the cognitive agent. But semantic externalism is a weaker thesis that is ultimately neutral on the location of content-bearing states; there is nothing in it to suggest that cognitively relevant content-bearing states are themselves to be found anywhere outside of the head. This is not to say that there are no interesting connections to be drawn between semantic externalism and extended cognitivism. Wilson (1994), for example, has argued against a certain computationalist argument for semantic internalism on the grounds of the possibility of what he calls 'wide computation' – a view that is, to all intents and purposes, a specifically computational version of extended cognitivism. Notwithstanding such connections, semantic externalism neither entails nor is entailed by extended cognitivism.

[94] For example, Putnam (1975), especially chapters 8 and 12, and Burge (1979).

§5.1.3 Some Historical Context

As surprising as this may sound to some, extended cognitivism, at least in the version defended here, can be seen as a very natural extension of orthodox cognitive science.[95] To see this, consider Turing (1950), regarded by most commentators as one of the founding contributions to cognitive science. In that paper Turing speculated that human thinking can be simulated or implemented using computations over symbolic representations in a digital computer. Turing's influence lies primarily in two theses that subsequent cognitive scientists were to draw from this idea (neither of which does Turing directly draw). The first was that human thinking *is* simply such computation and (thus) that the mind *is* a digital computer. This, in turn, gave rise to two parallel research projects: the first, to try to explain human cognition in terms of such computations, and the second, artificial intelligence – the attempt to get computers to reproduce certain aspects of human cognition. What is, perhaps, ironic is that cognitive science, whether in a traditional computationalist mould or otherwise, has only relatively recently woken up to the possibility that thinking might also be comprised of processes that *span* both humans and computers (or, for that matter, other information processing devices). In an important respect extended cognitivism is an attempt to substantiate this claim; that is to say, in the specific version defended in this dissertation, that thinking *is* (at least sometimes) comprised of processes involving *both* humans *and* information technology.[96]

The second important thesis drawn from Turing's idea founded functionalism. Functionalism claims, as it were, that cognition *is* as cognition *does*. In other words, aside

[95] This is not, of course, to say that its claims will not be startling.

[96] Although the arguments presented in this dissertation are most persuasive when seen in computationalist terms, extended cognitivism can sustain non-computationalism readings. See for example, Bechtel (1997) for a connectionist version.

from its independent scientific interest, it doesn't matter *how* cognition is implemented – what matters is the role it plays within the system.[97] Cognitive processes (and the states involved in such processes), functionalist have argued, are functional kinds. As such, any two processes that can solve the same cognitive problems is a cognitive process of the same type regardless of whether it is implemented in neurons or silicon.[98] The Turing Test, can be seen as an application of this functionalist principle. More to our present concerns, it is manifestly in the full spirit of functionalism that Clark and Chalmers offer what they describe as their *parity principle*:

> 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 recognising 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! (Clark and Chalmers, 1998, p. 8)

This principle is the argumentative heart of Clark and Chalmers (1998). Where disputes arise between advocates of extended cognitivism and critics largely concerns whether the parity is ever satisfied.

§5.1.4 Processes, States, and Systems

Before we proceed to an examination of some examples, there are a number of general relations involving cognitive processes, cognitive states, and cognitive systems (cognitive objects) that we will need to place out in the open. These relations have been left somewhat implicit in the dissertation up to this point, but will now need to be stated explicitly. As will become clear when we turn to some of the criticisms raised against extended cognitivism, many such objections founder through failing to distinguish between, or to be sensitive to, relations between cognitive processes, states, and systems.

[97] Of course, the most direct influence of Turing was on 'machine state' functionalism, such as in Putnam (1975), chapter 18. These modelled their views more closely around the concept of a Turing machine. The notion of functionalism to which I am here drawing attention is often called 'psycho-functionalism'.

To begin with, and speaking quite generally, we can say that a *process* inheres in a *system* (a *systemic object*) and is comprised of a succession of *states* of that system. Expanding on this a little, we shall understand a system (or systemic object) to be spatio-temporally extended and material. Depending on how one carves one's ontology, a system (e.g., a car engine) may be arbitrarily seen as a single object (an engine), a part of an object (a car), or a collection of objects (a combustion chamber, pistons, etc.). The significant ontological unit of present concern, however, is the systemic object considered as a single system and fixed in relation to a given process. On this model, a *state* is a spatially extended time-slice of a system. A *process* consists of a succession of such states with an (often more or less vague) beginning and an (often more or less vague) termination.

Focussing slightly more specifically, where the process is *goal*-directed (for instance, cognitive processes) the termination state will be a 'target' state or 'goal' state for that process. For example, the process of making a cup of tea typically begins somewhere around getting up to go to the kitchen to put the kettle on, and finishes somewhere around the appearance of the goal state – the presence of a hot cup of tea that is ready to drink. Such goal-directed processes acquire a *functional* description (kettle-filling, tea-making, multiplication, etc.) according the function that they fulfil. Moreover, such a goal-directed process P may (and typically will) be amenable to decomposition into *functional sub*-processes, P_1, P_2, P_3, etc., that may be either concurrent or sequential constituents of P. For example, making a cup of tea (at least in my house) can be decomposed into filling a kettle with water, plugging the kettle in, etc. Other processes may accompany P without being sub-process of P if they make no functional contribution

[98] Or even populations of Chinese – see Block (1999).

to *P*. For example, scratching one's elbow or chatting with a friend whilst waiting for the kettle to boil will *not* be sub-processes of making a cup of tea precisely because they make no functional contribution to fulfilling that goal.

Note, moreover, that the functional description *independently* attributable to a sub-process (that is, outside of its particular role in *P*) will generally *not* be the same as the functional description for the process of which it is a part. Yet the sub-process remains a part of the broader process all the same. Filling a kettle is not *in itself* a tea-making process, but it is certainly *part* of a tea-making process.

The next point is central and should be clear given Clark and Chalmers' parity principle: the claim of extended cognitivism is, first and foremost, a claim about cognitive *processes*. All other related claims concerning such things as cognitive states and cognitive systems (or, as I shall sometimes put it, systemic cognitive objects), owe their cognitive character and status to the relation of such claims to claims about cognitive processes. For example, a state gets to be a *cognitive* state in virtue of being part of a cognitive process; a system gets to be a *cognitive* system because of a cognitive process that inheres in that system.

With these points in mind, and taking 'cognitive process' to be primitive, let us introduce the following definitions. For some cognitive process *P*:

P_s is a ***sub-cognitive process*** iff P_s is a functional constituent sub-process of *P*.

As just pointed out in more general terms, it is important to emphasise that P_s need not itself be *independently* a cognitive process, any more than filling a kettle with water needs to be independently a tea-making process. To suppose that all sub-processes of a process of kind *K* must also be independently of kind *K* is to commit a kind of fallacy of division. It is for this reason that I choose the term 'sub-cognitive process' rather than, say, 'cognitive sub-process' – the latter phrasing suggests that the sub-process will have an independently cognitive character.

Equally, P_s not being independently a cognitive process no more impugns the status of P_s as *part of a cognitive process* than not being independently a tea-making process impugns the status of filling a kettle with water as part of a tea-making process. To suggest otherwise would be akin to what Ryle (1949) calls a category mistake. To take one of his examples, a military march-past may consist of the battalions, batteries and squadrons *of* a division. But the military division is not some *extra* unit alongside its elements, nor for that matter, is it any one of its parts (see especially, Ryle, 1949, pp. 17-18). Similarly, it would be a mistake to suppose that a cognitive process consists of anything over and above its (perhaps, when considered independently, non-cognitive) sub-processes. The central lesson is that the cognitive status of certain processes are attributable *holistically* to entire processes according to their performance of some cognitive function. Sub-processes within the broader process need not be independently of the same functional kind as the broader process, but they *will* be of that broader kind *derivatively* – that is, in virtue of their constitutive functional roles within the broader process. None of the sub-processes need themselves be fulfilling a cognitive function, though of course in many cases they will. It is for this reason that someone filling a kettle can quite truthfully respond, if asked what they are doing, that they are making a cup of tea. With this in mind, we can define a systemic cognitive object as follows:

O is the ***systemic cognitive object*** (= *cognitive system*) for *P* iff *O* is the system in which *P* inheres.

It is worth noting that talk about 'the brain' as a systemic cognitive object (=cognitive system) is, therefore, largely a matter of convenience only. From the point of view of *particular* cognitive processes, such as in-the-head arithmetical calculations, the *entire* brain, I take it, is not the object of interest and does not constitute a cognitive system. Under the above definition, those parts of the brain not involved in this or that cognitive process will not be a part of the cognitive system for this or that cognitive process. A systemic cognitive object consists of just those parts in which a given cognitive process inheres. With that in mind:

O_p is a ***partial cognitive object*** iff O_p is a part of a cognitive system, *O*.

Since, as noted above, sub-cognitive processes need not be independently cognitive to be a part of a cognitive process, partial cognitive objects need not themselves constitute systemic cognitive objects. Whether they do, in fact, constitute systemic cognitive objects depends on whether the sub-process that inheres in the partial cognitive object is independently a cognitive processes. At least in general, an individual neuron is not, I take it, a systemic cognitive object precisely because there will not be a cognitive process that inheres in that single neuron. As with a kettle in relation to making a cup of tea, none of this impugns it being part of a cognitive process.

Paired with systemic and partial cognitive objects, we also have the following:

S is a ***systemic cognitive state*** iff S is a state of a systemic cognitive object, O.

S_p is a ***partial cognitive state*** iff S_p is a state of a partial cognitive object, O_p.

The central claim of extended cognitivism can now be expressed more precisely. In its general form its central thesis amounts to the following (a stronger, specifically representational, version will be examined in the next chapter):

> **Extended Cognitivism:** Some cognitive processes of an individual cognitive agent include sub-cognitive processes that lie outside the standard organismic boundaries of the cognitive agent.

This claim, in turn, leads to a number of interrelated claims, including a) the systemic cognitive object for certain processes includes not only the cognitive agent (traditionally conceived) but also objects in the external world, and b) in certain cases, external objects (such as pocket calculators and notebooks) will be partial cognitive objects with partial cognitive states, precisely to the extent that they participate in cognitive processes of an individual cognitive agent.

§5.2 – SOME EXAMPLES

With these clarifications in mind, let us begin by considering a few putative examples of extended cognition from the extended cognitivist's standpoint.

§5.2.1 Pen and Paper Addition

Consider a commonplace example of problem-solving: adding two three-digit numbers together.[99] The goal-state here is to arrive at a correct answer for the addition of those two numbers, and the explanatory goal is (roughly) to explain how we manage to get the right answer. There are several methods that one might use to solve such a problem. One method is to engage in a bit of 'mental arithmetic', that is, to do the addition *all in one's head.* Other methods involve the use of various kinds of information technology: a computer, a pocket calculator, an abacus, or pen and paper. Consider the last of these – using pen and paper. This involves a finite number of iterable steps. First, the digits of the two numbers are written down in a specific form (units over units, tens over tens, etc.). In order to keep the digits to be summed distinct from their sum, a line is usually drawn underneath the digits. Next, starting with the rightmost column, the human cognitive agent initiates an internal process – an addition of the pair of digits in that column. Having figured out the additive result of these two digits the result is then written below the line, rightmost digit under rightmost column, remembering to carry, if necessary, into the next column. This process then turns to the next column to the left, and so on, until all digits are added. What we have provided here is a first-gloss functional decomposition for that process as it occurs using pen and paper. The functional role of the pen is as a recording device writing a symbol to a particular location. The paper functions as a medium for storing the external representations. Together, they function much as does short-term memory when undertaking in-the-head arithmetical calculations – recording and storing information for later retrieval and subsequent processing. Many cognitive scientists, especially those endorsing a computationalist framework, have suggested that human thinking is essentially like that – the execution of a 'program'. Notwithstanding whether internal human thinking is computational in exactly that way, the important point to emphasise is that under a given functional description of the task, whether we solve the problem using just neurons or using pen and paper (plus neurons) is a matter largely of implementational detail. What makes these methods essentially the same cognitive process is a matter of the equivalent functional role each plays in an agent's cognitive economy. For a given range of arithmetical problems, we can solve the same problem using either our brains in isolation from the world, or using our brains in conjunction with the world. Hence, given a broadly functionalist view together with the parity principle, we *ought* to say that since using only one's neurons to add two three-digit numbers is a cognitive process, so is doing it using pen and paper (plus one's neurons).

§5.2.2 Tetris and Scrabble

Clark and Chalmers defend their position by considering three methods of human problem-solving drawn loosely from the game of Tetris. In each case the problem-solving task is to answer questions concerning the 'fit' of various two-dimensional shapes that appear on a computer screen into 'sockets' also appearing on the screen. The three methods are as follows:

(1) The cognitive agent mentally rotates the shapes in their heads so as to imaginatively align them with the sockets.

[99] Arithmetical calculation examples appear in many discussions both computationalist and otherwise For example, Johnson-Laird (1993), chapter 3, Rumelhart, et al (1986), p. 46, and Clark (1989), p. 133. It is put to use in the service of a more externalist cause in, for example, Wilson (1994) and Bechtel (1997).

(2) The cognitive agent can choose either to mentally rotate the shapes as above, or else to physically rotate the image on the screen by pressing a rotate button. (The latter method, Clark and Chalmers not unreasonably suggest, will likely be faster.)

(3) The cognitive agent is armed with a neural implant that can perform the rotation operation as fast as rotating it on screen. The cognitive agent must choose whether to use good old-fashioned mental rotation or the implant as each makes different demands on attention and other concurrent brain activity.

Clark and Chalmers suggest that the differences between these methods are superficial with respect to their status as cognitive processes, and that (hence) each case ought to be considered as cases of cognitive processing. Case (1) is clearly uncontroversially cognitive and involves standard in-the-head cognition. Case (3), they suggest, is on a par with (1) inasmuch as it, too, is entirely in the head. And case (2) involves the same sort of computational structure as case (3). The only particularly interesting difference is that in case (2) *the processing is distributed between agent and computer* instead of occurring entirely within the boundaries of the brain or skin of the agent. Case (2), they claim, is a case of extended cognition.[100]

Consider another example that Clark and Chalmers mention: the playing of a game of Scrabble.[101] The primary cognitive activity involved in Scrabble is the turn-by-turn search for the 'best' move as judged by a variety of factors including point score, letter distribution, and so on. Leaving the evaluative aspects to one side, the process of finding

[100] A reminder – Clark and Chalmers use the term 'active cognition' to describe such cases. I shall continue to use 'extended cognition' and treat the phrases as interchangeable.

[101] An example they draw from Kirsh (1995)

the best move will involve the exploration of the possible tile combinations available in one's rack of tiles. Of these only a relatively small sub-class will constitute 'solutions' – that is, a way of ordering the tiles in one's rack such that at least some of them form a valid word or words in conjunction with the current state of the board. One of these solutions will be judged the best play using some of the evaluative criteria just mentioned. Paralleling Clark and Chalmers, we can imagine this cognitive task being undertaken in a couple of different ways.

In the first case, the player would 'shuffle the tiles in their head'; the cognition in this scenario involves the manipulation of various cognitive elements – cognitive representations of the tiles – through various combinations and re-combinations, until, after a certain time, a best solution is determined and played. In this case we would likely have no problem in characterising this process as cognitive, and there is, prima facie, strong grounds to suppose that this involves cognitive representations. The cognitive representations and the operations performed upon them are used in place of the physical tiles and the possible ordering operations that might be performed upon them, and to that extent function as representations of both tiles and possible operations on them.

The second method (and a far more common, precisely because of its greatly reduced neuronal load) is for the player to physically move the tiles around on the rack, noting various solutions (legal words) as they go, either internally in neuronal memory, or perhaps on paper. This process continues until a best solution is determined and played. Often this will involve the use of various search stratagems, such as separating off common word endings such as '-ing' or '-er' as an aid to efficient exploration of the possibilities. This method is a paradigm case of extended cognition and what Kirsh and Maglio (1994) call an 'epistemic action' – that is, the structuring of the environment to facilitate cognition (as distinct from 'pragmatic action' which is undertaken primarily in order to approach some physical goal). In a computerised version of Scrabble, there is even a button that randomly shuffles the tiles.

§5.2.3 Points of Agreement

Before we launch into a consideration of critical objections to extended cognitivism, we should note that there is a large measure of agreement on the empirical facts concerning such examples. Critics have not been shy in conceding that human cognition is manifestly and sometimes profoundly enhanced by the use of such 'props' or 'cognitive tools' as written language, notepads, and computers. For example, Adams and Aizawa acknowledge that 'calculators, slide rules, and computers provide tools that enable us to perform logical and mathematical operations more quickly and reliably than we might when relying exclusively on the limited resources in our brains' (Adams and Aizawa, 2001, p. 44). Rupert (2004, p. 393), for his part, raises no objection to the 'hypothesis of embedded cognition' which he describes as the view that 'cognitive processes depend *very* heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place'. For example, it is readily agreed that adding numbers using pen and paper has several advantages over doing it in one's head. As Adams and Aizawa point out:

> Since the numbers can be written one above the other, one can rely on vision to keep the ones, tens, and hundreds places coordinated. In addition, since one can write down the number to be carried above the column to which it will be carried, this would remove the burden of remembering the number to be carried. Further, by recording one's work at each step, one is spared the task of remembering where one is in the calculation . . . . It is because the use of pencil and paper generally provides a faster and more reliable method of computing products that one so frequently turns to it. (Adams and Aizawa, 2001, p. 43)

Aside from the extended cognitivist adding that it is not our *vision* on which we rely to keep the numbers straight, but the *paper*, the above can be agreed on by all parties to the dispute. Cases such as addition, where our neuronal cognitive load is lightened by the use of information technology, are neither isolated nor are they a particularly recent development in human cognition. There are a large number of cognitive tasks in which (as Dennett (1996) puts it) 'offloading' some of the work into the environment renders the task considerably more manageable.[102] Often the task is entirely quite impossible to solve using neurons alone but easy with the aid of informational props or technologies. Playing chess is within the capabilities of most people so long as they have the position clearly in view, but playing chess in one's head is considerably more difficult. It strains the neuronal cognitive capacities of most people to hold a complete chess position in their head, let alone to analyse the position for the best move. Conversely, playing chess with an 'analysis board', on which one can 'test run' various possible moves, increases one's cognitive effectiveness considerably.

Where critics and protagonists diverge largely concerns how to *interpret* these facts. Where critics draw the cognitive line concerns whether the states and processes of such cognitive tools are *constitutive* components of the cognitive states or processes of the individual who uses them. Similarly, all critics (so far as this author is aware) concede the reasonableness of the parity principle (although, as we shall see in §5.5.2, certain arguments pressed against extended cognitivism appear to be at odds with the acceptance of the principle). Yet, for a variety of reasons, critics deny that this parity is ever actually satisfied. They resist the inference that strikes proponents of extended cognitivism as

[102] See Dennett (1995), p. 176ff. Proponents would, of course, emphasise that this is *cognitive* work that is offloaded, whilst critics would, presumably, demur.

obvious: that the processes and states of these external 'props' or 'cognitive tools' are as much a part of our cognition as anything that goes on purely inside our heads.

§5.3 – THE COUPLING-CONSTITUTION FALLACY

The first objection we consider, featuring particularly prominently in the critical armoury of Adams and Aizawa is the 'coupling-constitution fallacy'. This fallacy, they suggest, 'is the most common mistake that extended mind theorists make' (Adams and Aizawa, forthcoming a, p. 2). In general terms, they characterise the coupling-constitution fallacy as

> a tacit move from the observation that process X is in some way causally connected (coupled) to a cognitive process Y to the conclusion that X is part of the cognitive process Y. The pattern of reasoning here involves moving from the observation that process X is in some way causally connected (coupled) to a process Y of type Φ to the conclusion that X is part of a process of type Φ.[103]

To illustrate the fallaciousness of such a move they give a number of examples including the following:

> Consider the bi-metallic strip in an ordinary thermostat. The expansion and contraction of this strip is closely coupled to the ambient temperature of a room and the air conditioning apparatus for that room. Nevertheless, this gives us no reason to say that the expansion and contraction of the strip extends beyond the limits of the strip and into the room or air conditioner. The Watt governor provides another example. The combustion of fuel in the governed engine is tightly coupled to the rotation of the weighted arms, yet the process of combustion does not extend beyond the bounds of the engine. (Ibid.)

Adams and Aizawa are entirely correct to suggest that the above examples clearly invoke a fallacious inference. But, so I argue below, they are *not* correct to suggest that the extended cognitivist makes this kind of move. Their mistake occurs on three fronts.

[103] Adams and Aizawa, forthcoming b, p. 8. A similar objection is raised by Keith Butler (1998). Referring to a specifically computationalist example in Wilson (1994), Butler suggests that: 'the only computation that is going on is going on on the inside. To the extent that the system is computing anything, it is only the internal portion that is doing any computing. It is the internal mechanisms that drive the process. The . . . external environment functions merely as input to the only system that is doing any computing, namely, the nervous system of the animal in question.' (Butler, 1998, p. 213.)

§5.3.1 A Central Misunderstanding

In the first place, Adams and Aizawa demonstrate an important misunderstanding of the basis of the extended cognitivist's position. They mistakenly suppose that the extended cognitivist argues for the cognitive status of certain external processes on the basis of their close *causal attachment* to *internal cognitive* processes, much as a cold might spread beyond an individual's body by close causal contact with other individuals. But this no more drives the extended cognitivist argument than it drives equivalent attributions in wholly internal cases. No one, I take it, supposes that the grounds for considering internal short-term memory processing as cognitive are its close causal attachments to other internal cognitive processes, such as internal deductive inference. In point of fact, the extended cognitivist's basis for considering certain external processes cognitive stems from their playing *an equivalent functional role in explaining cognitive behaviour* as internal cognitive processes. This is exemplified by the parity principle in which external processes are identified as cognitive not in virtue of close causal attachment to internal cognitive processes, but in virtue of being a functional equivalent of a sub-cognitive part of an internal cognitive process.

§5.3.2 Processes and Functional Descriptions

Even were Adams and Aizawa correctly understanding the externalist's motivations, their arguments fail on two further counts. To see this let us recall a general point made in §5.1.4: goal-directed processes generally, and cognitive processes in particular, are identified according to a functional description. Typically this will involve sequences of intermediate processes, but the functional description attributable to the broader process will not (or at least, not *generally*) be *independently* attributable to the sub-processes that comprise the broader process. A making-a-cup-of-tea kind of process is not a putting-the-kettle-on kind of process, nor (at least independently) vice versa. The process of walking, for example, comprises many sub-walking processes, *none* of which is itself a walking process. *Walking* is not something independently performed by the tensioning of an Achilles tendon, nor the contraction of a leg muscle, not the firing of a nerve fibre, etc.: it is a process performed by the entire *system* comprising these elements (and many more).

Given this, it becomes vitally important that we be clear as to *which* cognitive behaviour we are identifying and at which level of functional decomposition. In quite general terms, it depends on what we are interested in explaining. For example, if we are interested in explaining *heat generation* (in an engine, say), then we need only look in the combustion chamber because the process that explains that behaviour (*combustion*) takes place there. *That* process begins with the ignition of the fuel and ends when the fuel is spent or the oxygen exhausted and doesn't extend beyond the combustion chamber. But now suppose we are interested not in heat generation but in *power generation*. In that case looking within the combustion chamber will only provide a *part* of the explanation, because only a *part* of the process of power generation is going on in there (the heat-generating part). There is no independent power generation going on in the combustion chamber. The process of power generation is a process going on *throughout the engine*: in the combustion chamber, in the pistons, the drive-shaft, the fly-wheel, etc.

The central point is this: the general character of the explanatory processes as well as where we locate those processes depends on what we are interested in explaining. In the case of cognition, we are generally interested in explaining how a certain task or problem is solved. In the case of the three central examples mentioned above we are interested in explaining (respectively) the capacity to a) add two three-digit numbers, b) determine the fit of shapes to sockets, and c) find a good word in Scrabble.[104] If we are interested in a sub-cognitive process for that process, such as how two digits are summed, we may find that we need appeal only to neuronal processes occurring entirely within the head of the cognitive agent. However, if we are interested in explaining how two three-digit numbers are summed, we may well find that appealing to processes occurring entirely within the head will provide an inadequate account. Rather, we may find – and in the case of addition using pen and paper the extended cognitivist says that we *will* find – that we will have to appeal to processes going on outside the head of the cognitive agent. The process of adding two three-digit numbers together begins with the apprehension of the addition problem and ends with the determination of an answer. No mere part of that whole process (whether it is performed entirely in the head or with the aid of pen and paper) is itself the process of adding those two three-digit numbers together.

Adams and Aizawa's first error, therefore, stems from focussing at precisely the wrong level of functional decomposition to appropriately capture the cognitive processes in the putative examples. Cognition, in the central examples, is more appropriately analogous not to the expansion or contraction of a bi-metallic strip, nor the combustion of fuel, but to the process of *maintaining a constant room temperature* and the process of *power generation*, respectively. The expansion and contraction of bi-metallic strips is not, independently of its role in the broader process, a maintaining-constant-room-temperature process, but only a *part* of such a process. Similarly, the combustion of fuel is not, independently of its role in the broader process, a power-generating process, but only *part* of such a process. Adams and Aizawa's confusion may stem from the fact that, somewhat

[104] I am assuming here that it is success that is of primary interest. In many cases, of course, failures will be

unusually for most functional decompositions, cognition is a process whose various sub-processes will often be functionally characterised as cognitive. For example, adding together in one's head two digits appearing in the right-hand column of a pair of three-digit numbers is as cognitive as adding together the entire sum.

§5.3.3 Two Senses of Cognitive Object and the Object/Process Distinction

Yet a third mistake that Adams and Aizawa occasionally make in discussing the coupling-constitution fallacy, is to slide between a *process* and the *objects* that participate in a process. To see how this occurs recall that extended cognitivism's central claim is that an individual's cognitive processes sometimes include sub-cognitive processes that extend beyond the boundaries of their brain or skin. The thesis is, at bottom, a claim about *processes*, not about *objects*. As noted in §5.1.4, objects count as cognitive on the extended cognitivist account *only* in relation to particular processes that inhere in them.

With this in mind, note how Adams and Aizawa deal with Clark's example of writing an academic paper (which Clark (2001, p. 132) suggests is the result of the system comprising his brain and the computer). They concede as common ground that the brain and the tools are jointly responsible for the product, the journal article. However:

> This . . . does not require that both the brain and the tools constitutes a single *cognitive* process. It is the interaction between the spinning bowling ball and the surface of the alley that between them lead to all the pins falling. Still, there is no 'extended bowling ball' meshing with the alley nor do we see any particular *intimacy* between a bowling ball and the alley.[105]

So far as the extension of the bowling ball is concerned, the extended cognitivist would not dream of suggesting otherwise. The bowling ball, after all, is an *object* – an analogy with a (let's suppose, systemic) cognitive object such as the brain. But no one is

of equal or even greater cognitive interest.

suggesting that *brains* extend into the world! What, by that analogy, the extended cognitivist *is* saying is that there is a *bowling process* that inheres amongst all the objects Adams and Aizawa mention. The bowling ball would be a 'partial bowling object' for that process. By the same token, neither is the extended cognitivist suggesting that the *spinning* of the bowling ball extends beyond the bowling ball – it is surely only the bowling ball that is spinning. Nevertheless, the *bowling* process involves more than the spinning of the bowling ball; it includes processes of surface and air friction, together with other mechanical processes from biomechanical processes of the bowler through to those of the colliding pins.

§5.4 – INTRINSIC VS. DERIVED INTENTIONALITY

A second objection pressed by Adams and Aizawa, and aired also by several other critics (for example, Butler, 1998, p. 180), concerns the distinction between *intrinsic* (sometimes called 'original' or 'non-derived') and *derived* intentionality (or content). This distinction carves intentional or content-bearing entities or states into two kinds. Standardly, those entities that are *intrinsically* intentional, that is, intentional *in and of themselves*, are said to have *original* intentionality. Conversely, those that owe their intentionality to some entity with original intentionality (or to their originally intentional states) have *derived* intentionality. A standard example of the former would be someone's (internal) belief that it is raining, and of the latter, the content of 'es regnet' as meaning that it is raining (whose intentionality is said to be derived from the linguistic use of Germans to that effect). This distinction is taken to threaten extended cognition on the basis of two claims. Firstly, the claim goes, only in the head (or at the very least, in the

[105] Adams and Aizawa (forthcoming b), p. 8. The mention on 'intimacy' is a reference to a term used by

body) states or processes have intrinsic intentionality. All external intentional states are standardly said to derive their intentionality from the intentionality of certain intrinsically intentional brain/mind states. As Adams and Aizawa put it:

> Strings of symbols on the printed page mean what they do in virtue of conventional associations between them and words of language. Numerals of various sorts represent the numbers they do in virtue of social agreements and practices. The representational capacity of orthography is in this way derived from the representational capacities of cognitive agents. By contrast, the cognitive states in normal cognitive agents do not derive their meanings from conventions or social practices. Despite possible interpretationist perversions to the contrary, it is not by anyone's convention that a state in a human brain is part of a person's thought that the cat is on the mat.

Let the above be as it may. By itself this claim poses no insuperable threat to extended cognitivism. The difficulties for the extended cognitivist arise from combining this with a second claim: 'that cognitive states must involve intrinsic, non-derived content.'[106] Suitably interpreted (see below), it is the combination of these claims that threatens extended cognitivism by restricting the cognitive to internal intrinsically intentional states.

These two claims open up three strategies for resisting the objection. The first, and perhaps the most radical, is to deny that there really is any such distinction to be made. The second is to accept the distinction and accept that external intentional objects are all derivatively intentional, but to reject claims that no intrinsic content is involved. The third strategy is to reject the supposed significance of original intentionality as a 'mark of the cognitive', as Adams and Aizawa put it. Although I shall here prefer the second and third strategies, the first strategy is at least arguable and is worth pausing to consider.

Haugeland (1998), p. 217.

[106] Adams and Aizawa (2001), p. 48. There are many who echo this. Fodor (1987), p. 99, for example, is prepared to insist 'that only mental [i.e. brain-bound] states should turn out to have semantic properties *in the first instance*'. Searle, for reasons very much at odds with Fodor's, makes similar claims. See Searle (1992) and elsewhere.

§5.4.1 First Strategy: Denying the Difference

The most famous advocate of the first strategy is Daniel Dennett.[107] In various places he has emphasised his contention that there is no fundamental difference between original and derived intentionality.[108] *All* our intentional states are derived, hence:

> A shopping list in the head has no more intrinsic intentionality than a shopping list on a piece of paper. What the items on the list mean (if anything) is fixed by the role they play in the larger scheme of purposes. (Dennett, 1987, p. 318.)

Dennett sees the demand for original intentionality as a member of a species of demands that assume that some property must be either grounded in some basic or original version of that property or else lead to an infinite regress of 'as if' properties. An example of this style of thinking might be regarding *being a thing of value*. It is often supposed that unless there is something of intrinsic value, everything will have only 'as if' instrumental value, never *real* value. That a choice between infinite regress or foundational property is not the only option can be illustrated by the property of *mammalhood.* Dennett suggests that a mistaken argument might go as follows: every mammal has a mammal for a mother, hence there must have been an entity that had *intrinsic* mammalhood from which all other mammals are descended. This, Dennett suggests, is a mistake inasmuch as Mammalhood is a property with vague boundary conditions and hence one that 'bleeds' into our evolutionary history with no clear 'prime mammal'. Dennett is not denying that all mammals must have had a non-mammal for an *ancestor* (it is not intentional states 'all the way down'), only that the transition from non-mammal to mammal was gradual and spread across a number of generations rather than across a single generation. In place of such reasoning Dennett suggests a *finite* regress. In the case of intentional states, therefore,

[107] Clark (forthcoming) expresses considerable sympathy with Dennett's views on the matter, although, as I do below, he prefers not to argue against the objection on such grounds.

he suggests that intentional states and processes are decomposable into simpler systems, with at most a more or less arbitrary dividing line partitioning 'bottom-level' intentional states from the non-intentional states upon which they supervene. Dennett's position is not without its critics, of course, but since I shall prefer the second and third strategies we shall leave such arguments as they stand.[109]

§5.4.2 Second Strategy: Denying that no Intrinsic Content is Involved

As just noted, Adams and Aizawa (2001) propose as a mark of the cognitive 'that cognitive states must involve intrinsic, non-derived content'. But this claim requires considerable clarification – how involved must the intrinsic content be in the process for it to be cognitive? Must *every* aspect of a cognitive process involve intrinsic content? Or is it only that *somewhere* in the cognitive process there must be some state with intrinsic content?

If we read this claim in the first and stronger sense it is surely highly implausible. In a recent response to Adams and Aizawa (2001), Clark (forthcoming) reads the claim in precisely such a way and raises the question, 'must everything that is to count as part of an individual's mental processing be composed solely and exclusively of states of affairs of this latter (intrinsically content-bearing) kind? I see no reason to think that they must.'[110] In a recent restatement of their arguments, Adams and Aizawa (forthcoming a) apparently agree. Despite being quite candid that their original claim 'has some calculated openness

[108] Most obviously and directly in Dennett (1990)

[109] See, for example, Searle (1992), pp. 212-3, and Adams and Aizawa (forthcoming c) for their objections. In addition, well known Sorites-style problems arise when attempting to present arguments involving vague predicates. I leave it to the reader to select their own preferred solution.

[110] Clark (forthcoming), p. 10. Clarks example of cognition involving derived content – imagined Euler circles – is (in this author's view) rightly criticised by Adams and Aizawa, but since Adams and Aizawa reject the strong reading of their claim, this issue is moot and will not concern us.

about it', they nevertheless chide Clark for misunderstanding them and go on to make clear that their claim should not be interpreted in the strong sense:

> The hypothesis [that cognition must involve intrinsic content] has this latitude, since we think that while we have good reasons to believe in the existence of intrinsic content, we have no good reasons to think that cognitive states must consist entirely of intrinsic representations or that cognitive states must be, in their entirety, content bearing. (Adams and Aizawa, forthcoming a.)

But once we interpret their claim in the weaker sense that they apparently intended, it becomes considerably less clear why the extended cognitivist should be at all concerned. At least on the face of it, none of the putative examples of extended cognition mentioned above fail to involve intrinsic content, given that that is what human brain states possess. All of them involve a cognitive process that putatively inheres in a systemic cognitive object consisting of a human brain together with its environment. That systemic cognitive object will contain at least some partial cognitive states (brain states) with non-derived content. With the exception, perhaps, of those who follow Dennett in denying or questioning whether there is intrinsic intentionality in the first place, I can think of no extended cognitivist who would take issue with this weaker reading of the claim.

Yet Adams and Aizawa take this weaker reading of their claim to pose a problem for the extended cognitivist. Why? The following extract may provide a clue:

> if you have a process that involves no intrinsic content, then the condition rules that the process is non-cognitive. In fact, that is exactly what the condition is used to show in our 2001 paper. The images on the CRT of the Tetris video game are not representations of blocks to be rotated, they are the blocks to be rotated. (Adams and Aizawa, forthcoming a.)

Recall, now, that the cognitive behaviour to be explained involves the answering of questions concerning the 'fit' of various two-dimensional shapes that appear on a computer screen into 'sockets' also appearing on the screen. Assuming that the cognitive agent manages to answer better than chance, we presumably will want to explain how they manage to do that. For *that* process there is every reason to suppose that intrinsic content is involved, since the process surely involves certain perceptual, reasoning, and motor-coordinating states of the cognitive agent's brain that are, prima facie, strong candidates for carrying such content. But judging by the quote, Adams and Aizawa are unconcerned with *that* cognitive process or with explaining *that* cognitive behaviour. Instead, they appear to want to apply their principle to a particular sub-cognitive process involving just the blocks on the CRT screen. Even down to this level of functional decomposition, however, their demand may yet be satisfied. The process of rotating the blocks is a process co-ordinated by the agent in interaction with the blocks on the CRT screen. It directly involves the intrinsically intentional states of the cognitive agent both through directing his or her button-pressing actions and through his or her perception of the rotations. Once more, therefore, there would be every reason to suppose that the process involves intrinsic content.

§5.4.3 Third Strategy: Denying the Significance of the Distinction

At *some* level of functional decomposition, of course, we will find sub-cognitive processes supervening on partial cognitive objects whose partial cognitive states are in no respect carrying intrinsic content. If we focus, say, on the processes rotating the CRT blocks after the agent presses the rotate button and before they perceptually process the rotated image on the screen, we will presumably find no intrinsic content. But it is hard to see this as particularly surprising or alarming. After all, even in the case of entirely internal cognitive processes, there is nothing surprising or alarming in the suggestion that at some level of decomposition we will find brute causal processes that are not themselves characterised in cognitive terms.

Adams and Aizawa appear to be making a similar mistake to that noted in §5.3.2, of supposing that the status of specific sub-cognitive processes as genuine parts of a broader cognitive process depends upon every such part having a cognitive status (i.e., by involving intrinsic content) *independently* of the broader process. This is surely an unreasonable demand; as unreasonable as demanding that filling a kettle with water, independently of its role in the process of making a cup of tea, must be a tea-making process.

The extended cognitivist is not suggesting that the blocks on the CRT screen constitutes a *systemic* cognitive state. The claim is that the CRT blocks count as *partial* cognitive objects because they are a constituent *part* of a systemic cognitive object comprising the cognitive agent *together with* the CRT screen. This, in turn, stems ultimately from the status of the rotations of the blocks on the CRT screen as *part* of a cognitive process that explains the cognitive behaviour at issue. Adams and Aizawa admit that they have no reason to think that cognitive processes or states must consist *entirely* of intrinsic representations or that cognitive processes or states must be *entirely* content-bearing. It therefore remains unclear on what basis the intrinsic vs. derived content distinction can be used to oppose the extended cognitivist's claim that the rotation of the CRT blocks are part of a cognitive process.[111]

§5.5 – SCIENTIFIC KINDS AND EXPLANATORY UNITY

Another objection concerns the contribution (or alleged lack of it) that extended cognitivism can make to the scientific understanding of cognition. In Adams and Aizawa

(2001) the argument has two distinct strands. The first strand emphasises the sheer *diversity* of processes and media that would have to be incorporated within a single cognitive science, were extended cognitivism taken seriously. For example, it would include photo albums, Rolodexes, computer databases, strings around the finger, address books, sets of business cards, bulletin boards, date books, personal information managing software, palmtop computers, hand drawn maps, and so on. This imponderably diverse collection leads Adams and Aizawa to conclude that no useful cognitive science will emerge from taking extended cognition seriously. They suggest that:

> in contrast to intracranial processes, transcranial processes are not likely to give rise to interesting scientific regularities. There are no laws covering humans and their tool use over and above the laws of intracranial human cognition and the laws of the physical tools. (2001, p. 61.)

By extension, Adams and Aizawa suggest, internal cognitive processes and states, and (partly) external 'cognitive' processes and states do not constitute an explanatory kind. In what amounts to a mirror-image of the central argument of the extended cognitivist, they argue that whatever superficial similarities there might be between internal cognition and extended 'cognition', there are deeper dissimilarities that make them very different natural kinds with only the former being properly cognitive.

At the heart of Adams and Aizawa's argument lies a deeper claim concerning the kind of science that they insist cognitive science must be if it is to be a science at all. They argue that 'the cognitive must be discriminated on the basis of underlying causal processes' (2001, p. 52.) This condition is, they suggest, a second 'mark of the cognitive'.

[111] Objections falling somewhere between the strong and weak readings considered here might be offered. For example, it might be suggested that every part of a cognitive process that is also itself a cognitive process or state must involve intrinsic content, if it has any content. Such objections fail to provide any grounds for concern to the extended cognitivist just so long as the principle appealed to (as here) does not rule out there being sub-cognitive processes that do not involve intrinsic content being a proper part of a cognitive process.

§5.5.1 The Diversity of Extended Cognition

Let us begin by noting that the sheer diversity of the phenomena does nothing *by itself* to suggest that an underlying unity is unlikely to be found. What matters more is the degree of diversity at the level of description appropriate for that discipline. Consider, for comparison, both chemistry and biology. Both of these sciences, to an extent perhaps greater than any that might be suggested in respect of extended cognition, deal with an astonishing diversity in the forms and behaviours of their subject matters. Nevertheless, those forms and behaviours turn out to be well captured (to varying degrees) by certain general underlying processes, and amidst the dazzling diversity can be found various scientific kinds. In the face of such diversity, pessimism would have been quite forgivable in the early stages of either of these sciences, but it turns out that such pessimism would have been unfounded.

Adams and Aizawa, however, take their position to be grounded in more than mere pessimism, but instead in the empirical evidence. They argue (forthcoming a) that the empirical evidence supports the following claim: that there are processes that

a) are plausibly construed to be cognitive

b) occur within the brain

c) do not occur outside the brain

d) do not cross the bounds of the brain

This claim is ambiguous, firstly between *instances* and *kinds*. Interpreted as ranging over instances of the cognitive the claim is entirely unobjectionable. Taken at face value it is quite true and unobjectionable to say that there are *instances* of processes that are plausibly construed as cognitive, occur within the brain, do not occur outside the brain, and do not cross the bounds of the brain. An in-the-head arithmetical calculation of 154 + 17 would be an obvious example. I take it, therefore, that it is intended as ranging over *kinds* of processes. This also requires further disambiguation between *particular* kinds of cognitive processes, and *all* kinds of cognitive processes. Of the first of these, the extended cognitivist can once again concede the claim as unobjectionable. There quite possibly are particular kinds of processes that are plausibly cognitive, occur within the brain, do not occur outside the brain, and do not cross the bounds of the brain. Putting aside whether they should count as cognitive, perhaps conscious processes and states are a case in point. The cognitive externalist can still argue that *some* kinds of cognitive processes do, on occasion, cross the bounds of the brain – arithmetical calculation, for example. That, after all, is all they are claiming. It will then become an interesting question as to *why* some kinds of cognitive processes can cross the boundaries of the brain and others can't, but that is not a counter to the extended cognitivists' claim unless the explanation for that difference is later taken to distinguish the cognitive from the non-cognitive. Clearly, then, only the strongest interpretation of the claim is of concern to the extended cognitivist: the claim that *all* kinds of cognition that are plausibly construed as cognitive, occur within the brain, do not occur outside the brain, and do not cross the bounds of the brain. Since that is the only interpretation that poses a genuine challenge to the extended cognitivist, it behoves us to ask what empirical grounds Adams and Aizawa can put on offer for the claim that *all* cognitive processes occur only in the brain.

Adams and Aizawa (forthcoming a) place great emphasis on the claim that 'the brain processes information according to different principles than do brain-tool combinations'. In support of this they reel off a list of consumer electronic devices including CD players, MP3 players, FM radios, AM radios, digital cameras, inkjet printers, cell phones, watches, personal computers, and walkie-talkies. All these devices, like brains, are information processors, they admit.[112] They then go on to list a number of things that (they suggest) humans brains can do that these objects cannot *because*, they say, the brain processes information differently from these devices. Amongst those capacities they list linguistic processing and facial recognition over a range of environmental conditions. They go on:

> This is why the brain is crucial for humans to drive cars, where these other devices are not. The differences in information processing capacities between the brain and a DVD or CD player is part of the story why you can't play a DVD or CD with just a human brain. These differences are part of the reason you need a radio to listen to FM or AM broadcasts. It is these differences that support the defeasible view that there is a kind of intracranial processing, plausibly construed as cognitive, that differs from any extracranial or transcranial processing. (Forthcoming a.)

But it is surely somewhat churlish to try to base an argument against extended cognitivism on the manifest functional differences between human brains and DVD players, of all things. Those difference are explained, in the first place, not by the fact that they process information differently (which is true insofar as it goes), but from the fact that they process information differently *because* the latter is not designed to drive cars or perform face recognition and the former is not designed to receive radio signals.[113] Such functional differences no more demonstrate a deep explanatory disunity between DVD players and brains than do manifest functional differences demonstrate a deep explanatory disunity between hearts and brains. There is virtually nothing that a heart can do that a brain can do, or vice versa. Most pertinently, hearts cannot cogitate and brains cannot pump blood. But as we can all agree, according to standard biological science the heart and the brain *do* share a deep explanatory unity in several respects, including such

[112] Let us note, in passing, that in raising this point they have implicitly, if inadvertently, brought a very large measure of theoretical unity to a large portion of the domain that they complain of as lacking such a unity.

[113] Of course, brains are not designed to drive cars either, but *cars* are designed to take advantage of the various capacities that our brains together with our bodies (by design) provide for us.

commonalities as both being the products of millions of years of selective evolutionary pressure, and both being products of cellular activity. Once we move to consider technological devices that are *designed* to perform (or at least capable of performing) some of the functions otherwise performed by the brain – from the symbolic recording capabilities of pen and paper to the sophisticated information processing of digital computers – then the differences that exist in the *way* that they perform such tasks typically take on an appropriate air of irrelevancy. A digital computer processes information via electrical impulses travelling through a silicon chip – a human brain processing information via electrical impulses travelling along synaptic fibres. These are differences that make a difference in some functional respects, but not in others.

§5.5.2 Causal Kinds vs. Functional Kinds

As mentioned earlier, the deeper claim that drives much of Adams and Aizawa's thinking on this issue is a commitment to cognition being a subject to *causal* explanation ranging over *causal* natural kinds. Cognition, if it is to be explained at all, is to be explained by appeal to causally individuated natural kinds and processes. Consequently, the kinds of external devices that provide the mainstay of the extended cognitivist's arguments are, they seem to suggest, automatically barred from providing a basis for the scientific explanation of cognition. As they see it, '[t]ools do not constitute a natural kind; tools are, after all, artefacts' (forthcoming a).

There are a number of ways to respond to this claim. In the first place, let us note that an insistence on cognitive explanation being fundamentally causal in character stands rather incongruously with their professed acceptance of the manifestly *functional* parity principle. As that principle expresses it, the cognitive status of an external process stems from direct *functional* analogy with a process that, were they to take place in the head, we would count as cognitive.

Secondly, it is questionable whether there is any *fundamental* explanatory difference between our use of tools and our use of other more 'natural' artefacts such as eyes or hearts. Eyes are as much tools for seeing as binoculars; their differences stem largely from the fact that the research and design process for the former extends over a considerably longer period and operates without the benefit of direct human cognitive intervention. Moreover, as Beth Preston has noted, at least prima facie strong arguments can be made that tool use,

> constitutes [with language] the other major form of cognitive mediation between individual and world. For example, a spoon embodies in its very shape aspects of our knowledge of the physical propensities of liquids, and therefore is the peculiarly appropriate mediator of the interaction between individual and world in situations where this knowledge comes into play.[114]

But putting aside whether tool use can (or should) be seen as a 'natural' or 'legitimate' explanatory kind, there are strong grounds to suggest that a science of cognition must deal in *functional* kinds rather than causal kinds if it is to be of any explanatory value.

To see this note firstly something that Adams and Aizawa are sensibly willing to concede: that we cannot explain significant human capabilities without appeal to information processing technology. Our capacity to engage in complex mathematical calculations is, they freely admit, heavily dependent upon such technological resources as calculators or pencils and paper. Putting aside for a moment whether we deign to call the disciplinary study of such capabilities a 'science', and further putting aside whether the role of such technology is 'cognitive', the obvious fact remains that appeal to such technology *does* serve an important explanatory role for human behavioural capacities. So at the very least, the appeal to information technology does help to explain certain human capacities including, for example, our ability to perform complex arithmetical calculations with very large numbers. To try to explain that capacity would be like trying to explain our capacity to travel faster than sound without appeal to aeronautical technology. Combining this with the relatively uncontentious claim that information technology is constituted by various functional kinds, we are immediately in a position to make cogent explanatory appeal to functional kinds for a certain portion of human behaviour. But once this much is conceded as a legitimate explanatory strategy, it becomes considerably less clear why anyone should demand that the explanation of cognition must proceed solely by appeal to causal processes and causally individuated kinds.

Adams and Aizawa are quite correct, of course, to point out that there is a remarkable diversity amongst information technology when viewed in terms of the details of their physical processes. They do not constitute especially unified causal kinds, and their interesting functional properties do not generally track processes that can be type-identified causally. But, contrary to what Adams and Aizawa might like to suggest, it is far from anathema to orthodox cognitive science to demur from giving explanations in terms of causal laws or to type-identify its kinds causally. For example, one of *the* fundamental explanatory concepts in orthodox cognitive science is, as we have noted, information, in the truth-dependent sense of the term. Yet such information is not, as Dretske is at pains to emphasise, an essentially causal concept:

> The flow of information may, and in most familiar instances obviously does, depend on underlying causal processes. Nevertheless the informational relationships

[114] Preston (1998b), p. 514. See also, for example, Preston (1998a) for the explanatory similarities between appeals to biological and artefactual kinds.

> [between source and receiver] must be distinguished from the system of causal relationships existing between these points. (Dretske, 1981, p. 26.)

Certainly, causal relations help to underwrite explanatorily significant informational relations, but information flow supervenes on an unlimited range of causal relations across an unlimited range of causal kinds.

Moreover, manifest diversity at the level of causal processing in no way precludes there being explanatorily significant processes and kinds at some higher and more abstract – indeed, *functional* – level of description. As Dennett (1998, chapter 5) has noted, there may well be what he calls 'real patterns' that perform a vital role in our explanations of such phenomena. Similarly, Putnam (1994, chapter 23) argues that the fact that certain phenomena may be deducible from lower-level causal processes does not show that the appropriate kind of explanation for such phenomena will be at that causal level rather than, say, at some higher, functional level. Indeed, as Clark (2005, p. 18) points out, internal or external cognitive processes involving information-bearing states need not be similar in terms of their detailed implementation:

> It is simply that, in respect of the role that the long-term encodings play in guiding current response, both modes of storage [i.e., internal and external] can be seen as supporting dispositional beliefs. It is the way the information is poised to guide reasoning . . . and behaviour that counts.

Placing the explanatory emphasis on causal laws and causally individuated kinds, as Adams and Aizawa might wish, fails to appreciate that the domain of cognitive behaviour is, at least in part, a *functional* domain. Cognition is a domain fashioned by both nature and artifice in order to solve problems that typically demand flexible or plastic behaviour. As already noted, it is a domain more akin to, say, power-generation than to combustion, and as such, its scientific study is perhaps more closely comparable to the mixed science of engineering than to a purely causal science such as physics or chemistry.[115]

§5.6 – LOCUS OF CONTROL

The final objection that we will consider in this chapter (further objections, more directly concerning the role of external representations, are considered in the next chapter) is likewise intended to undercut the putative analogy between internal and external cases, and concerns the matter of *control.* Considering some of Clark and Chalmers' examples, Butler (1998, p. 212) suggests,

> there can be no question that the locus of computational or cognitive control resides inside the head of the subject. . . . [T]he decision to engage in the task, the drive to complete it, and the way it is to be carried out all involve internal processes in a way quite distinct from the way external processes are involved. If this feature is indeed the mark of a truly cognitive system, then it is a mark by means of which the external processes [Clark and Chalmers] point to can be excluded.

Let us allow, for the moment at least, that the substantive premise here is correct – that the locus of control for cognitive behaviour lies exclusively within the brain of the cognitive agent. The central issue, then, becomes whether the concluding conditional of the above quote has any truth to it. Is such control a distinctive mark of the cognitive? It is hard to see why anyone should think that it is. Whilst control is undoubtedly an important feature of *agency* (and hence, *cognitive* agency) there is no reason to suppose that a process or state must be within an individual's control in order to count as one of their cognitive processes or states.

[115] Should the critic of extended cognitivism be looking for a burgeoning science examining the joint problem-solving properties of humans and computer, it typically travels under the banner of HCI: human-computer interaction. Several journals are dealing with these (amongst related) issues. For example, 'Human Computer Interaction', 'International Journal of Human-Computer Studies', and 'Computers in Human behaviour'. Many connections with recent research can be found in Clark (2003). See also Gorayska and Mey (1996).

In the case of cognitive *states* this is relatively clear. In the case of most perceptual states, for example, control is, at most, extremely limited. Whilst we typically have substantial control over various actions that influence our perceptual states, such as where we look, what we focus our attention on, and unconscious control over such things as eye saccades, we do not generally have any control over what we see or hear when we do so. The content of our perceptual states are, to that extent, generally beyond our control. Yet no one, I take it, would claim that such states are not cognitive on such grounds. What about *processes* of control? As Clark (forthcoming) suggests, if adopted as a general discriminatory mark of the cognitive it risks being reduced to absurdity should we apply such a principle *within* the brain of the cognitive agent. Clark asks (rhetorically, it must be admitted):

> Do we now count as *not part of my mind or myself* any neural subsystems that are not the ultimate arbiters of action and choice? Suppose only my frontal lobes have the final say – does that shrink the real mind to just the frontal lobes!? What if . . . no subsystem has the 'final say'. Has the mind and self just disappeared? (Clark, 2005, p. 24)

Further questions can be raised not merely against its adequacy as a 'mark of the cognitive' but against the explanatory utility of such a control principle. Consider cases of complex reciprocal interaction between a cognitive agent and their environment. In such cases attributing the locus of control within this or that cognitive agent appears to serve little or no explanatory role. Consider a jazz band improvisation.[116] Perhaps a strong claim can be made that each musician remains, in some important sense, the locus of control for the individual notes that they play. Yet the degree of reciprocal interaction between the musicians in the ongoing improvisation is sufficiently complex and dynamic that attributions of control to this or that musician will serve a limited role in the explanation

[116] Examples along these lines can be found in Haugeland (1998), chapter 9 as well as Dreyfuss (1972).

of the resulting performance. Insofar as cognition is involved in such behaviour, locating the control within the traditional organismic boundaries of particular cognitive agents seems to be a non-issue.

Lastly, let us note that from a *functional* perspective it is not at all essential that a cognitive agent be *directly* involved in the enactment of each and every step of the process for it to count as part of his or her cognitive processing. In some cases, of course, the cognitive agent *will* be continuously involved in the ongoing process. In-the-head arithmetical calculation is clearly such a case, but so is, to take a putatively extended example, moving tiles around on a Scrabble rack. In both cases the cognitive agent is cognitively engaged throughout the process and at each step. Regardless of whether one feels comfortable ascribing cognitive status to that process of shuffling the tiles there should be little difficulty in allowing that it is *all* something that the cognitive agent is doing. It is the cognitive agent who, say, picks up a randomly selected tile, moves it to a different randomly selected location, and so on. But now note that in the computerised version of the game the cognitive agent can simply click on a button and the tiles are shuffled *automatically by the computer*. From a functional perspective this fact makes little or no difference. It does not matter whether the cognitive agent is *directly* involved in the shuffling of the tiles, or whether the tiles are shuffled by a playing partner, or a computer. Generalising a little, given a functionalist approach to cognition, what seems to be significant in a cognitive process is the functional role that a process plays in the cognitive economy of the agent. So long as one stands in the appropriate agential relations to the process, the operations of pocket calculators, computers, and so on are, I suggest, as much part of one's cognitive processing as one's neuronal activities.

This claim, I suspect, may require some considerable digestion. To perhaps ease matters a little, compare this with agency more generally. Attributions of agency are *not* restricted simply to the processes over which the agent has *ongoing* control. For example, a claim by a gunman that it was the *bullet* that killed the victim not *him* will get short shrift. It will not do for the gunman to protest that all he did was pull the trigger; that the rest was nothing to do with him at all. Attributions of agency might even coherently be extended beyond the death of the agent. Someone who plants a landmine but who dies before it explodes is as much the agent responsible for the explosion as if they stood at the side of the road and detonated the mine directly. Whilst I shall not presume to give a general account of agency, what seems to be important here is that the agent initiates a process with the desire to bring about a certain outcome, and that that outcome is the result of (and can reasonably be expected to be the result of) such a process. Similarly, I suggest, with *cognitive* agency. What matters is not that the agent has direct and ongoing control over each and every step or over each event in a cognitive process, but that the agent *initiates* the process with the desire to bring about a certain solution, and that the outcome is the result of (and can reasonably be expected to be the result of) such a process.

§5.7 – CONCLUSION

In this chapter I have laid and clarified the central extended cognitivist thesis. This has involved both distinguishing it from certain positions that are more or loss closely associated with it (and sometimes confused with it), as well as defining a number of aspects of cognitive processes, states, and systems that have lain implicit in the dissertation up to this point. After examining some key putative examples of extended cognition (as well as noting some important points of agreement between critics and protagonists), I then examined and rejected four of the central objections ranged against extended cognitivism.

The first of these objections – the 'coupling-constitution fallacy' – was found to be based on a misunderstanding of the central motivation for the extended cognitivist position. Notwithstanding that, further faults were found primarily in a failure to focus on the relevant behaviour and an unprincipled demand that the sub-cognitive processes of a cognitive process be independently cognitive. The next objection focussed on the alleged significance of the distinction between intrinsic and derived intentionality and three strategies were proposed for dealing with the objection. Only states or process with intrinsic content, it is alleged, can count as cognitive. The first response – Dennett's denial of the reality of the distinction – was outlined but passed over in favour of less radical alternatives. It was argued that once we focus on the appropriate level of functional decomposition, the challenge is defused either by a satisfaction of the demand for intrinsic content or, once again, by the objection making an unprincipled demand that the sub-cognitive processes of a cognitive process be independently cognitive. The third objection questioned the scientific credentials of extended cognitivism by suggesting that it is incapable of delivering causally individuated explanations and that only causally individuated explanations are appropriate. This objection was criticised on two main grounds. The first ground was that the empirical evidence does not obviously support a claim of the strength required to worry the extended cognitivist. The second ground was that cognitive explanation does not need to be fundamentally causal. It was argued that explanatory unity across causally diverse phenomena may be found at a functional level. The final objection considered in this chapter emphasised the importance of the locus of control for a cognitive process as something residing within the head of the cognitive agent. This was rejected as being neither a necessary aspect of cognition, nor, at least in any direct or ongoing way, an essential feature of agency more generally.

In the next chapter we shall continue the defence of extended cognition primarily through an examination of its application in the realm of belief-desire explanation. There we shall encounter a slightly stronger form of the extended cognitivist thesis – one that attempts to cement extended cognitivism's credentials within the mainstream of cognitive science through the explicit use of representational explanation. During the course of that defence a number of remaining criticisms of extended cognitivism will be examined and rejected.

CHAPTER 6 – EXTENDED INTENTIONAL STATES

In the previous chapter I developed and detailed the general thesis of extended cognitivism: that an individual's cognitive processes or states sometimes extend beyond the boundaries of their brain or skin. In this chapter I examine and defend a stronger version of extended cognitivism, one that includes a commitment to specifically *representational* extended cognitive processes and states.

In the first section (§6.1), and as something of a preamble, I discuss the pervasive role of *cognitive technology* in enhancing our cognitive powers. This is followed (§6.2) by defence of a specifically representationalist interpretation of the examples considered in the preceding chapter: pen and paper arithmetical calculation, Clark and Chalmers' Tetris example, and Scrabble. In the next section (§6.3) I characterise extended cognitivism in respect of external belief states and intentional explanation – the central focus of the chapter.[117] The key example, drawn from Clark and Chalmers (1998), is that of Otto, a sufferer of a mild form of Alzheimer's disease, and his notebook. Following Clark and Chalmers, I argue that the role that Otto's notebook plays in his cognitive economy qualifies it as a repository of his beliefs. I then consider and ultimately reject three significant objections raised against this claim. The first (§6.4) – the *cognitive bloat* objection – occupies the bulk of the chapter. Its central thrust (which is conceded) is that extended cognitivism implies the extension of belief states to counter-intuitive cases. Problems are identified (§6.4.1) with Clark and Chalmers original handling of the objection and so rather than resisting the extension to counter-intuitive cases I argue in favour of embracing that extension. I argue for this in three stages. Firstly, (§6.4.2) I re-affirm the often-made point that naturalistically inclined explanations do not kow-tow to ordinary usage. On the contrary, ordinary usage, if it is to be taken on board, must be grounded in theoretically well-motivated intuitions. To the extent that fears about cognitive bloat may be fuelled by the undeniable resistance of ordinary usage, and to the extent that no such theoretically adequate grounding intuitions are provided for that usage, they can and ought to be set aside. Secondly, (§6.4.3) I give a brief argument in support of the claim that believing is a kind of functional relation (more specifically, a causal-functional relation) between agents and content-bearing states, and that as such the location of such states in relation to the agent's body is of only secondary significance – secondary, that is, to the satisfaction of that causal-functional relation. The second objection – the *Otto 2-step* – (§6.5) argues that the use of external representational resources is a 2-step process, with only the internal being properly intentional. I reject this primarily through a distinction between the explanatory roles of occurrent vs. non-occurrent (i.e., dispositional) beliefs. The third objection (§6.6) argues that the role of *perception* provides a principled 'barrier' between cognitive processes and states occurring within the head and other (putatively non-cognitive) processes or states occurring outside of the head. Finally, (§6.7) I flesh out my account by reference to a set of putatively necessary and sufficient conditions for non-occurrent belief (occurrent belief being left largely unanalysed) that are clearly satisfied by certain external states. These conditions are, in some measure, modifications of conditions set out in Clark and

[117] There is a familiar ambiguity when talking about beliefs: i.e., between the content of the belief and the content-bearer. The issue of extended cognitivism and its challenge via cognitive bloat concerns the location of such content-bearers and as such, when referring to some cognitive agent's beliefs, I should be understood as referring to the content-bearer.

Chalmers (1998) – modified, in part, so as to avoid potential counter-examples and in part with an eye to greater clarity and precision.

§6.1 – COGNITIVE TECHNOLOGY

Many of the putative candidates of extended cognition involve, as I shall understand the phrase, the use of *information technology* – that is, technology that directly involves or facilitates the storage, retrieval, or manipulation of representations. In its more modern usage 'information technology' tends to be used to refer to specifically computerised technology, but the digital computer is but the latest in a long line of advances in information technology dating back, perhaps, to the external symbolism of cave paintings. The significance of information technology to our cognitive powers cannot be overstated. As Donald (1991) has argued, symbolic representation is the 'principle cognitive signature' of human beings and our use of external representation is the most significant manifestation of symbolic representation in specifically human cognitive evolution. With the invention of external symbolic storage (especially writing) human beings became capable of circumventing, at least partially, the limitations of biological working memory, whilst creating a wide range of new storage, retrieval and processing possibilities (see especially Donald, 1991, p. 308ff). The status of pocket calculators and computers as examples of information technology is clear, and under the suitably broad construal given above so is the status of pen and paper as well as writing more generally.

§6.2 – IN DEFENCE OF EXTERNAL REPRESENTATIONS

In many cases, for example written numerals in pen and paper arithmetical calculations, there are prima facie strong grounds to consider the external states as representational, satisfying all of the central characteristics of representations identified in §3.1. For other cases matters are considerably less evident. Why, for example, should we count the manipulation of tiles on a Scrabble rack or the manipulation of shapes on a computer screen as process involving the manipulation of *representations*? Adams and Aizawa raise essentially this concern. In relation to the Tetris case (§5.2.2) they suggest that

> [i]n case (1) [the in the head case], the agent presumably uses mental representations of the blocks and their on-screen rotations in cognitive processing. By contrast, in case (2) [the distributed case], the blocks on the screen that are physically rotated by pushing the button are not representations at all . . . . They do not *represent* blocks to be fit together; they *are* the blocks to be fit together.

Similarly, in Scrabble when relying on purely internal resources[118] we standardly appeal to internal cognitive representations as stand-ins for the physical tiles, but in the case of putative extended cognition involving the manipulation of the tiles it would seem, at least at first blush, that such an imputation is idle, for the cognition operates directly with the objects that, in the internal cognitive activity, required representational stand-ins. The activity, therefore, might seem not to require representations, at least so far as the physical manipulations of the tiles are concerned. The point is that if we appeal to representations in the internal case but not the external case then the supposed analogy breaks down and the parity principle (§5.1.3) fails to apply.

But the differences between pen and paper arithmetical calculations and such examples as moving tiles on a Scrabble rack or rotating Tetris blocks are considerably less than might, at first, appear. The representational status of each stems from a key feature common to all the above cases: they are all *rule-governed* activities. *All* concrete rule-

[118] Actually, even in the supposedly cognitively internal case one is typically not relying on purely internal resources inasmuch as one is continually studying the rack as one cogitates, much as one might a chess board in a game of chess. But I pass over this.

governed activities, such as moving physical tiles around on a scrabble rack or moving a material chess piece from one square to another on a wooden board, *can* be seen as representations of operations within the abstract problem-space determined by those rules. In the case of Scrabble (and, mutatis mutandis, Tetris, chess, etc.) the problem space is fixed by the rules of the game with the concrete tiles representing values in the abstract problem-space. The moving of physical tiles around on a tile-rack represents an exploration of the problem space in search of a valid solution. Similarly, in the case of pen and paper arithmetical calculations the problem-space is fixed by the rules of arithmetic. The digits represent numbers and manipulations of those digits represent arithmetic operations over numbers; if the calculations are correct, the operations will be 'legal moves'. This fact renders it virtually obligatory for there to be an explanatory appeal to an *abstract model* of the problem-domain in which the physical tokens are representations of positions within the problem domain and the concrete operations are representations of transitions between such positions.

To explore the example of Scrabble a little further, we might describe each possible unique tile combination as a location within the Scrabble problem-space. Some of these locations will be, relative to the conditions of the board (again abstractly characterised), *solutions* in the problem-space, and one will be judged the best solution. Such an abstract multi-dimensional description of Scrabble may be quite complex depending on how many of the subtleties of the game are included, but the basic outlines will be something like the following. The 'board' is a 15 by 15 array that will take up two dimensions. A third dimension will be taken up by the values for each position in the array. For the start position the values in the array will initially all be set to zero – in other words, we start with an empty board. Assuming that the alphabet is Roman, each location on the board may, as the game progresses and according to the rules, take one of 151 values (any of the 98 regular tiles, 2 blanks that may take any of 26 values each, plus an empty space). A further predicate specifies the set of tiles for each player's rack. The specification of a language (English, say) further determines a set of valid words that, together with formation rules (such as that each new word, after the first word, must utilise at least one letter already on the board), determines a binary relation (i.e. legal moves) from any board position and rack. The initial play then selects one of these locations in the problem-space, given the rack position, which, according to the specified language and formation rules, simultaneously eliminates certain erstwhile solutions and opens up others thereby changing the range of potential paths through the problem space. The game then proceeds by a turn-by-turn selection of valid solutions through the problem-space. A sequence of such solutions constitutes a particular game or a 'path' through the problem space.[119]

There is no question that we are not particularly accustomed to viewing the game of Scrabble in such abstract terms, and for precisely that reason we are not accustomed to giving a physical instantiation of a game or the extended cognitive activities typically involved in it, a representational gloss.[120] Neither is it to suggest that in all cases where we can view a concrete process as a representation of an abstract domain that we *should* view it as such. The initial point of note here is that the game of Scrabble and the extended cognitive processes that it typically involves *can* be given a representational gloss simply by shifting our focus to a different descriptive level – to a more abstract domain.

[119] This is not intended as a complete rendition of the abstract game of Scrabble, but simply to illustrate the manner in which concrete Scrabble games can be treated as representations of an abstract domain.
[120] This is less the case in certain kinds of games where certain sequences of moves are identified in representational terms. For example, it is relatively natural to think of a chess board as a representation where one is, say, playing through the tenth game of Fischer-Spassky 1972 World Championships.

Why, though, are we sometimes more and sometimes less inclined to see physical activities or behaviour as representations of some other, more abstract structure? As suggested earlier, it stems largely from our explanatory needs. In the case of beliefs, the motivation for ascribing similar representational contents to the beliefs of two individuals lies in the explanatory leverage this gives us concerning their similar behaviours. As Pylyshyn points out, we can explain the similar behaviours of two individuals with greater coherence and simplicity by ascribing to them a belief with the same representational content, than we can by treating their behaviours idiosyncratically. The appeal to such structures allows us to step above the details of the implementation of these beliefs in the specific neuronal make-up of individual brains to greater explanatory effect. Similarly, with chess (or Scrabble, or Tetris, etc.) an appeal to an abstract model of the problem domain helps to organise and explain superficially idiosyncratic manoeuvres, without losing ourselves in the obscuring details. Whether someone examines a possible mating attack in their head, over a board using carved wooden pieces, or on a computer screen on which pieces are represented by letters, viewing these physically disparate processes as representations of an abstract domain allows us to provide an explanatory unity to such behaviours. The characterisation of such processes and states as representations of operations or states within the abstract problem-domain plays essentially the same unifying explanatory role in our cognitive explanations as does the appeal to representational content in the case of belief-desire intentional explanation. Viewed in such a manner, the movements of physical tiles on a Scrabble rack in search of a good move, just as much in-the-head operations over internal neuronal representations, will play essentially the same representation-manipulating explanatory role.

Further justification for treating a concrete operation as a representation of an abstract domain can be found wherever we take ourselves to learn something about an abstract domain through the manipulation of the concrete structures. In the case of arithmetic this is relatively clear. We typically take ourselves discover facts about the arithmetical domain (for example, what the sum of several numbers is) by performing concrete operations over concrete objects (for example, written digits). Similarly with chess. Performing concrete operations using physical tokens allows us to discover, for example, whether it is possible in the game of chess to arrange eight queens on a board such that none can capture any of the others, or whether a knight can visit every one of the sixty-four squares in succession landing only once on each square (the so called 'knight's tour'). Our tendency to endorse a representational reading of such concrete operations at least partly reflects our interest in the abstract domain in question. It is, I take it, partly for such reasons that for both arithmetic and chess we have considerably more well-developed theories than we do for Scrabble.

§6.3 – OTTO'S NOTEBOOK

With the general representational credentials of extended cognition (hopefully) in place, let us turn our attention to the imaginary case of Otto and his notebook. Otto is an individual whose neuronal belief-storage is not functioning properly (due to mild Alzheimer's disease) and who, therefore, as a replacement for his faltering neuronal-based beliefs, keeps a notebook containing various pieces of information.[121] Whenever Otto encounters information that he suspects might prove useful, he writes it down in his notebook. One such piece of information recorded in his notebook is that New York's Museum of Modern Art (MoMA) is located on 53rd Street. Clark and Chalmers argue that when Otto hears of an interesting new exhibit at that museum, decides to pay it a visit, and then looks in his notebook to determine that it lies on 53rd Street that this is, in all *cognitively* relevant respects, similar to Inga, an entirely normal cognitive agent who merely consults her neuronal-based beliefs in the usual way. In such a case it is wholly appropriate, they suggest, to attribute to Otto the same belief as we do to Inga, but that it just happens that Otto's belief (more specifically, the content-bearer of his belief) lies beyond his skin. It is argued that all cognitively relevant features present in Inga's case are paralleled in Otto's, including such things as portability of the content, ready accessibility, reliability, and automatic endorsement. Otto's circumstance, moreover, is not that unusual. A police officer giving evidence in court will be credited as presenting reliable testimony even though they will standardly refer to their notebook rather than (or perhaps in addition to) internal neuronal resources.[122] We regularly rely on written lists, address books, and so on, as either backups to our internal informational resources, or as primary informational resources.

[121] In Clark and Chalmers (1998) they occasionally talked about Otto's notebook as constituting his 'memory', but in Clark (forthcoming) this is dropped and the focus is exclusively on its role as belief storage. Perhaps one good reason for this shift is the often supposed facticity of memory. Otto or Inga can only remember what was actually the case, whereas belief in non-factive. More significantly, belief and memory can come apart. Otto or Inga can remember without believing what they remember, and believe something without remembering it. Focussing on the notebook as belief-storage, therefore, avoids problems that would otherwise occur regarding the automatic endorsement criterion (§6.4). Aside from direct quotes from Clark and Chalmers (1998) my presentation and discussion reflects the Clark (forthcoming) presentation.

[122] An example drawn from Houghton (1997), p. 168.

§6.4 – THE PROBLEM OF COGNITIVE BLOAT

Our first objection has been raised by both Adams and Aizawa (2001), and very recently by Rupert (2004). The objection claims that counting Otto's notebook as a repository of some of his beliefs leads to an unacceptable mental or cognitive 'bloat'. For example, Adams and Aizawa (2001, p. 57) present the challenge in terms of a dilemma:

> If Clark and Chalmers opt for the simplistic view that anything that is causally connected to a cognitive process is part of the cognitive process, then there is the threat of cognition bleeding into everything . . . . The threat is of pancognitivism, where everything is cognitive. This is surely false. If, on the other hand, Clark and Chalmers opt for some more discriminating mark of the cognitive . . . then it is far from clear that this will allow the cognitive to cross the boundaries of the brain without extending to the whole of creation.

No one, of course, is advocating the simplistic view on offer, least of all Clark and Chalmers. As was noted in §5.3, advocates of extended cognitivism do not hold that 'anything that is causally connected to a cognitive process is part of the cognitive process'. For one thing, states or processes that do not carry content relevant to the cognitive task at hand can presumably be immediately excluded. For example, the presence of oxygen in the blood together with its transportation to the brain is causally connected to standard examples of human cognition, but no one considers such physiological processes to be cognitive. One clear reason for this is that haemoglobin does not generally carry representational content in any way relevant to the cognitive task in question. Thus, the threat of pancognitivism wherein 'everything becomes cognitive', a view which is, just as Adams and Aizawa suggest, obviously false, is to be immediately discarded.

It is, therefore, on the second horn of the dilemma that the debate has been focused, with the extended cognitivist attempting to provide some 'discriminating mark of the cognitive' that allows for the inclusion of such things as the contents of Otto's notebook, but excludes the more deeply counter-intuitive or controversial examples. Precisely which examples are counter-intuitive varies to some extent, but in general proponents and critics alike have tended to agree that such things as the contents of Otto's encyclopaedia in his basement and the contents of his local telephone directory in his desk drawer are at best highly problematic and should probably be excluded from counting amongst the content-bearing constituents of Otto's beliefs.[123] Thus, turning specifically to the case of Otto and his notebook, the criticism from cognitive bloat, which is, in essence, an attempted *reductio ad absurdum*, can be summarised in the following valid argument:

> 1. If extended cognitivism is correct, then the contents of Otto's notebook are amongst Otto's beliefs.
>
> 2. If the contents of Otto's notebook are amongst Otto's beliefs, then so are the entire contents of the encyclopaedia in his basement, the telephone directory in his desk drawer, and so on.
>
> 3. It is false – indeed, absurd to suggest – that amongst Otto's beliefs are the entire contents of encyclopaedias or telephone directories.
>
> 4. Therefore, the contents of Otto's notebook are not amongst Otto's beliefs.
>
> 5. Therefore, extended cognitivism is false.

Recognising the threat, Clark and Chalmers (p. 17) put on offer the following four criteria that appear to be satisfied by Otto's notebook, and yet block cognitive bloat:

> 1. The information store is a constant in the agent's cognitive life. For example, whenever the information in Otto's notebook would be relevant, he will rarely take action without consulting it.
>
> 2. The information will be directly available without difficulty.
>
> 3. Upon retrieval the information will be automatically endorsed.
>
> 4. The information will have been consciously endorsed by the agent at some point in the past, and is there because of that endorsement.

[123] Clark (1997a), p. 217, for example, explicitly rejects the suggestion that encyclopaedias in basements count amongst the repositories of beliefs.

Of these criteria only the fourth – the 'past-endorsement' criterion – provides a substantial barrier to cognitive bloat. The encyclopaedia, for example, could well be a constant in an agent's life, directly available without difficulty, and endorsed automatically. Online internet databases coupled with mobile-wireless access are rapidly rendering the satisfaction of the first three criteria a realistic possibility or for some, perhaps, an actuality. Commenting on these criteria Rupert presses the importance of the past endorsement criterion especially strongly by considering the case of a telephone directory service. In such a case, he says:

> the first three criteria imply that virtually every adult, Otto included, with access to a telephone and directory service has true beliefs about the phone numbers of everyone whose number is listed. The directory assistance operator is a constant in Otto's life, easily reached; when the information would be relevant, it guides Otto's behaviour; and Otto automatically endorses whatever the operator tells him, about phone numbers, anyway. It is absurd to say that Otto has beliefs about all of the phone numbers available to him through directory assistance (that is beliefs of the form, 'John Doe's phone number is ###-####'), so long as he remembers how to dial up the operator. To say so would be to depart radically from the ordinary use of 'belief' (similar remarks apply to 'know': given ordinary usage, we would not say that Otto knows my phone-number to be such-and-such). (Rupert, 2004, pp. 402-3. Footnote references omitted.)

Hence, Rupert concludes, inclusion of the past-endorsement criterion is well advised, for it saves the extended cognitivist the embarrassment of having to say that, for any arbitrarily chosen person in the directory, Otto has an accurate belief that their phone number is such-and-such. Moreover, the idea behind the fourth criterion may seem tempting enough. Otto is the author of his notebook, whereas he is, presumably, not the author of his telephone directory or his encyclopaedia. It is, after all, *Otto's* notebook, in as much as Otto himself was consciously involved in the laying down of the content-bearing states that he later consults, and it is this that distinguishes its contents from the contents of the encyclopaedia or the telephone directory as amongst *Otto's* beliefs.

§6.4.1 Some Problems with the Past-Endorsement Criterion

Alas, for the extended cognitivist, notwithstanding its apparent adequacy in stopping cognitive bloat, there are some deep problems with the past-endorsement criterion. To be fair to Clark and Chalmers, in their paper they are themselves very cautious in their endorsement of it and express some reservations, particularly concerning the apparent incorporation of historical elements into attributions of belief.[124] Moreover, as Rupert points out (p. 402), it undermines the spirit (if not the letter) of extended cognition to appeal to Otto's prior *conscious* endorsement given that consciousness is not supposed to be a distinctive mark of the cognitive. The worry is that by reintroducing consciousness into the story the brain-world barrier will once again assume cognitive significance antithetical to the goals of extended cognitivism. And if this were not enough, conscious endorsement does not seem to be a necessary feature of belief for as Clark and Chalmers themselves suggest, perhaps beliefs can be acquired by such means as subliminal perception or some other nefarious means of belief tampering. One might even imagine the entire contents of a telephone directory, none of which having been given prior conscious endorsement by Otto, being transcribed into a silicon chip and then planted in his brain such that he could access that information as reliably, effectively, automatically and transparently as Inga accesses her neuronal memory.[125] No doubt it is such concerns that explain why Clark has more recently dropped all mention of the past-endorsement criterion. For example, in a recent book (Clark, 2001b, p. 156) he suggests that

> It is quite proper to restrict the props and aids that can count as part of *my* mental machinery to those that are, at the very least, reliably available when needed and used

[124] See especially their footnote 5, p. 17. In conversation Clark has also recognised a degree of *ad hoc*-ery about it.

[125] Clark and Chalmers (1998), p. 17. It has been suggested that ordinary perception regularly gives rise to beliefs that are not consciously endorsed, at least if we understand conscious endorsement as involving some kind of reflective process.

> (accessed) as automatically as biological processing and memory. . . . Easy availability and automatic deployment seem to be what matters here.

A similar absence can be found in Clark (2005, pp. 6-7) where, although he does not tackle the problem of cognitive bloat in any direct way, he suggests that the first three criteria suffice to exclude, for example, the book in his library at home, presumably – he does not specify – on the basis of it not being directly available when required. It remains unclear, however, how the first three criteria are supposed to exclude (if, indeed, they *are* supposed to exclude) such repositories of information as telephone directories, that are – or so it seems – easily accessible, automatically endorsed, and available for use as and when required. Without a suitable replacement for the past-endorsement criterion, such allegedly counter-intuitive examples remain.

In trying to defeat premise (2) of the cognitive bloat argument, defenders of extended cognitivism such as Clark and Chalmers have, it seems to me, placed themselves in an apparent dilemma: rejecting the past-endorsement criterion fails to leave one with an intuitively acceptable basis on which to eliminate the contents of such things as encyclopaedias and telephone directories from amongst the repositories of the content-bearing constituents of beliefs. On the other horn, accepting the criterion both rules out certain plausible cases of belief (such as those acquired by subliminal perception) and imports consciousness back into the cognitive story in an apparently *ad hoc*, unprincipled, and possibly damaging way.[126]

The solution I suggest is this: rather than trying to stop, via a fourth criterion, the extension of the content-bearing constituents of belief states to putatively counter-intuitive

[126] Perhaps one might replace the prior conscious endorsement with the criterion that the agent be causally responsible for the content-bearing constituent of their belief. This is a suggestion reportedly of William Lycan's in conversation with Robert Rupert. (Rupert (2004), p. 403, fn.26.) Whilst potentially interesting, this will be moot for the purposes of this paper, since I argue below that a fourth criterion is not needed.

cases, the extended cognitivist should concede premise (2) and argue instead against premise (3). I propose to resist the claim that it is absurd to suggest that amongst Otto's beliefs are the entire contents of encyclopaedias or telephone directories on two main fronts. Firstly, I shall suggest that if we appeal to ordinary usage (as Rupert does) in order to dismiss the extension of cognitive states to the contents of telephone directories and the like, then the extended cognitivist is owed some intuitive or theoretically cogent grounds for this usage. I consider what seems to me to be the most likely candidate for such grounds and further suggest that ordinary usage is considerably less clear and rather more conflicted on the matter than it might at first seem. At least until we are suitably furnished with some viable alternative grounding intuitions, ordinary usage provides a considerably less decisive guide to appropriate belief attribution than a critic needs to suggest.

Secondly, I shall argue on more theoretically oriented grounds that the role of belief attribution in intentional explanation demonstrates no clear basis for excluding the contents of encyclopaedias and telephone directories and the like from amongst the constituents of our beliefs. On the contrary, a close examination of what it to have a belief augurs in favour of precisely such an extension.

§6.4.2 Ordinary Usage

Rupert's explicit grounds for premise (3) rest on a direct appeal to ordinary usage, and it has to be conceded that ordinarily usage very clearly mitigates against saying that someone knows some arbitrary person's telephone number given (in addition to their automatic endorsement, etc.) that they can dial directory services. The interesting and salient question, however, is why ordinary usage comes out against an extension of belief to the contents of telephone directories, etc. Is it based simply on an unschooled and possibly biased habit, or is it perhaps underpinned by a robust and principled intuition? Given the broadly naturalistic explanatory framework presupposed by the thesis, it goes without too much saying that we ought to be cautious about placing too much weight on ordinary usage. There are many examples, both historical and current, where, by our best explanatory lights, ordinary usage carves the world in very incongruous and unprincipled ways. For example, ordinary usage generally places tomatoes amongst the vegetables rather than amongst the fruits, notwithstanding good explanatory reasons to the contrary.[127] Ultimately, ordinary usage can be no substitute in our explanatory endeavours for a careful examination of the relevant facts – a matter over which we shall cast a more careful eye in the next section.[128] Nevertheless, insofar as – but, I would add, *only* insofar as – ordinary usage may be based on robust and principled intuitions does it deserve to be reckoned with.

So what intuitions, if any, might underlie our ordinary common-sense resistance to externally located attributions of belief? And are they robust enough or principled enough to serve the critical ends to which Rupert seeks to put them? There are, perhaps, a number of possible candidates for intuitions driving our ordinary usage with respect to belief attribution and I cannot hope to cover all such candidates here. I will, therefore, restrict myself to a consideration of what seems to me to be the most likely candidate (at least for linguistic cognitive agents[129]), namely this: that someone's having a belief that *P* at time *T* is conditional on whether they are, at *T*, capable of sincerely affirming *P*. Fairly clearly, this intuition would adequately account for our ordinary usage resistance in attributing to a

[127] Tomatoes, I would guess, are thought of as vegetables primarily because they lack the sweetness typically associated with fruit. In other words, the (scientific) falsity of our beliefs about tomatoes may stem, in part, from the scientific irrelevancy of a criterion often associated with fruit.

[128] To be fair to Rupert, he is well aware that ordinary usage (together with any unschooled intuitions that may fuel such usage), will count for little within a naturalistic explanatory framework (see Rupert, 2004, p. 390).

cognitive agent beliefs relating to the entire contents of telephone directories. Suppose, for example, that I had the relevant telephone directory right in front of me and someone asked: 'What is Rob Rupert's telephone number?' Prior to consulting the directory I would be quite incapable of providing a sincere response, whilst after consulting it, at which point the information will have made its way into my brain, I would be entirely capable.

The problem with this as a grounding intuition for ordinary usage is that it is not particularly robust – in other cases ordinary usage seems incompatible with such an intuition. All of us, I suspect, will be familiar with having a sense of knowing someone's name whilst being temporarily entirely incapable, no matter how hard one might try, of saying what it is. This may sometimes continue for days or even weeks (perhaps accompanied by periodic spells of the mouthing of various sounds in order to try to retrieve the information from its hidden depths). Often the name will pop up when least expected – sometimes, it never does. In any case, at the end of the struggle, and assuming the name is remembered correctly, it would seem entirely in keeping with ordinary usage to attribute knowledge retrospectively – that is, to say, notwithstanding the difficulties one may have had recalling it, that one knew the name *all along*. And if one knew it all along then, presumably, one also believed it all along.

Moreover, in addition to retrospective knowledge attribution, to some degree ordinary usage sanctions the attribution of knowledge to individuals prospectively as well. We sometimes confer a degree of 'epistemic credit'[130] in lieu of a capacity to provide, in short enough order, a sincere and accurate answer to a question. For example, were I to

[129] For non-linguistic putative believers the considered criterion will not do. I leave that for the critic to worry about.

[130] My use of this phrase is unrelated to that in Clark and Chalmers (1998).

ask a telephone directory operator 'Do you know Rob Rupert's telephone number?' it does not, it seems to me, do significant violence to ordinary usage for the operator to respond 'Yes I do sir. Just give me a moment whilst I look it up.' In such cases the knowledge attribution goes proxy for the claim that the cognitive agent is capable of delivering a sincere and (in the case of knowledge attributions, justifiable and accurate) response to the question within an acceptable period of time. How long it takes before such epistemic credit becomes 'defaulted', so to speak, is generally left rather vague and will be, to a large extent I am sure, dependent on context. The salient point here is that ordinary usage often sanctions attribution of belief independently of the having of particular capacities at the time to which belief is attributed. Attempts might, of course, be made to revise the intuition to take account of such cases. If, however, we weaken the intuition to improve its robustness by allowing the temporal restrictions to vary with context, it becomes considerably less clear that ordinary usage has much to say, at least on any *principled* grounds, against including the contents of telephone directories and the like. For example, if it were suggested that someone having a belief that *P* at time *T* is conditional on whether they are, at some more or less vague and indeterminate time *U*, capable of sincerely affirming *P*, then it fails to account for our ordinary usage, for such a condition would be satisfiable in the case of the contents of telephone directories.

It would, of course, be rather presumptuous to suppose, given only the inadequacy of one candidate grounding intuition, that no such intuitive and principled grounds can be given for the ordinary usage to which Rupert alludes, but if there are any such grounds I suggest that it remains for the critic to provide them. My suspicions are that any resistance in the way of ordinary usage to the attribution of beliefs under such circumstances will be found to stem from a question-begging bias based on the fact that telephone directories are repositories of information entirely external to our bodies.

§6.4.3 Believing as a Causal-Functional Relation

Let us leave ordinary usage where it lies and turn our attention instead to rather more theoretically motivated and, from the point of view of extended cognitivism, rather more positive considerations. In what follows I shall assume a number of things without further defence. In the first place I take a naturalistic and realistic stance with respect to beliefs. That is, I shall suppose that there are such things as beliefs, that they supervene on the material, and that they are causally efficacious in our behaviour. In so doing I shall ignore issues surrounding the status of folk psychological explanation. This is not, I hasten to add, because I suppose the status of folk psychology to be unimpeachable or free of problematic issues, but simply because my concern here is not to defend folk psychology; it is to argue that it extends (warts and all) very naturally to there being beliefs lying external to the body or brain of the cognitive agent.

In addition, instead of trying to give a characterisation of *belief* qua object, I shall approach the issues via an examination of the nature of *believing* or, in other words (to take a particularly common locution), the *having* of a belief by a cognitive agent. From such a perspective, a belief can be provisionally defined in terms of being a content-bearing state with respect to which a cognitive agent stands in a believing relationship. The central task, as I shall approach it, is to characterise the nature of the believing relationship. Any and all states that stand in that relationship to a cognitive agent will, ipso facto, be beliefs. The central reason for focussing primarily on believing rather than belief, qua object, is that it more effectively places at centre stage the functional and explanatory role that beliefs play in our intentional behaviour. Beliefs, I suggest, are much as Ramsey puts it: 'maps by which we steer.' And like maps, they are content-bearing structures that are type-identified most appropriately in terms of the general functional role that they play in our cognitive lives.

§6.4.3.1 Kinds of Relations: Physical, Corporeal, Functional and Causal-Historical

In order to better appreciate the general irrelevance of whether belief-states are internally located, it will be well to examine more closely what general *kind* of relation characterises the way that individual cognitive agents stand with respect to their beliefs (or, for that matter, mutatis mutandis, other of their intentional states). The nature of the believing relation, I argue, relegates the internal location of standard examples of beliefs to, at most, a secondary and contingent significance. To see this, let us first distinguish four important ways in which an object *O* stands in relation to an individual *S*, which I shall label respectively, the *physical*, the *corporeal* (a kind of physical relation), the *functional*, and the *causal-historical.*

Physical relations are perhaps the simplest to describe: *O* stands in a physical relation to *S* when *O* stands in a spatial relation to *S*. 'Spatial' relation, might, therefore, do just as well. This relation is, of course, trivially satisfied for every spatially located object. Of considerably more interest for our present concerns, however, are the physical relations between *S* and those objects *O* that can be said to be 'on or about *S*'s person'. For example, *S* might stand in such a relation to a scar on his right leg, his socks, his two arms, the pen in his pocket, or his appendix. Let us call this subclass of physical relations, *corporeal* relations. For many objects, at least, this involves more or less vague boundary conditions; for example, there is no particularly determinate distance that *S* will be from a pen, before *S* can no longer be said to stand in a corporeal relation to it.

An object *O* stands in a *functional* relation to *S* when it is the function of *O* to bring about one of *S*'s functionally specifiable goals.[131] More precisely what this functional relationship amounts to will depend on the item in question and the functionally specifiable goals. For many things, such as *S*'s right arm, this will mean *S* being able to control it in some functionally appropriate way (for lifting, etc.). But for other things, *S*'s apartment say, it will mean little more than *S* being able to access it for use as and when desired. In general, functional relations require that there be some functionally appropriate causal relation between *S* and *O*. Let us call this subclass of functional relations that require a causal relation, *causal-functional* relations. Not all functional relations are causal-functional relations. For example, there need be no ongoing causal connection between a dog owner and a guard-dog in order for the dog to fulfil its functional role for its owner. Sometimes the mere presence of something is sufficient to fulfil a functional role for an agent. Again, as with corporeal relations, functional relations come by degrees with somewhat vague boundary conditions.

The link between functional relations and corporeal relations is often very close in that *S* standing in a functional relation to *O* is *typically* causally dependent upon *S* standing in a corporeal relation to *O* – of *O* being on or about *S*'s body. For example, in general, one will not stand in a functional relation to a pen unless one also stands in a corporeal relation to a pen. Pens are not generally capable of fulfilling their functional role unless one has them relatively close enough at hand. Conversely, and only slightly less generally,

[131] This relation is intended to be neutral between various kinds of functions, such as Cummins functions or historical functions. Additionally, there may be, of course, some *O* for which *S* performs a function. I take it that such relations are of no present concern.

the pens in which one stands in a corporeal relation are precisely those in which one stands in a functional relation – pens that do not work one soon discards. Importantly, for our present purposes this is not always the case. One can stand in a functional relation to something even though it may be physically at some considerable remove from oneself; for example, where *S* is a model aircraft enthusiast and *O* is a remote-controlled model aircraft, or where *S* is a NASA scientist and *O* is NASA's Mars Exploration Rover. Conversely, *S* can also stand in a corporeal relation to an object *O* without standing in a functional relation to *O*. For example, where *S* has an appendix (there being no function that *S*'s appendix serves for *O*).

Lastly, we have any of various *causal-historical* relations in which an object *O* stands to S. For example, *S* having a mother in London, England. One causal-historical relation of particular importance for the present chapter is the *authorial* relation – i.e., when someone is the originator, creator, or author of something. It is in this sense that we might say, for example, that Picasso stands in authorial relation to many paintings in MoMA. This is so precisely because Picasso created those paintings.

So which of these broad kinds of relations characterise believing? Or, in other words, which best characterises the relation between individual cognitive agents and the content-bearing states to which they stand in a believing relation? The first I consider is the causal-historical relation, for it seems closest to that implicated in Clark and Chalmers' past-endorsement criterion: the beliefs in Otto's notebook were held to be such (in part) because *Otto* endorsed them and wrote them down as a conscious act. He is, *prima facie*, the *author* of the content-bearing inscriptions of the notebook. In appealing to authorship here we need to be clear that Otto is not, at least not generally, the author of the *contents* of his beliefs, only, at most, the content-*bearers*. Consider, to take an everyday in-the-head example, the belief that the Taj Mahal is in India. Most of us can claim little creative control over its content since it is most likely something that we read in a book or saw on television as a child, and pretty much immediately and uncritically accepted. Otto, we may note, is very likely to be in much the same position, having presumably accepted (perhaps via a guidebook or local passer-by) that MoMA is on 53rd Street, with similar automaticity and lack of critical attention. In recording the location of MoMA in his notebook, what Otto does is create his own relatively private token expressing a content that he more or less uncritically accepts. This is true, I suggest, of most of our everyday beliefs, the contents of which are, in general, very much second-hand in that respect.

Where our claims to be authors of our beliefs stand on rather more solid ground concerns the extent to which we are intimately involved in the causal processes that result in the existence of such content-bearers. Even as mere 'copyists', we are typically involved in a very special and intimate way with the causal process through which the content-bearing neuronal states come into being in our brains, and Otto is presumably likewise intimately involved in the causal story concerning how the content-bearers gets to be in his notebook. But we have already had cause to note that various nefarious means of getting such states into our brains (and into the appropriate attitudinal relationship with ourselves) cast considerable doubt on any supposed necessity of our direct personal or causal involvement but without, thereby, undermining either the claim that we have such beliefs or the role that appeal to such beliefs might go on to play in explaining our behaviour. As with Inga's brain, content-bearing states could be surreptitiously inserted into Otto's notebook without undermining in any way its subsequent role in explaining Otto's intentional behaviour. Quite irrespective of the role each of us may play in the creation of our beliefs, they remain amongst our beliefs just so long as they are available to play the relevant explanatory role in our behaviour. Note that such content-bearing states may count amongst our beliefs even if they are never *actually* called upon in the determination of behaviour. What matters here is their functional availability such they *would* serve to determine behaviour where appropriate. Otto (like Inga) believes that MoMA is on 53rd street regardless of whether that belief is ever called upon to guide behaviour.[132]

Turning now to corporeal relations, we manifestly stand in precisely this relation to our in-the-head content-bearing states, very much as Otto stands, most of the time at least, to the contents of his notebook.[133] The question before us, however, is whether, or to what extent, corporeal relations are relevant to the attribution of beliefs to cognitive agents in the explanation of their intentional behaviour. Any importance attributed to corporeal relations is, I argue, superficial, and parasitic upon the fact that it is typically associated with cognitive agents standing in a causal-functional relation to such states.

Consider the case of Olga who, like Inga, was informed of the location of MoMA and who likewise tokens a content-bearing neuronal state to this effect in the usual way. Additionally, both token a type-identical attitudinal disposition towards the contents of their respective neuronal states, i.e., were either to access their respective content-bearers, each would endorse its content and behave accordingly. The sole difference between Olga and Inga is that due to some unfortunate damage to certain pathways of Olga's brain, she has lost all *causal* access to the said neuronal state, and so no longer stands in the relevant

[132] One might wish to distinguish between *mere* belief – that is, any and all beliefs that may play an explanatory role in one's intentional behaviour – and those beliefs that play a metaphysically significant role in personal identification – presumably a subclass of mere beliefs. Sometimes, for instance, one may act on someone else's testimony, whilst at the same time distancing one's *self* from such beliefs. Such beliefs are not one's *own*, in some such metaphysically significant sense. Such distinctions do not affect the present issue, however. From the point of view of intentional explanation surreptitiously inserted beliefs will count amongst an intentional agent's beliefs, notwithstanding that they may not be amongst *his* or *her* beliefs in some more metaphysically loaded personal sense of belief.

causal-functional relation to that state. The said content-bearing neuronal state remains unchanged by the damage that occurred elsewhere in Olga's brain. It seems, therefore, reasonable to maintain that both Olga and Inga continue to stand in a relevantly similar *corporeal* relation to a content-bearing state carrying the content that MoMA is on 53rd street.[134] Despite this, it would be explanatorily idle to suggest that she still tokens any such belief, precisely to the extent that the neuronal state is rendered incapable of playing any further role in explaining her intentional behaviour. Given that the lesion renders the content-bearing state permanently inaccessible to Olga, she no longer stands in the relevant causal-functional relation to it, and so in all cognitively relevant respects, she to have that state as a belief. This is true even though she continues to have the relevant attitudinal disposition towards that state. Despite such a disposition, so long as the state is permanently incapable of guiding her behaviour it cannot count amongst her beliefs, not even amongst her dispositional beliefs.[135] Her believing, therefore, is not a corporeal relation, but a causal-functional one.

The above argument is, perhaps, somewhat 'quick and dirty', and I shall add some detail to the causal-functional relation of believing below (§6.7) by offering some revised conditions for belief attribution. For the moment, however, let us note that this conclusion

[133] He probably won't take it into the shower with him. But again, though, the corporeal relation has vague boundaries so it is not especially clear how much such occasions matter.

[134] Some people's intuitions differ here. At least one commentator has suggested that a state or structure can be considered content-bearing only so long as there is an active means of accessing that content. Olga's neuronal state formerly carrying the content that MoMA is on 53rd St. would lose its content-bearing status as soon as she has acquires her unfortunate brain damage. But such a view would have the (in my view counter-intuitive) consequence that prior to the discovery of the Rosetta Stone hieroglyphic inscriptions had not been content-bearing for millennia. Additionally, notwithstanding Olga's ongoing inability to retrieve the content from that state, it might yet remain possible for it to be retrieved, in principle at least, by someone armed with the neurological equivalent of Olga's Rosetta Stone.

should be no great surprise given the general demands of intentional explanation. In order for the standard folk-psychological attribution of belief states to individuals to have any explanatory purchase in respect of their intentional behaviour, we need to suppose that the representational states or structures are available to do the functionally appropriate kind of causal work. The reason why Olga does not believe that MoMA is on 53rd street is that, unlike Inga and Otto, she does not have, in the appropriate causal-functional sense, the relevant content-bearing state. Inga and Otto, by contrast, stand in the functionally appropriate relations to their respective content-bearing states such that, together with the appropriate attitude of endorsement and the relevant desire, their accessing those states would help to explain their museum-going behaviour.

§6.5 – THE OTTO 2-STEP

A second objection to the Otto example is what Clark (forthcoming) calls the 'Otto 2-step' and which he cites as the most common objection to the Otto example. Here's how Clark characterises it: 'all Otto actually believes (in advance) is that the address is in the notebook. That's the belief (step 1) that leads to the looking (step 2) that then leads to the (new) belief about the actual street address.'[136] Houghton (1997) considers a similar objection:

> It is gratuitous, the objection goes, to attribute intentional states to a person unless and until the content of the states is fully internalised. For only when the content is internalised can the states play a role in explaining the subject's behaviour and psychological development. Beliefs and desires cannot serve as mental causes and move us to act while their contents are located wholly outside us.

[135] The stipulated permanence of its inaccessibility is intended to remove the wiggle room and vagueness that otherwise occurs if it is only temporarily inaccessible. As with temporarily forgetting someone's name (§6.4.2) in many contexts we might be inclined to continue to attribute to Olga the relevant belief just so long as she *eventually* regains access, or regains access within some vague time period. Notwithstanding such explanatory 'promissory notes', however, so long as the content-bearer would be inaccessible were it needed, it cannot play a role in intentional explanation of her behaviour and counts amongst neither her occurrent nor dispositional beliefs. For more on dispositional (i.e., non-occurrent) beliefs, see §6.7.

[136] Clark, forthcoming, p. 7. See also Clark and Chalmers,1998, p. 13.

The thrust of this criticism is that it threatens to remove all explanatory power from extended cognitivism. Externalised 'beliefs' become cognitively inert and explanatorily idle.

Before seeing our way past this objection, let us note a couple of responses that do not work. One might be to suggest that it proves too much inasmuch as it would render all internally stored non-occurrent beliefs explanatorily idle unless or until brought to consciousness, 'before the mind'.[137] But this seems to miss the central thrust of the objection. It is not that a belief must be brought to *consciousness* to be explanatorily relevant – *un*conscious beliefs might be just as explanatorily relevant to one's behaviour. The central thrust of the objection is that beliefs (whether conscious or unconscious) must be *inside* the cognitive agent to do any explanatory work in respect of our behaviour.

Clark (forthcoming) presents another response (actually a restatement of the response in Clark and Chalmers [1998]) which is less than convincing. Clark asks why we do not depict Inga in similar terms, i.e., as having (initially) simply the belief that the location of MoMA is in her neuronal belief-storage system, and then after retrieval of that information, as having a second belief as to the location of MoMA. Clark suggests that this explanation adds spurious complexity to the Inga case. Surely, Inga does not rely on any beliefs about her belief-storage as such, she just uses it, transparently as it were. He goes on:

> But ditto (we may suppose) for Otto: Otto is so used to using the book that he accesses it automatically when bio-memory fails. It is transparent equipment for him just as biological memory is for Inga. And in each case it adds needless and psychologically unreal complexity to introduce additional beliefs about the book or biological memory into the explanatory equations. (Clark, 2005, p. 8.)

[137] Houghton (1997), p. 170 tries this.

There are a couple of worrying aspects to this response. The first is that the transparency or otherwise of Inga's *use* of her neuronal belief-storage system or Otto's *use* of his notebook is beside the point of the objection. A critic may concede that Otto uses his notebook (as Inga uses her brain) transparently, that is, without further appeal to beliefs, without in any way retracting the objection. The objection is that the explanatory work is done in two stages neither of which appeals to an externalised belief. Firstly, there is an internal belief that the information is in his notebook, and that explains why Otto goes to his notebook looking for the address of MoMA. Once that is initiated, the actual use of the notebook and the subsequent retrieval of the information it contains may be as transparent as Clark might wish to suggest. The rest of his behaviour – Otto heading out to MoMA – is then explained by the second belief generated by the (we may suppose, transparent) accessing of the notebook. Notwithstanding whether Otto *uses* his book transparently – i.e., without conscious appeal to beliefs (or desires) – a critic may nevertheless appeal to his prior belief that the information is in the notebook in the explanation of why that notebook-consulting process was *initiated.*

In order to properly address the critics objection, therefore, the appeal to transparency or automaticity needs to be in relation to the *initiation* of Otto's notebook-consulting behaviour, not the execution of that behaviour. In other words, we would need to suppose that Otto initiates his notebook-consulting behaviour without an explanatory appeal to a belief. Granting Clark this claim, however, there yet remains what I take to be the central thrust of the objection: regardless of whether we appeal to beliefs and desires, to unconscious (and transparent) neural processes, or what have you, we must nevertheless appeal to something *inside Otto* to explain his behaviour, both for his notebook-consulting behaviour and his MoMA visiting behaviour. In the case of belief-desire explanation, our explanatory strategy must surely somehow bring Otto's belief *together with* his desire to visit MoMA before it can do any explanatory work. But his *desire* is (presumably) in Otto's head, and never leaves his head. Therefore, before an appeal to Otto's belief (that MoMA is on 53rd street) can combine with his desire (to visit MoMA) so as to explain his behaviour, the belief must *also* be inside Otto's head (similarly, mutatis mutandis, for other styles of cognitive explanation). What remains outside Otto, the critic might yet insist, cannot play an explanatory role in Otto's behaviour.

§6.5.1 – The Speeding Ottoman

Clark and Chalmers seem rather oblivious to the force of the above concern – a result, I suggest, of a failure to adequately distinguish between the explanatory roles of occurrent and non-occurrent (i.e., dispositional) beliefs. A belief is *occurrent* if and only if it is currently playing a role in an intentional process. It does not matter, therefore, *how* it plays its current role. It may do so either consciously or unconsciously; explicitly or implicitly (see §4.1.2). Such beliefs may play a role in the production of overt behaviour (such as the belief that there is beer in the fridge in fridge visiting behaviour) or not, as the case may be (for example, the belief that it will rain next Sunday in the formation of the thought that it might be a good idea to cancel the proposed picnic). I shall add some greater precision to the notion of *non-occurrent* belief in §6.7, but in broad outline we can say that a non-occurrent belief is one that is functionally available to play a role in an intentional process, but is not currently playing such a role. Non-occurrent beliefs, therefore, are never conscious, for to be conscious is (I take it) to be *presently occurring* (hence, occurrent) in the consciousness of a cognitive agent. Again, like occurrent beliefs, non-occurrent beliefs may be available to play their role either explicitly or implicitly.

Clark and Chalmers, however, appear to treat occurrent and dispositional beliefs on an explanatory par. They suggest, for example, that '[w]e are happy to explain Inga's action in terms of her occurrent desire to go to the museum and her *standing* belief that the museum is on 53rd street, and we should be happy to explain Otto's action in the same way' (1998, p. 13, emphasis added). In the case of Inga, both the *standing* – i.e., *non-occurrent* – belief as well as the *occurrent* belief that is directly driving her museum-going behaviour lie within Inga's head. It is in part because they are both located within Inga's head that we might (incautiously) appeal to her non-occurrent belief in the explanation of her behaviour. But the extended cognitivist, if they are to respond effectively to the Otto 2-step objection, cannot afford to be less than precise.

To see this, consider the case of Ottoman, a conscientious and law-abiding citizen, who is caught speeding down the highway at 145 kmh having uncharacteristically failed to consult his speedometer.[138] When stopped by a police officer and asked whether he believed that he was speeding he replies – quite sincerely and accurately – that he did not; that he did not consult his speedometer because he was (regrettably) exceptionally distracted after an argument with his spouse. Fairly clearly, it would be antithetical to the demands of intentional explanation to continue to insist, given his desires and his behaviour, that Ottoman really believed that he was speeding.[139] But given only Clark and Chalmers' first three criteria, it becomes difficult, at least without further qualification, to resist the claim that the information carried by his speedometer must count amongst his belief states. It seems entirely reasonable to say that Ottoman will automatically endorse

[138] My thanks to Dominic Lopes for this example and for pressing my on this point. Similar examples abound, such as missing an appointment through forgetting to check one's watch.

the information on his speedometer whenever it is consulted, that that information is directly available without the slightest difficulty, and that his character is such that he is rarely to be found driving without regularly and frequently consulting it.

As suggested, the solution here rests on clearly distinguishing occurrent from non-occurrent beliefs and particularly in taking note of their different explanatory roles and their associated different conditions of attribution. It is only *occurrent* beliefs that play a *direct* explanatory role in driving intentional processes and behaviours, and they play that role only to the extent that they are occurrent. For example, the intentional explanation of Ottoman's behaviour, given his desires, appeals directly only to his occurrent belief that he was not speeding. This is entirely consistent with him also having a non-occurrent belief very much at odds with his occurrent belief – a circumstance particularly familiar in cases of temporarily forgetting something relevant to one's behaviour. By contrast with occurrent beliefs, non-occurrent beliefs, by their very nature, play only a more indirect explanatory role; a role that lies at one remove from the role of occurrent beliefs; namely, in being amongst the primary sources of one's occurrent beliefs. Although they do not usually state their criteria specifically in such terms, what Clark and Chalmers attempt to provide is best characterised as a set of conditions for non-occurrent belief rather than occurrent belief.[140] With this firmly in mind, it becomes clearer that Ottoman's speeding poses no direct challenge for the extended cognitivist, for whom it remains open to maintain that whilst Ottoman occurrently believed that he was not speeding, he non-

[139] There are, of course, cases – for example addictive behaviour – where folk-psychological explanation has to make considerable appeal to implicit ceteris paribus clauses in order to preserve such central generalisations as 'if S desires P and believes that doing A will achieve P, then S will do A'. I bypass such issues for the aforementioned reason that I am not concerned to defend folk-psychological explanation but simply to argue for its natural extension to externalised beliefs.

[140] Clark and Chalmers switch between talking about 'standing' belief, 'non-occurrent' belief, and simply 'belief' without much caution. For example, when they present their criteria they present them simply as criteria for 'belief'. Yet these criteria are surely more appropriately applied specifically to non-occurrent belief.

occurrently believed that he was doing 145 kmh – a state that would have resulted in a corresponding occurrent belief were it not for the exceptional circumstance. Similarly, the extended cognitivist can concede the thrust of the Otto 2-step objection – that belief-desire intentional explanation depends on bringing occurrent beliefs together with occurrent desires, but without tarnishing in any way the central explanatory role that external content-bearers play in non-occurrent belief.

§6.6 – THE PERCEPTION OBJECTION

An objection very closely related to the Otto 2-step places emphasis on the *interceding* role of perception between internal cognitive processes or states and putative external cognitive processes or states. The objection is that Inga, unlike Otto, does not depend on perceptual processes to make her way to MoMA, thereby creating an important disanalogy between the two cases. Butler (1998, p. 211) suggests that 'the very fact that the results are achieved in such remarkably different ways suggests that the explanation for one should be quite different from the explanation for the other'. Otto, he points out, has to look at his notebook, but Inga does not have to look at anything.

This objection can be examined on both counts – namely, whether it *is* in fact disanalogous, and whether it is an *important* disanalogy, given that the disanalogy holds. Clark (2005, p. 26) gives a response in the first respect when he points out that relative to Otto and his notebook considered as a single, extended cognitive system, 'the flow of information is wholly internal and functionally akin to introspection'. The functional analogies between imagistic introspection and perception have been supported by a considerable body of research. Drawing on such research, Kosslyn (1994, p. 74) argues that '[o]nce a pattern of activation is evoked in the visual buffer, it is processed in the same way, regardless of whether it was evoked from input from the eyes (perception) or from memory (imagery)'. Of course, such functional parallels may be absent in many cases. In the case of Inga, there is no suggestion that she literally calls before her mind an image of (let's say) the book in which she first read of the location of MoMA and then scans it. There remain, therefore, plausible functional differences between Inga and Otto at least insofar as the specific cognitive mechanisms that they employ are concerned.

Turning, then, to the second count, it remains unclear why functional differences appearing at a fine-grained level of analysis should constitute a damning objection to extended cognitivism. For example, whether we gather the information that there is a fire in the building via someone shouting 'Fire!' and using language (and aural) processing, or by seeing flames and using visual processing, will make little difference to the broad structure of the belief-desire explanation of our fleeing behaviour. The basic explanatory strategy for such behaviour is (roughly) that we came to believe that there was a fire, we do a bit of quick practical reasoning to determine that fleeing would be the behaviour best placed to maximise the fulfilment of our desires. Where we get the belief is a feature of the finer-grained functional analysis, and at *that* level of functional analysis differences certainly do appear. Rupert (2004, p. 409ff.) makes much of essentially this issue in the case of memory. Internal memory, he points out, exhibits certain features absent from putative cases of external memory (external belief-storage, as we have been terming it here). But such differences do not make a salient difference to their status as external belief-storage to the extent that they fulfil the same functional role within the cognitive economy of the system concerned. Aural, visual, and olfactory systems, for instance, are remarkably different in various respects, but each performs a *perceptual* role for the system of which they are a part.

Rupert resists the move to a broader level of functional description by pointing to the fact that diversity amongst various kinds of internal memory systems is yet unified, at least to a substantial extent, by certain features that are absent in putative external cases.[141] But to place emphasis on certain unifying features of *internal* memory begs the question against the extended cognitivist who sees external belief-storage as fulfilling a broader functional role than might be captured by reference to the exhibiting of such properties in the case of many specifically internal memory systems. This in no way denies (as Rupert seems to think – see p.420) that such differences between kinds of belief-storage systems will be of great explanatory interest. Otto and Inga may well exhibit interesting differences in their behaviour as a direct result of the different ways in which they store their memories and the different ways in which their respective memories function. Some of those differences may well be explained in terms of the fact that Otto has to look at his notebook. But none of that gives any clear grounds to not categorise Otto's notebook as a member of the broad category of belief-storage devices.

§6.7 – SOME REVISED CONDITIONS FOR BELIEF

In the above I have laid great emphasis on our having beliefs in virtue of our standing in a certain causal-functional relationship with content-bearing states or structures. As a result, the supposed significance of having one's beliefs located within one's body or brain will, I hope, have begun to recede. In the light of such considerations, and recalling the distinct explanatory roles of occurrent and non-occurrent beliefs, I close with some revised criteria for the counting of content-bearing states amongst our beliefs. In any case, more can now be said to flesh out and strengthen the explanatory parallels

[141] Rupert (2004), p. 419. For instance, Rupert cites the Rescorla-Wagner law.

between internal and externally located belief states by taking a closer look at those functional relations. Leaving aside the problematic past-endorsement criterion, Clark and Chalmers provide us with three other criteria designed to satisfy this end, but certain refinements prove to be necessary before we can present extended cognitivism with respect to belief in its best light.

For one thing, Clark and Chalmers' first criterion suggests that the information store, be it a neuronal belief-storage system or a paper and ink notebook, will be a constant in the agent's cognitive life. But this will not quite do as it stands. For one thing it differs slightly from the other criteria in that it focuses on a functional aspect of the *repository* of beliefs – of such things as Otto's notebook and, by parallel, Inga's neuronal belief-storage system – namely, that the object is a constant in the agent's cognitive life. Whether Otto's notebook is a functional equivalent of Inga's neuronal belief-storage system is, of course, entirely relevant to the issue of extended cognition, but its status as such will presumably be secondary to, if not wholly dependent upon, the status of its contents as functional equivalents of Inga's beliefs stored neuronally. What we are after, therefore, at least in the first instance, are the functional requirements of belief.

For *occurrent* beliefs some broad-stroke variation of the standard folk-psychological intuition should suffice. For example, we might say that some content-bearing state or structure *C* counts amongst *S*'s occurrent beliefs if and only if *C* plays a direct causal role with *S*'s desires in one of *S*'s current intentional processes. As mentioned earlier, occurrent beliefs can be either conscious or unconscious. In Ottoman's case, it is quite plausible that his occurrent belief that he was not speeding was unconscious, since he was consciously occupied with his recent argument with his spouse.

For *non*-occurrent (i.e., dispositional) beliefs I propose the following amendments to Clark and Chalmers' first three conditions: with respect to some intentional behaviour *A*, some content-bearing state *C* may be said to count amongst *S*'s non-occurrent beliefs if and only if:

1) *S* has a disposition such that where the content that *C* carries is functionally relevant to *S* doing *A* then *S* will retrieve from *C* its content.
2) *S* can directly retrieve the content carried by *C* without difficulty.
3) If *S* retrieves the content carried by *C*, then *C* will directly cause *S* to have an occurrent belief with that content.

The first of these criteria replaces Clark and Chalmers' first criterion and differs from it primarily in focusing not on the *repository* of content-bearers (brains, notebooks, etc.) but specifically on the central target – the content-bearer. Brains and notebooks derive their status as repositories of the content-bearers of beliefs from the disposition of cognitive agents towards those contents. Being a dispositional criterion it also includes an important ceteris paribus clause that renders it usefully weaker than Clark and Chalmers first criterion. For instance, Clark and Chalmers' criterion suggests that Otto will not take action without consulting his notebook. This, presumably, was intended to help preserve parallels with the special role that brains play for in most people's cognitive lives. But Otto might be a sensibly cautious fellow who makes multiple back-ups of his notebook, perhaps in other notebooks or on a computer, say. Each and any of these repositories may contain the content-bearers of his non-occurrent beliefs, with none being privileged over the others. Similarly, we need to allow that someone may have some non-occurrent belief even when, on certain occasions, other sources provide for an occurrent belief with an equivalent content. For example, Otto may encounter a passing New Yorker who, seeing him with map in hand, asks him if he needs directions somewhere, and who tells Otto where MoMA is located. A more common example occurs when we require a little time to retrieve some information from our belief-storage system (such as the example mentioned earlier of recalling someone's name) and have our neuronal belief-storage system pre-empted by someone else providing the information.

The second criterion is similar to Clark and Chalmers' only expressed in slightly more precise terms. Its importance lies primarily in allowing us to exclude cases where we have a content-bearer but cannot easily retrieve its content. For example, where we have something in a language that we do not understand, or when the agent is very drunk. How difficult access must be before it fails this condition involves us in an unavoidable degree of vagueness.

The third criterion is intended to render more precise and perspicuous the essence of Clark and Chalmers somewhat opaque 'automatic endorsement' criterion. It does so by capturing the relationship between non-occurrent beliefs and the occurrent beliefs – endorsing, I take it, implies occurrently believing – that are directly responsible for intentional behaviour. The purpose of this criterion is to avoid over-extending non-occurrent belief to content-bearers that do not give rise automatically to occurrent beliefs. These include content-bearers located in sources that we do not generally trust (e.g., most internet content, the contents of *The National Enquirer*) and those located in a source that we generally trust but which express propositions in direct conflict with other more strongly held occurrent beliefs (e.g., the readings of a speedometer stuck on '0' when we are manifestly speeding down the highway). Such content-bearers are causally inert inasmuch as they will not cause us to form occurrent beliefs with their content and, therefore, will not count amongst our non-occurrent beliefs. For this reason, 'automatic' endorsement does not amount to *unconditional* endorsement, nor to endorsement regardless of content. The point of the 'automatic' qualification is to avoid *overly* conditional endorsement such as that which arises arise only after significant and prolonged critical evaluation. 'Automatic', therefore, is here to be understood as requiring no further reflection beyond, at most, a (typically brief and unconscious) examination of the coherence of the content with occurrent beliefs.[142]

If these conditions are, as I suggest, adequate to the task of characterising which states or structures should count amongst our non-occurrent beliefs, then it should be clear that the states or structures contained within encyclopaedias and telephone directories – just as much as neuronal states or structures – *can* and sometimes *do* contain the content-bearers of our non-occurrent beliefs. We often have precisely the relevant kind of functional access to the contents of such technological resources.

§6.8 – CONCLUSION

In this chapter I have examined and defended a stronger version of extended cognitivism, one that includes a commitment to specifically *representational* extended cognitive processes and states. After discussing the pervasive role of cognitive technology in enhancing our cognitive powers, and defending a specifically representationalist interpretation of the examples considered in the preceding chapter, I introduced the example of Otto and his notebook. Following Clark and Chalmers, I argued that the role that the contents of Otto's notebook plays in his cognitive economy qualifies those

[142] I am grateful to Eric Hochstein for forcing on me clarity regarding the 'automaticity' of endorsement.

contents as amongst his beliefs. I then considered and rejected three significant objections raised against this claim. Firstly, I argued in favour of embracing the central thrust of the cognitive bloat objection and for the inclusion amongst our beliefs of certain counter-intuitive cases such as encyclopaedias and telephone directories. I defended this position in two ways. Firstly, I re-affirmed the often made point that naturalistically inclined explanations do not kow-tow to ordinary usage. Secondly, I argued in support of the claim that the having of a belief is a kind of causal-functional relation between agents and content-bearing states, and that as such the location of such states in relation to the agent is of only secondary significance – secondary, that is, to the satisfaction of that causal-functional relation. The second of the objections – the *Otto 2-step* – argued that the use of external representational resources is a 2-step process, with only the internal being properly intentional. I rejected the objection primarily through a distinction between the explanatory roles of occurrent vs. non-occurrent beliefs. The third objection argued that the role of perception provides a principled barrier between cognitive processes and states occurring within the head and other (putatively non-cognitive) processes or states occurring outside of the head. I argued that such objections focus on characteristics of no significance for the broad functional kinds at issue. Finally, I fleshed out my account by reference to a set of putatively necessary and sufficient conditions for non-occurrent belief based loosely on those set out in Clark and Chalmers (1998). On that account it becomes apparent that external content-bearers often do fulfil precisely the explanatory role that we expect of non-occurrent beliefs. The cognitive extension of our belief states out into the world, I conclude, is considerably more in keeping with intentional explanation and our intuitions about beliefs than it might, at first, appear.

CHAPTER 7 – CONCLUSION

Over the past several chapters we have journeyed over some considerable terrain. Beginning with an examination of the general character of cognition, we have gone on to examine the nature of cognitive representation, embedded cognition, para-information, and extended cognition, the last of which in both a general and explicitly representational form. Within these general topics we have had cause to light upon a number of other issues, including the general character of dispositions, the role of information in cognition, arguments for anti-representationalism, and the conditions for belief attributions. The central goal, of course, has been to elucidate an account of extended cognitivism and to defend that thesis against many of the arguments ranged against it, and two chapters have been devoted specifically to that task.

§7.1 – SUMMARY OF THE MAIN ARGUMENTS

The dissertation was divided loosely into two parts. The first part – chapters 2 and 3 – laid the groundwork for an understanding of the key concepts that are essential to assessing the cogency of the claim that an individual's cognition sometimes extends beyond the boundaries of their brain or body: *cognition*, *cognitive representation*, and *information*. The general strategy in these chapters was to move from the outside in, as it were; that is, to begin with the broader issues and then to focus progressively more narrowly on the issues and concepts that arise out of the broader analysis. Specifically, I began with an examination of the general character of cognition, which in turn raised questions concerning the nature of cognitive representation and the role(s) of information in cognitive explanation. The second part – chapters 4 through 6 – developed and defended the central substantive claims relating to extended cognitivism. The strategy there was to move from a defence of weaker claims to stronger claims. Thus, I began with embedded cognition, moved next to a general account of extended cognition, and finished with a defence of a specifically representational version of extended cognitivism. Chapter 7 – this chapter – offers some concluding remarks.

§7.1.1 Chapter 1 – Introduction

This chapter provided a brief history of the origins and evolution of the dissertation and introduced, in a basic way, the thesis of extended cognitivism. The strategy for the defence of the central thesis was explained and the broad structure of the dissertation was laid out. A number of important assumptions were put in place and a number of issues were identified as orthogonal to the thesis and hence excluded from further consideration. The first chapter closed with a summary of the chapters to come.

§7.1.2 Chapter 2 – Cognition

The dissertation began with an examination of the general character of *cognition*. The issue, it turned out, divided into a pair of questions. The first concerned the general characters of cognitive *behaviours* (i.e. the domain to be explained); the second, the general character of cognitive *processes* (i.e., the underlying processes that explain the phenomenal domain). It was argued that here, as in other sciences, the explananda and the explanans are interdependent. That is, that explanatory theory influences the domain of phenomena to be explained and vice versa. It was argued, therefore, that it would be unwise to place too much emphasis on either of these questions at the expense of the other.

Regarding the class of cognitive *behaviours* it was argued that many attempts to address the issue are largely uninformative: they tend to rely on translation into near synonyms equally in need of analysis, merely list a set of paradigmatic examples, or appeal to a very diverse body of disciplines nominally collected under the banner of cognitive science. The strongest answer to the question of the phenomenal domain focussed on the *plasticity* and *flexibility* of behaviour which were then examined. It was argued that although these concepts do not provide necessary or sufficient conditions for delimiting the domain, they nevertheless do provide useful markers.

Regarding cognitive *processes* it was argued that specifically computationalist theories were in danger of letting theoretical commitments dictate the domain of cognitive behaviour, and that a broader perspective would be more prudent. It was argued that such a broader perspective could be found in the characterisation of cognitive processing as involving *information* processing. The concept of information, it was argued, does double duty for a pair of concepts – what I describe as the *truth-neutral* and *truth-dependent* senses of the term. Each of these play a significant (but distinct) role in cognitive explanation. Truth-neutral information – a synonym for representational content – plays an important role in connecting similar cognitive behaviours across physically diverse stimuli. This carries with it a theoretical commitment to the role of representations in cognitive processing. Truth-dependent information plays an important role in shaping our behaviour to so *successfully* or *appropriately* fit the world by reliably reflecting how things stand. Certain objections to the information processing approach were considered and it was suggested that this largely grew out of the lack of an agreed upon understanding of what a cognitive *representation* is.

§7.1.3 Chapter 3 – Cognitive Representation

Having arrived at a qualified and preliminary endorsement of a standard line that cognitive processing involves information processing – i.e., the processing of *representations* – the general nature of representation as it features in cognition was therefore examined in greater detail. This was the central concern of chapter 3 – to come to a clearer understanding about what it is for something to be the kind of thing that can play a representational role in a cognitive process. In other words, the central concern was with elucidating and defending a description of cognitive representation in a broad enough way to cut across various partisan divides, especially those between anti-representational and representational approaches.

A cognitive representational state or structure, I agued, is something that:

- is *about* something, in something like the standard intentional sense.
- *stands in for* something, but in a way that it plays some *informational* role.
- represents through and in virtue of its *properties*.
- is, under an interpretation, *isomorphic* in some determinate way or other, with what it represents or whatever it has the function to represent.
- requires an *interpretation* to determine the isomorphism from amongst an indeterminate number of possible candidates.
- is (thus) capable is *misrepresenting* what it represents.

With this characterisation of cognitive representation in place, it was argued that i) anti-representationalist arguments do not demonstrate a need or even a desirability to dispense with explanations that appeal to cognitive representations, and ii) there remains a clear need for an explanatory appeal to cognitive representations in order to account for significant swathes of cognitive behaviour, such as conceptual thinking and 'offline' reasoning. It was concluded that whilst one cannot dismiss the suggestion that some (perhaps even many) cognitive behaviours do not require explanatory appeal to representations, overly strong anti-representationalist arguments to the effect that *all* cognition requires no appeal to representation can be rejected.

§7.1.4 Chapter 4 – Para-information and Embedded Cognition

In chapter 4 we began our journey out into trans-cranial and trans-corporeal cognitive world. The primary focus of this chapter was an examination of a pair of closely related concepts: what I called *para-information* and *embedded* cognition. Certain cognitive behaviours, I argued, require an explanatory appeal to features of the world – para-informational states or structures – that enable the cognitive agent to extract task-relevant information. I defended the claim that cognition is often deeply *embedded* in a cognitive agent's environment, meaning that it is often dependent upon and shaped by external para-information. In making such an explanatory appeal, I suggested, this does not imply that we must view such features of the environment as being *cognitive*, however. The claim that much human cognition is embedded is, therefore, a weaker claim than that much of it is extended.

The relationship between embedded cognition and para-information was then crystallised through a brief analysis of the nature of dispositions, and hence (qua dispositions), of embedded cognitive *capacities*. Dispositions, it was argued, supervene upon the salient properties of the systems to which they are notionally attributed *together with* those of the systemic elements of their activating conditions. By extension, embedded cognitive capacities should be seen as supervening upon cognitive agents considered not in isolation from, but *together with*, cognitively significant features of their environments.

§7.1.5 Chapter 5 – Extended Cognition

In chapter 5 I laid out and clarified – first in a rough way and later with greater precision – the extended cognitivist thesis: that certain processes occurring beyond the boundaries or a cognitive agent's body, are as much parts of that agent's cognitive processing as anything going on in their brain or within the bounds of their body. The focus was on an elucidation and defence of the *general* thesis of extended cognitivism – one that is silent on the underlying character of extended cognitive processing. This involved both distinguishing it from a number of positions with which it is more or less closely associated (and with which it is sometimes confused), as well as defining a number of general claims concerning cognitive processes, states, and systems that have lain implicit in the dissertation up to this point. After examining some key putative examples of extended cognition (as well as noting some important points of agreement between critics and protagonists), I then examined and rejected four of the central objections ranged against extended cognitivism.

The first of these objections – the 'coupling-constitution fallacy' – was found to be based on a misunderstanding of the central motivation for the extended cognitivist position. Notwithstanding that, further faults were found primarily in a failure to focus on the relevant behaviour and on an unprincipled demand that the sub-cognitive processes of a cognitive process be independently cognitive.

The next objection focussed on the alleged significance of the distinction between intrinsic and derived intentionality: only states or process with intrinsic content, critics have alleged, can count as cognitive. Three strategies were proposed for dealing with that objection. The first strategy – Dennett's denial of the reality of the distinction – was outlined but passed over in favour of less radical alternatives. It was argued that once we focus on the appropriate level of functional decomposition for a given cognitive process, the challenge is defused either by a satisfaction of the demand for intrinsic content (the second strategy) or, once again, by the objection making an unprincipled demand that the sub-cognitive processes of a cognitive process be independently cognitive (the third strategy).

The third objection questioned the scientific credentials of extended cognitivism by suggesting that it is incapable of delivering causally individuated explanations and that only causally individuated explanations are appropriate. This objection was criticised on two main grounds. The first ground was that the empirical evidence does not obviously support a claim of the strength required to worry the extended cognitivist. The second ground was that cognitive explanation does not need to be fundamentally causal. It was argued that explanatory unity across causally diverse phenomena may be found at a functional level.

The final objection considered in chapter 5 emphasised the importance of the locus of control for a cognitive process as something residing within the head of the cognitive agent. This was rejected as being neither a necessary aspect of cognition, nor, at least in any direct or ongoing way, an essential feature of agency more generally.

§7.1.6 Chapter 6 – Extended Intentional States

In chapter 6 I examined and defended a stronger version of extended cognitivism that includes a commitment to specifically *representational* extended cognitive processes and states. I began with an examination of the important role of *cognitive technology*, broadly understood, in human cognition. This was followed by defence of a representationalist interpretation of the central examples considered in the preceding chapter: pen and paper arithmetical calculation, Clark and Chalmers' Tetris example, and Scrabble. In the next section I characterised extended cognitivism in respect of external belief states and intentional explanation – a central focus of the chapter. The central example, drawn from Clark and Chalmers (1998), was that of Otto, a sufferer of a mild form of Alzheimer's disease, and his notebook. Following Clark and Chalmers, I argued that the role that Otto's notebook plays in his cognitive economy qualifies it as a repository of his beliefs. I then considered and ultimately rejected three objections raised against this claim.

The central thrust of the first objection – the cognitive bloat objection – is that extended cognitivism implies the extension of belief states to counter-intuitive cases. Problems were identified with Clark and Chalmers original handling of that objection and so rather than resisting the extension to counter-intuitive cases I argued in favour of embracing that extension. I supported this position in part by emphasising the general immunity of naturalistic accounts to criticism from intuition or ordinary language usage, and in part by arguing for a causal-functional understanding of the relation of believing.

The second objection – the Otto 2-step objection – suggested that the appeals to external states or processes lack any explanatory purchase; all the explanatory weight is carried by internal processes and states. This was rejected by distinguishing between the explanatory roles of occurrent versus non-occurrent beliefs. It was argued that Otto's notebook contains only his non-occurrent beliefs, whose explanatory role lies at one remove from those of his occurrent beliefs. This point was drawn out in a later section of the chapter through the provision of a set of conditions for non-occurrent belief attribution.

The third objection argued that the role of perception provides a principled 'barrier' between cognitive processes and states occurring within the head and other (putatively non-cognitive) processes or states occurring outside of the head. Analogies between notebooks and brains, critics suggest, founder on the inevitable intercession of perceptual processes in the former but not the latter. This was rejected primarily on the grounds that differences at a fine-grained level of analysis offer no grounds to reject functional parity at a broader and more appropriate level of functional analysis.

§7.1.7 Chapter 7 – Conclusion

In this chapter I summarise the arguments of the previous chapters and explore some avenues for future research.

§7.2 – SOME CLOSING SPECULATIONS

During the course of the dissertation several issues of less central importance to defence of the thesis were treated relatively briefly or not at all. Many of these deserve a more expansive and detailed examination in their own right, and indeed, many extended cognitivists have explored some of them. In this section I shall offer a few speculations and present some summary positions concerning some of these matters.

One very general claim that extends well beyond the philosophy of mind or cognition and into broader metaphysical issues concerns the analysis of dispositions offered in chapter 4. It was argued there that dispositions are best seen as supervening on the properties of what is notionally the dispositional system taken together with the systemic elements of the activating conditions. So far as my relatively limited knowledge of the literature on dispositions can discern, this offers a relatively novel kind of dispositional analysis that has not been properly explored.

Following very much on the heels of that account of dispositionality is, I suggest, a nascent theory of representational content. As noted in the introduction, the naturalisation of representational content does not directly concern the central thesis, but it would be an impressive asset to the account given in the preceding chapters should it be found to contribute towards a viable theory of representational content. Earlier versions of this dissertation contained some considerable speculation along precisely those very lines, and whilst this is not the place to rehearse such arguments in their entirety, the barest outlines can be given.

States or structures have determinate content (a meaning) in virtue of a disposition notionally attributable to cognitive agents, but properly supervening on properties of the content-bearer taken together with an 'environment' which provides an interpretation for the content-bearer.[143] Words, sentences, and our own internal thoughts acquire a determinate content in essentially this way. Either the 'interpretive environment (in part, the local geography in the case of Alice examined in chapter 4) or the content-bearer (the contents of Otto's notebook in chapter 6) may be external to the traditional boundaries of the cognitive agent. For example, what determines the meaning of the written sentence 'the cat is on the mat' is that there is a disposition notionally attributable to certain cognitive agents to interpret such sentences in a particular way. Properly speaking, however, such dispositions supervene on the properties of the cognitive agent together with those of their environment, and especially those of the (in this case external) content-

[143] The scare quotes around 'environment' are intended to indicate that the term needs to be very broadly construed, for it may well be internal to the traditional boundaries of the cognitive agent as occurs in purely in-the-head cognition.

bearer. Fairly clearly, for example, monoglot Hungarians have different dispositions towards the above quoted sentence and do not interpret it as saying that the cat is on the mat, if they interpret it as saying anything at all. Equally, however, properties of the content-bearer to which the cognitive agent is sensitive make a difference to the meaning. Firstly, there is the syntactic structure of the sentence. 'The cat is on the mat' does not have the same meaning as 'the mat is on the cat' in part, because of the ordering of its parts. Secondly, there are the letters of the words themselves. The word 'mat' does not have the same meaning as the word 'cat' in part because they start with different letters. Both of these properties make an obvious and direct contribution to determining the meaning of the written sentence.

There are, of course, difficulties with this (as with any extant) account of representational content. One prominent set of difficulties facing dispositional accounts of content are presented in Kripke (1982). But questions can be raised against the specific notion of dispositionality upon which such objections depend (see, for example, Martin and Heil, 1998). Another objection concerns the fear that an infinite regress of interpretations beckons.[144] In its usual form the regress argument is this: if the content of a representational state or structure always requires an interpretation, then all representations require an interpreter external to the representational state or structure. But such an interpreter will presumably have internal mental states that, if they are indeed representational, must themselves be in need of interpretation, and so on. This, I suggest, can be substantially addressed by noting that 'interpretive environment' is specified ultimately in relation to a content-bearer not a cognitive agent. As noted, the 'interpretive

[144] Versions of this concern can be found, for example, in Dennett (1978), p. 122, or Millikan (1984), pp. 89-90.

environment' (as well as the content-bearer) might be entirely internal to the cognitive agent. Thus, no additional *interpreter* external to the representation is required.

Another metaphysical issue to which extended cognitivism is highly relevant concerns the extension of the *self* into the environment beyond the skin. Andy Clark (2003) has explored this issue with considerable imagination, and argues that we should consider ourselves 'natural-born cyborgs'. Unfortunately, he provides (so far as I can discern) no rigorous analysis of 'self', and examines the notion in rather impressionistic terms, and so his arguments for an extended self are less easy to assess than his arguments for extended cognition. As indicated in the introduction, the development of a sufficiently substantive conception of the self, required to rigorously sustain such a claim, lay beyond the necessities of the dissertation. Cognition, as I have examined it, can be viewed as a 'self-less' commodity given any relatively rich understanding of the term 'self'. However, insofar as agency is central to the self, then one can point out (§5.6) that agency generally extends quite naturally beyond the bounds of one's skin. Most obviously, insofar as one's cognition is central to the self, then the arguments for an extended self are already made. However, such claims involve conceptions of the self that are a little too thin to satisfy many, and perhaps with good reason. Control or agency seem important, but neither of these are obviously essential for cognition; and what role has consciousness in our self-conception? Granted, we may be much more than our consciousness, but it is at least highly questionable whether we can be a 'self' without consciousness being involved in some respect or other.

There are a number of issues that were avoided in the dissertation primarily in order to avoid certain complications or distractions that might arise from them. One such issue concern that 'group-mind hypothesis' advocated by the likes of Hutchins (1995) and

Wilson (2001 and 2004) and discussed briefly in chapter 5. I am very sympathetic to this position. However, defending it adequately would, in my opinion, require addressing a complex of issues concerning the ontological and explanatory status of collective entities such as cultures and societies. Moreover, it would require a considerably more concerted effort to justify non-reductive naturalism than seemed appropriate or possible within the confines of this dissertation. (Wilson [2004] undertakes essentially these tasks and, in my opinion, does an excellent job.) But the case for the extension of cognitive processing and states into the world can be made most persuasive, I believe, when restricted to the cognition of an individual cognitive agent. Individuals are, after all, already considered cognitive agents in their own right, whereas collections of individuals, a critic might argue, are cognitive in sense entirely derivative of individual cognition.

There is an extension of the thesis defended here that is considerably more immediate than to the group-mind hypothesis. I have focussed my defence on examples exclusively involving the extension of cognitive processes and states into *inanimate* external cognitive resources such as notebooks, pocket calculators, and the like. But it will hardly have escaped attention that this might be extended to include *animate* external resources, especially other cognitive agents. For example, Clark and Chalmers (1998) suggest that in the case of an unusually interdependent couple, the states of the one might play the same kind of role as the notebook does for Otto. Most of us will, from time to time, *consult* and *trust* the advice of experts, frequently acting directly on their pronouncements. That is, we often stand in relations to the pronouncements of other cognitive agents in much the same way that we do to the contents or output of cognitive technology. Such matters bring us squarely into the realms of social epistemology. Extended cognitivism does nothing if not add increased weight to the importance of the domain of social epistemology, and the prospects of mutually beneficial contribution seem likely. At the same time, of course, by extending the boundaries of the individual cognitive agent, it might be seen as simultaneously bringing into question the significance of standard ways of distinguishing social epistemology from individual epistemology.

These and many other matters stemming from this dissertation I leave for another time and for future research.

BIBLIOGRAPHY OF WORKS CITED AND CONSULTED

Abelson, R.P. (1986). Beliefs Are Like Possessions. *Journal for the Theory of Social Behaviour*, *16* (3), 223-250.

Adams, F. and Aizawa, K. (2001). The Bounds of Cognition. *Philosophical Psychology*, *14* (1), 43-64.

Adams, F. and Aizawa, K. (forthcoming a). Defending the Bounds of Cognition. In R. Menary (Ed.), *The Extended Mind.* Aldershot, Hants: Ashgate Publishing Ltd.

Adams, F. and Aizawa, K. (forthcoming b). Why the Mind is Still in the Head. *Cambridge Handbook on Situated Cognition.* Eds. P. Robbins and M. Aydede. April 15, 2005 draft retrieved from http://personal.centenary.edu/~kaizawa/cv.htm as 'Challenges to Active Externalism'.

Adams, F. and Aizawa, K. (forthcoming c). Defending Non-Derived Content. *Philosophical Psychology*. Retrieved from http://personal.centenary.edu/~kaizawa/cv.htm.

Aizawa, K. (draft). Clark's Conditions on Extended Cognition are Too Strong. April 15, 2005 draft, retrieved from http://personal.centenary.edu/~kaizawa/cv.htm.

Amadae, S.M. (2003). *Rationalizing Capitalist Democracy : The Cold War Origins of Rational Choice Liberalism*. Chicago: University of Chicago Press.

Anderson, J. R. (1980). *Cognitive Psychology and its Implications*. San Francisco: W. H. Freeman and Company.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. Cambridge, Mass.: MIT.

Bechtel, W. (1997). Embodied Connectionism. In D.M. Johnson and C. Erneling (Eds.), *The Future of the Cognitive Revolution* (pp. 187-208). Oxford: OUP.

Bechtel, W. (1998). Representations and Cognitive Explanations: Assessing the Dynamicist's Challenge. *Cognitive Science*, *22* (3), 295-318.

Bechtel, W. (1999). The Case for Connectionism. In W.Lycan (Ed.), *Mind and Cognition* 2nd ed. (pp. 153-169). Oxford: Blackwell.

Beer, R.D. (1995). A Dynamical Systems Perspective on Agent-Environment Interaction. *Artificial Intelligence*, *72*, 173-215.

Beer, R. D. (2000). Dynamical Approaches to Cognitive Science. *Trends in Cognitive Sciences, 4*, 91-99.

Bigelow, J. and Pargetter, R. (1987). Functions. *Journal of Philosophy, 84*, 181-196.

Block, N. (1999). Troubles with Functionalism. In W.Lycan (Ed.), *Mind and Cognition* (2nd ed.) (pp. 435-439). Oxford: Blackwell.

Boden, M. (Ed.) (1996). *The Philosophy of Artificial Life*, Oxford: OUP.

Brand, M. and Harnish, R.M. (Eds.) (1986). *The Representation of Knowledge and Belief.* Tucson, AZ.: University of Arizona Press.

Brandom, R.B. (1994). *Making It Explicit*. Cambridge, Mass.: Harvard.

Brooks, R.A. (1986). A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation, 2*, 14-23.

Brooks, R.A. (1991a). Intelligence Without Representation. *Artificial Intelligence, 47*, 139-159.

Brooks, R. (1991b). New Approaches to Robotics. *Science, 253*, 1227-1232.

Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy, 4*, 73-121.

Butler, K. (1998). *Internal Affairs*. Dordrecht: Kluwer Academic.

Chemero, A. (2000). Anti-Representationalism and the Dynamical Stance. *Philosophy of Science, 67* (December), 625-647.

Cherry, E.C. (1951). A History of the Theory of Information. *Proceedings of the Institute of Electrical Engineers, 98* (III), pp. 383-393.

Chomsky, N. (1965). *Aspects of a Theory of Syntax*. Cambridge, Mass.: MIT.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia.

Churchland, P. S., Ramachandran, V. S., and Sjenowski, T. J. (1994). A Critique of Pure Vision. In C. Koch and J. L. Davis (Eds.), *Large-scale Neuronal Theories of the Brain* (pp. 23-60). Cambridge, MA: MIT.

Clark, A. (1989). *Microcognition*. Cambridge, Mass.: MIT.

Clark, A. (1997a). *Being There*. Cambridge, Mass.: MIT.

Clark, A. (1997b). The Dynamical Challenge. *Cognitive Science, 21* (4), 461-481.

Clark, A. (1998). Embodied, Situated, and Distributed Cognition. In W. Bechtel and G. Graham, (Eds.) *A Companion to Cognitive Science* (pp. 506-517). Oxford: Blackwell.

Clark, A. (2001a). Reasons, Robots and the Extended Mind. *Mind and Language*, *16* (2), 121-145.

Clark, A. (2001b). *Mindware: An Introduction to the Philosophy of Cognitive Science.* Oxford: OUP.

Clark, A. (2003). *Natural Born Cyborgs*. Oxford: OUP.

Clark, A. (forthcoming). Memento's Revenge: The Extended Mind Extended. To appear in R. Menary (Ed.), *Papers on the Extended Mind.* Aldershot: Ashgate Press. Draft retrieved from http://www.cogs.indiana.edu/andy/Mementosrevenge2.pdf).

Clark, A. and Chalmers, D. (1998). The Extended Mind. *Analysis*, *58* (1), 7-19.

Clark, A. and Toribio, J. (1994). Doing Without Representing? *Synthese*, *101*, 401-431.

Cole, M. and Engeström, Y. (1993). A Cultural Historical Approach to Distributed Cognition. G. Salomon (Ed.), *Distributed Cognitions: Psychological and Educational Consideration* (pp. 1-46). Cambridge: Cambridge University Press.

Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, Mass.: Harvard.

Crane, T. (2003). *The Mechanical Mind.* (2^{nd} ed.). London: Routledge.

Cummins, R. (1975). Functional Analysis. *Journal of Philosophy*, *72*, 741-765.

Cummins, R. (1986). Inexplicit Information. In Brand and Harnish (Eds.), *The Representation of Knowledge and Belief* (pp. 116-126). Tucson, AZ.: University of Arizona Press.

Cummins, R. (1989). *Meaning and Mental Representation.* Cambridge, Mass.: MIT.

Cummins, R. (1996). *Representations, Targets, and Attitudes*. Cambridge, Mass.: MIT.

Dartnall, T. (1996). Retelling the Representational Story. *Communication and Cognition*, *29* (3/4), 479-500.

Davies, P.S. (2001). *Norms of Nature*. Cambridge, Mass.: MIT.

Dawkins, R. (1982). *The Extended Phenotype*. Oxford: OUP.

Dawson, M.R.W. (1998). *Understanding Cognitive Science.* Oxford: Blackwell.

Dennett, D.C. (1978). *Brainstorms.* London: Penguin.

Dennett, D.C. (1983). Styles of Mental Representation. *Proceedings of the Aristotelian Society, 83*, 213-226.

Dennett, D.C. (1987). *The Intentional Stance.* Cambridge, Mass.: MIT.

Dennett, D.C. (1990). The Myth of Original Intentionality. In In K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale and K.V. Wilkes (Eds.), *Modelling the Mind* (pp. 43-62). Oxford: Clarendon Press.

Dennett, D.C. (1995). *Darwin's Dangerous Idea.* London: Penguin.

Dennett, D.C. (1996). *Kinds of Minds.* London: Orion Books.

Dennett, D.C. (1998). *Brainchildren.* London: Penguin.

Dipert, R.R. (1995). Some Issues in the Theory of Artifacts: Defining 'Artifact' and Related Notions. *Monist, 78* (2), 119-135.

Dobbyn, C. and Stuart, S. (2003). The Self as Embedded Agent. *Minds and Machines, 13* (2), 187-201.

Donald, M. (1991). *Origins of the Modern Mind.* Cambridge, Mass: Harvard.

Dretske, F. (1981). *Knowledge and the Flow of Information.* Cambridge, Mass.: MIT.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, Content and Function* (pp. 17-36). Oxford: OUP.

Dretske, F. (1988). *Explaining Behaviour.* Cambridge, Mass.: MIT.

Dretske, F. (1990). Reply to Reviewers. *Philosophy and Phenomenological Research, 50* (4), 819-839.

Dretske, F. (1994). If You Can't Make One, You Don't Know How It Works. *Midwest Studies in Philosophy 19*, 468-462.

Dretske, F. (1995). *Naturalizing the Mind.* Cambridge, Mass.: MIT.

Dreyfuss, H. (1972). *What Computers Can't Do: A Critique of Artificial Reason.* Cambridge, Mass.: MIT.

Eliasmith, C. (1996). The Third Contender: A Critical Examination of the Dynamicist Theory of Cognition. *Philosophical Psychology, 9* (4), 441-463.

Eliasmith, C. (1997). Computation and Dynamical Models of Mind. *Minds and Machines, 7*, 531-541.

Eysenck, M.W. (1993). *Principles of Cognitive Psychology.* Hove, U.K.: Lawrence Erlbaum Associates.

Finke, R. and Pinker, S. (1982). Spontaneous Imagery Scanning in Mental Extrapolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 142-7.

Finke, R. and Pinker, S. (1983). Directional Scanning of Remembered Visual Patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 398-410.

Fodor, J. (1968). The Appeal to Tacit Knowledge in Psychological Explanation. *Journal of Philosophy, 65*, 627-40.

Fodor, J. (1975). *The Language of Thought.* Cambridge, Mass.: Harvard.

Fodor, J. (1980). Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioral and Brain Sciences, 3*, 63-109.

Fodor, J. (1983). *The Modularity of Mind.* Cambridge, Mass.: MIT.

Fodor, J. (1986). Information and Association. In Brand and Harnish (Eds.), *The Representation of Knowledge and Belief* (pp. 80-100). Tucson, AZ.: University of Arizona Press.

Fodor, J. (1987). *Psychosemantics.* Cambridge, Mass.: MIT.

Fodor, J. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition, 28*, 3-71.

Fodor, J. (1990a). *A Theory of Content and Other Essays.* Cambridge, Mass.: MIT.

Fodor, J. (1990b). Information and Representation. In P. Hanson (Ed.), *Information, Language, and Cognition* (pp. 175-190). Vancouver: UBC.

Fodor, J. (1994). *The Elm and the Expert.* Cambridge, Mass.: MIT.

Frege, G. (1884/1959). *Die Grundlagen der Arithmetik.* Breslau: Wilhelm Koebner. Translated by J.L.Austin as *The Foundations of Arithmetic* (revised ed.) Oxford: Balckwell.

Gallistel, C.R. (1998). Symbolic Processes in the Brain: The Case of Insect Navigation. In D. Scarborough and S. Sternberg (Eds.), *An Invitation to Cognitive Science Vol.4.: Methods, Models and Conceptual Issues* (2nd ed.) (pp. 1-51). Cambridge, Mass.: MIT.

Gardner, H. (1985). *The Mind's New Science.* New York: Basic Books.

Godfrey-Smith, P. (1992). Indication and Adaptation. *Synthese, 92*, 283-312.

Gould, S.J. and Vrba, E.S. (1982). Exaptation – A Missing Term in the Science of Form. *Paleobiology*, *8* (1), 4-15.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton, Mifflin.

Gorayska, B., and Mey, J.L. (Eds.). (1996). *Cognitive Technology : In Search of a Humane Interface*. New York: Elsevier.

Grice, P. (1957/1989). Meaning. In *Studies in the Way of Words* (pp. 213-223). Cambridge, Mass.: Harvard.

Grush, R. (1997). The Architecture of Representation. *Philosophical Psychology*, *10* (1), 5-23.

Gundersen, L. (2002). In Defence of the Conditional Account Of Dispositions. *Synthese*, *130*, 389–411.

Hanson, P. (Ed.). (1990). *Information, Language, and Cognition*. Vancouver: UBC.

Harnad, S.(1990). The Symbol Grounding Problem. *Physica D*, 42, 335-346.

Harnish, R.M. (2002). *Minds, Brains, Computers: An Historical Introduction to the Foundations of Cognitive Science*. Oxford: Basil Blackwell.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: Bradford/MIT.

Haugeland, J. (1995). Mind Embodied and Embedded. In *Having Thought* (pp. 207-237). Cambridge, Mass.: Harvard.

Haugeland, J. (1997). (Ed.). *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. Cambridge, Mass.: MIT.

Haugeland, J. (1998). *Having Thought*. Cambridge, Mass.: Harvard.

Heidegger, M. (1926/1962). *Being and Time* (trans. J. Macquarrie and E.Robinson). Oxford: Blackwell.

Heil, J. (1981). Does Cognitive Psychology Rest on a Mistake?. *Mind*, *90*, 321-342.

Houghton, D. (1997). Mental Content and External Representations. *Philosophical Quarterly*, *47*, 159-177.

Hume, D. (1978). *A Treatise of Human Nature*. Oxford: Clarendon.

Hurley, S.L. (1998). *Consciousness in Action*. Cambridge, Mass.: Harvard.

Hunt, R.Reed and Ellis, H.C. (1999). *Fundamentals of Cognitive Psychology* (6th ed.) Boston: McGraw-Hill.

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, Mass.: MIT.

Jacobson, A. Jaap (2003). Mental Representations: What Philosophy Leaves Out and Neuroscience Puts In. *Philosophical Psychology*, *16* (2), 189-203.

Johnson, D. Martel and Erneling, C.E. (1997). *The Future of the Cognitive Revolution.* Oxford: OUP.

Johnson-Laird, P. (1993). *The Computer and the Mind.* London: Fontana.

Johnston, M. (1992). How to Speak of the Colors. *Philosophical Studies 68*, 221–263.

Kelso, J.A.Scott. (1995). *Dynamic Patterns*. Cambridge, Mass.: MIT.

Kirsh, D. (1990). When is Information Explicitly Represented?. In P. Hanson (Ed.), *Information, Language, and Cognition* (pp. 340-365). Vancouver: UBC.

Kirsh, D. (1991). Today the Earwig, Tomorrow Man?. *Artificial Intelligence*, *47*, 161-184.

Kirsh, D. (1995). The Intelligent Use of Space. *Artificial Intelligence*, *73*, 31-68.

Kirsh, D. and Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, *18*, 513-549.

Kosslyn, S. (1980). *Image and Mind.* Cambridge, Mass.: Harvard.

Kosslyn, S. (1994). *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, Mass.: MIT.

Kosslyn, S., Ball, T., and Reiser, B. (1978). Visual Images Preserve Metric Spatial Information. *Journal of Experimental Psychology: Perception and Performance*, *8*, 142-7.

Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

Lakoff, G. and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lewis, D. (1969). *Conventions*. Cambridge, Mass.: Harvard.

Lewis, D. (1997). Finkish Dispositions. *The Philosophical Quarterly*, *47*, 143–158.

Marr, D. (1982). *Vision*. San Francisco: W.H.Freeman and Company.

Martin, C. (1994). Dispositions and Conditionals, *The Philosophical Quarterly*, *44*, 1-8.

Martin, C. and Heil, J. (1998). Rules and Powers. *Philosophical Perspectives, 12*, 283-312.

Matthen, M. (1988). Biological Functions and Perceptual Content. *Journal of Philosophy, 85*, 5-27.

Matthen, M. (2004). Features, Places, and Things: Reflections on Austen Clark's Theory of Sentience, *Philosophical Psychology, 17* (4), 497-518.

Matthen, M. (2005). *Sensing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford: OUP.

McClelland, J. and Rumelhart, D., et al. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2. Cambridge, Mass.: MIT.

Merleau-Ponty, M. (1942/1963). *The Structure of Behaviour*. (trans. A.L.Fisher). Boston: Beacon Press.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT.

Millikan, R. (1989). Biosemantics. *Journal of Philosophy, 86* (6), 281-297.

Millikan, R. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass.: MIT.

Millikan, R. (1998). Language Conventions Made Simple. *Journal of Philosophy, 95* (4), 161-180.

Millikan, R. (1999). Wings, Spoons, Pills, and Quills: A Pluralist Theory of Function. *Journal of Philosophy, 96* (4), 215-254.

Millikan, R. (2000). *On Clear and Confused Ideas*. Cambridge, UK.: CUP.

Millikan, R. (2001). The Language-Thought Partnership: A Bird's Eye View. *Language and Communication, 21*, 157-166.

Mumford, S. (1998). *Dispositions*. Oxford: OUP.

Nagel, E. (1961). *The Structure of Science*. New York : Harcourt, Brace and World.

Neisser, U. (1976). *Cognition and Reality*. San Francisco: W.H. Freeman and Co.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, Mass.: Harvard.

Northcutt, W. (2000). *The Darwin Awards*. New York: Penguin Putnam.

Papineau, D. (1987). *Representation and Reality*. Oxford: Blackwell.

Pfeifer, R., and Scheier, C. (1999). *Understanding intelligence*. Cambridge, Mass.: MIT.

Port, R.F. and Van Gelder, T. (Eds.). (1995). *Mind as Motion*. Cambridge, Mass.: MIT.

Preston, B. (1998a). Why is a Wing Like a Spoon? A Pluralist Theory of Function. *Journal of Philosophy, 95* (5), 215-254.

Preston, B. (1998b). Cognition and Tool Use. *Mind and Language, 13* (4), 513-547.

Putnam, H. (1975). *Mind, Language and Reality: Philosophical Papers Volume 2*. Cambridge: Cambridge University Press.

Putnam, H. (1988). *Representation and Reality*. Cambridge, Mass.: MIT.

Putnam, H. (1994). Reductionism and the Nature of Psychology. In *Words and Life* (pp. 428-440). Cambridge, Mass.: Harvard.

Pylyshyn, Z. (1980). Computation and Cognition: Issues in the Foundations of Cognitive Science. *Behavioural and Brain Sciences, 3*, 111-169.

Pylyshyn, Z. (1983). Information Science: Its Roots and Relations as Viewed from the Perspective of Cognitive Science. In F. Machlup and U. Mansfield (Eds.), *The Study of Information: Interdisciplinary Messages*. New York: Wiley.

Pylyshyn, Z. (1991). The Role of Cognitive Architecture in Theories of Cognition. In K.VanLehn (Ed.) *Architectures for Intelligence*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Ramsey, W.M. (1997). Do Connectionist Representations Earn Their Explanatory Keep? *Mind and Language, 12*, 34-66.

Ramsey, W.M., Stich, S.P. and Rumelhart, D.E. (Eds.). (1991). *Philosophy and Connectionist Theory*. Hillsdale, N.J.: L. Erlbaum Associates.

Rowlands, M. (1999). *Body in Mind*. Cambridge, England: CUP.

Rowlands, M. (2002). Two Dogmas of Consciousness. *Journal of Consciousness Studies, 9* (5-6), 158-180.

Rupert, R. (2004). Challenges to the Hypothesis of Extended Cognition. *Journal of Philosophy, 101* (8), 389-428.

Rumelhart, D. and Smolensky, P., et al (1986). Schemata and Sequential Thought Processes in PDP models. In J. McClelland, *et al., Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (*Vol. 2*) (pp. 7-58.). Cambridge, Mass.: MIT.

Ryle, G. (1949). *The Concept of Mind*. London: Penguin.

Scarborough, D. and Sternberg, S. (1998). *An Invitation to Cognitive Science Vol.4.: Methods, Models and Conceptual Issues* (2nd ed.). Cambridge, Mass.: MIT.

Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences, 3*, 417-424.

Searle, J. (1983). *Intentionality.* Cambridge, England: CUP.

Searle, J. (1984). Intentionality and Its Place in Nature. *Dialectica, 38* (2-3), 87-99.

Searle, J. (1992). *The Rediscovery of the Mind.* Cambridge, Mass.: MIT.

Searle, J. (1997). *The Mystery of Consciousness.* London: Granta.

Sellars, W. (1956/97). *Empiricism and the Philosophy of Mind*, edited by R. Brandom. Cambridge, Mass.: Harvard.

Shannon, C., and Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Silveira, J. (1971). *Incubation: The Effect of Interruption Timing and Length on Problem Solution and Quality of Problem Processing.* Unpublished Ph.D. thesis, University of Oregan. Discussed in H. Eysenck *Principles of Cognitive Psychology* (pp. 133-134). Hove, U.K.: Lawrence Erlbaum Associates.

Soames, S. (1998). Facts, Truth Conditions, and the Skeptical Solution to the Rule-Following Paradox. *Philosophical Perspectives, 12*, 313-348.

Stalnaker, R. (1984). *Inquiry.* Cambridge, Mass.: MIT/Bradford.

Stampe, D.W. (1977). Toward a Causal Theory of Linguistic Representation. In P. French, et.al. (Eds.) *Midwest Studies in Philosophy, 2*, Minneapolis: University of Minnesota Press.

Stephens, C. (2001). When is it Selectively Advantageous to Have True Beliefs? Sandwiching the Better Safe Than Sorry Argument. *Philosophical Studies, 105*, 161-189.

Sterelny, K. (1990). *The Representational Theory of Mind.* Oxford: Blackwell.

Sterelny, K. (forthcoming). Externalism, Epistemic Artefacts and the Extended Mind. R. Schantz (Ed.), *The Externalist Challenge: New Studies on Cognition and Intentionality.* New York: de Gruyter.

Sternberg, R.J. (1987). Intelligence. In R. Gregory (Ed.) *The Oxford Companion to the Mind* (pp. 375-381). Oxford: OUP.

Stich, S. (1978). Autonomous Psychology and the Belief-Desire Thesis. *Monist*, *61*, 573-591.

Stich, S. (1990). Building Belief: Some Queries about Representation, Indication, and Function. *Philosophy and Phenomenological Research*, *50* (4), 801-806.

Stillings, N., Weisler, S.W., Chase, H., Feinstein, J., Garfield, J. and Rissland, E. (1987). *Cognitive Science: An Introduction*. Cambridge, Mass.: MIT.

Symons, J. (2001). Explanation, Representation and the Dynamical Hypothesis. *Minds and Machines*, *11*, 521-541.

Thelen, E., and Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Mass.: MIT.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard.

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind*, *59*, 433-460.

Van Gelder, T. (1991). What is the 'D' in 'PDP'? A Survey of the Concept of Distribution. In W. M. Ramsey, S. P. Stich and D. E. Rumelhart (Eds.), *Philosophy and Connectionist Theory* (pp. 33-59). Hillsdale, N.J.: L. Erlbaum Associates.

Van Gelder, T. (1998). The Dynamics Hypothesis in Cognitive Science. *Behavioral and Brain Sciences*, *21*, 615-665.

VanLehn, K. (1991). *Architectures for Intelligence*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Varela, F.J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind*. Cambridge, Mass.: MIT.

Vygotsky, L. (1978). *Mind in Society*. Cambridge, Mass.: Harvard.

Vygotsky, L. (1986). *Thought and Language*. Cambridge, Mass.: MIT.

Wilson, M. (2002). Six View of Embodied Cognition. *Psychonomic Bulletin and Review*, *9*, 625-636.

Wilson, R. (1994). Wide Computationalism. *Mind*, *103*, 351-372.

Wilson, R. (2001). Group-Level Cognition. *Proceedings of the Philosophy of Science Association*, *68*, S262-S273.

Wilson, R. (2004). *Boundaries of the Mind*. Cambridge, England: Cambridge University Press.

Wheeler, M. (1996). From Robots to Rothko: The Bringing Forth of Worlds. In M. Boden (Ed.), *The Philosophy of Artificial Life* (pp. 209-236). Oxford: OUP.

Wheeler, M. and Clark, A. (1999). Genic Representation: Reconciling Content and Causal Complexity. *British Journal for the Philosophy of Science*, *50*, 103-135.

Wittgenstein, L. (1922/1961). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

Wittgenstein, L. (1953/1968). *Philosophical Investigations*. Oxford: Basil Blackwell.

Wittgenstein, L. (1975). *On Certainty*. Oxford: Basil Blackwell.

Wooldridge, D.E. (1963). *The Machinery of the Brain*. New York: McGraw Hill.

Yaniv, I., Meyer, D.E. and Davidson, N. S. (1995). Dynamic Memory Processes in Retrieving Answers to Questions: Recall Failures, Judgments of Knowing, and Acquisition of Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1509-1521.