Representation Rectified

Jack Wadham

Thesis submitted for the degree of

Doctor of Philosophy

(Philosophy)

June 2014

Department of Philosophy

The University of Sheffield

For Poppy

Thesis Abstract

This abstract is a little involved and technical. Those looking for a more leisurely, thematic overview of the thesis can turn to the Introduction, which follows shortly.

In the first chapter of the thesis, I examine Andy Clark's argument for the extended mind thesis (EM henceforth). As Clark acknowledges, his argument for EM relies on a brand of functionalism developed by Frank Jackson and David Braddon-Mitchell. Mark Sprevak claims to have developed a *reductio* not only of Clark's argument of EM, but of the functionalist position that Clark's argument presupposes. I show that this *reductio* can be blocked, because it rests on an optional presupposition (a presupposition that is anyway implausible). But by rejecting this presupposition, we end up blocking Clark's argument for EM as well as Sprevak's *reductio* of that argument. This result is bad news for Clark, but rather better news for the functionalists on whose work Clark's argument relies.

In chapters 2 and 3, I develop and apply a model-based theory of mental representation. The basic idea is there is a type of mental representation (what I call 's-representation') which is best understood by analogy with scientific models. In chapter 2, I develop a theory of content for s-representations, improving on the work of writers who have attempted to do so in the past. The theory of content I develop makes use of theoretical resources provided by those functionalist writers whose position I defended from Sprevak's *reductio* in chapter 1. In chapter 3, I give reasons for thinking that s-representations are biologically ubiquitous and cognitively significant. I then show how my theory of s-representation differs from similar rival accounts in the literature. Finally, I argue that my account has the resources to deal with certain sceptical challenges raised by anti-representationalists (those sceptical of the claim that a certain class of mental capacities can be explained in representational terms).

In Part II (chapters 4, 5 and 6), I apply the lessons learned from the first half of the thesis to develop some distinctive claims about the nature of visual perception. I do so by using Alva Noë's theory of perception as a spring-board. I argue that many of Noë's most notorious claims are false, but that there are still valuable resources to be gleaned from his theory. I then borrow and redeploy what is valuable in his theory (i.e. certain aspects of his 'virtual content' thesis and some of his claims about perspectival content). I do so, in part, by drawing on the s-representation story developed in Part I. I argue, in line with similar claims made by Rick Grush, that Noë's notion of 'sensorimotor knowledge' can usefully be treated as a form of s-representation. With a fully reconstructed version of Noë's theory in place, I show how it can make sense of some otherwise puzzling findings made by psychologists of perception.

Table of Contents

Thesis Abstract……v

Table of contents……vii

Acknowledgements……..ix

Introduction……xi

PART I……**xv**

Chapter 1 – Functionalism and the Extended Mind ……...**1**

Chapter 2 – A Theory of Content for S-Representation……..**29**

Chapter 3 – The Significance of S-Representations……....**71**

PART II……**95**

Chapter 4 –Perception and Action……..**97**

Chapter 5 – The Virtual Content Thesis……..…**117**

Chapter 6 – The Problem of Invisible Contents……**143**

Afterword……**157**

References……..…**159**

Acknowledgements

I should start by thanking the Arts and Humanities Research Council for funding my studies and making this project possible.

I'd also like to thank my two excellent and patient supervisors for all the support and advice. Thanks go to Rob Hopkins, my primary supervisor, for all the time and effort he devoted to helping me turn jumbled, inchoate ideas into something thesis-like. Without his insightful comments and words of encouragement I doubt I would have got very far. He deserves special thanks for being so generous with his time even after he'd stopped working for the University of Sheffield. And thanks go as well to Dominic Gregory, my secondary supervisor, for finding (and helping me fix) all manner of serious problems in earlier drafts, and for his extremely helpful advice on writing strategy. Thanks also for lending me the brewing tips and paraphernalia without which I might never have produced a more or less drinkable beer.

Thanks also to the many other members of the Sheffield philosophy department who have helped me on my way at various points during my time here. Particular thanks go to George Botterill, Chris Hookway and Jimmy Lenman for encouraging me to continue my philosophy studies to PhD level. Chris Bennett, Rosanna Keefe, Steve Laurence, Eric Olsen, (and no doubt others besides) all helped me out in various ways along the way, so thanks go to them too.

Thanks to Rich Healey, Steve Wright and Carl Fox for proof-reading chapter drafts for me at very short notice. And thanks to Angie Pepper and Jess Begon for answering my incessant questions on all matters thesis-formatting-related.

I am grateful also to the various audience-members to whom I have presented earlier drafts of this work: those at the *Evolution, Intentionality and Information* conference in Bristol, the *Reach of REC* conference in Antwerp, the *Dangerous Liaisons* conference in York, and those at the various Sheffield Grad seminars and summer seminars at which I have presented work. Since my memory is poor and the list of those to whom I owe thanks is long, I won't attempt to name names here.

Thanks to Mum, Gavin, Rohan, Susie and the rest of the family. I couldn't hope for a more supportive family and I thank them for being there in times of need.

All my friends in Sheffield deserve thanks for helping to make my PhD-years the happiest, funniest and most interesting in (my admittedly poor) memory.

And finally, and most importantly, to Poppy: for all the fun, all the laughs, for all the love and support, and generally for putting up with me, thank you.

Introduction

My thesis explores promising moves we might make in explaining certain mental capacities exhibited by humans and other animals. Perception is the capacity on which I will focus primarily. The aim is to show how the mental capacities under investigation can (in principle at least) be viewed as extremely complex physical processes (and nothing more besides). One strategy commonly employed by those aiming to explain the mental in physical terms is to think of the mind as just a very sophisticated computing system. I sketch out an alternative to this strategy.

The project began with my scepticism about computational theories of mind and with an enthusiasm for what is often referred to as the 'enactivist' literature. I was taken with the idea that mental processes like perception might be 'active' or 'embodied' in some theoretically significant and hitherto unappreciated sense, and might be amenable to explanation in non-representational terms. The position I eventually arrived at, and the view that I defend in what follows, is quite different from that with which I started.

I remain sceptical about computational theories of mind – theories according to which the functional architecture of the mind is in interesting ways analogous to that of a conventional computer. And it is the main project of the thesis to sketch out a promising alternative to this picture. According to the computational picture to which I provide an alternative, mental capacities (even quite basic mental capacities) are to be explained in terms of the processing of language-like, symbolic mental representations, in accordance with certain rules of inference. But, unlike many writers whose work I discuss, I want to provide an alternative to this view of the mind, without giving up on the idea that even quite basic mental capacities (like the basic perceptual capacities displayed by most animals) might best be explained, at least in part, by appeal to mental representations.

Mental representations are states of the mind that have the property of 'aboutness'. We can get a little clearer on what 'aboutness' is supposed to amount to by thinking of the sense in which non-mental representations display this property. A linguistic sentence can be about the world, in that when I say 'the cat is on the mat' I am saying something about the cat, and its position relative to the mat.

Similarly, a painting or a photograph of the cat, sat on the mat, ascribes to the cat the property of being on the mat. So all these forms of representation are capable of ascribing properties to objects. Another form of non-mental representation is a scientific model. To take an example that will feature prominently in what follows, an aeroplane tail in a wind-tunnel can tell us about certain states of affairs – it can tell us about how a similarly structured aeroplane tail would behave, while flying through the air, attached to an aeroplane. On the position I defend, we should think about mental representations primarily as model-like representations rather than as language-like. Borrowing a term from Ramsey (2007), I call model-like representations 's-representations'.

It should be noted that one could think of mental representations as both model-like and language-like. After all, one could build a model of the world out of lots of language-like symbols. Indeed, as we will see, computer models can be thought of in exactly the same way. But as we will also see, it is possible to have model-like representations that are not also language-like. And it is because this possibility is open that there is room for developing a representationalist alternative to computationalism, by thinking of mental representations by analogy with scientific models. One of the main aims of the thesis is to show that such an enterprise might be promising. My aim is not to present a refutation of the computational theory of mind, but rather to suggest that a promising alternative to it might be waiting in the wings.

I do so in the following steps. In chapter 2, I develop a theory of s-representation that can answer some of the key philosophical challenges that any successful theory of mental representation must address, distinguishing my own theory from what I take to be its less successful predecessors. In chapter 3, I show how the 's-representation' notion for which I have developed an account might be put to work in empirically significant ways. I also show how the account can meet certain sceptical challenges raised by those suspicious of representational explanations of mental phenomena.

In Part II (chapters 4, 5 and 6), I try to establish that the notion of s-representation can be put to work in making sense of (some of) Alva Noë's more interesting claims about the nature of perception. This is not the only aim of Part II,

however. The other aim is to identify and improve upon what I take to be Noë's keen insights, and to separate these off from what I take to be his more eye-catching (but less plausible) claims. S-representations play an important role in my attempt to do this, but much of my analysis could be accepted even by those who did not think much of the s-representation story developed in the first half of the thesis.

The two key features of Noë's theory that I take to be promising are his 'virtual content' theory and his explanation of our ability to see perspective-independent properties. The 'virtual content' theory is supposed to solve what I call the 'problem of retinal paucity', which can be stated briefly as follows. It seems to us that our visual field (that which sighted individuals experience more or less whenever we open our eyes) seems detailed and richly coloured all the way out to the periphery. But this phenomenological fact is puzzling, since only the central region of the retina is capable of registering detailed information about colour and form. And normally, the only light rays that hit this region of the retina are those reflected off objects at which our eyes are directly pointing. Given this, one might expect our visual periphery to seem sparse in colour and detail. But this is not how the periphery seems to us, or so the claim goes. Noë's 'virtual content' theory is supposed to solve this problem. I argue that, in its current form, the solution is inadequate. But I then show how it might be improved, so as to do the necessary work.

The second promising idea I look at is Noë's explanation of how we get from the perception of perspectival properties to the perception of perspective-independent properties. An example of a perspectival property is the elliptical aspect presented by a coin, when viewed from a certain angle. An example of a perspective-independent shape is the coin's circularity. A coin is circular regardless of the position from which it is viewed. But it only looks to have a certain elliptical perspectival shape from a certain perspective (or a certain range of perspectives). Noë's claim is basically that we see a coin's circularity by seeing its perspectival properties, and understanding how these perspectival properties change (or would change) as a function of changes in the spatial relations between us and the coin. When I move relative to the coin, its visible perspectival shape changes. And the exact manner in which it changes is (in part) determined by the fact that it has a

particular perspective-independent shape. When we (or our visual systems) somehow understand this fact, we become able to see the coin's non-perspectival shape *by* seeing the manner in which its perspectival properties change. More generally, we see perspective-independent properties by seeing perspectival properties and understanding how these perspectival properties vary as a function of movement. As I will argue, this basic picture is on the right lines, but stands in need of significant revisions.

Part II runs as follows. In chapter 4, I discuss various claims that Noë seems to endorse at some point or another about the relation between action and perception. I give reasons for finding most of these claims implausible. But I identify his perspectival content story as a view with promise, before delaying further discussion of it until chapter 6. In chapter 5, I discuss the 'virtual content' theory and make the necessary alterations to it. In doing so, I set out some of the key claims that will have a bearing on my eventual (chapter 6) discussion of the 'perspective-independent content' part of the story.

The reader may have noticed that I have not yet said anything about what I will be up to in chapter 1. In this chapter, I examine Clark's argument for the extended mind thesis. This is, roughly, the thesis that the mind is made up not just by the brain, but by parts of the body, and even parts of the environment at well. A more precise definition will follow, but this gives the rough idea of the thesis. I look at Clark's attempt to derive this thesis from the common-sense functionalist position advocated by Jackson and Braddon-Mitchell (2007). I argue that his attempt is unsuccessful, as is Mark Sprevak's attempt to turn Clark's argument into a *reductio ad absurdum* of the functionalism on which it relies. There is a sense in which this chapter swings free of the rest of the thesis. I do, in later chapters, discuss ideas that have often been taken to have a bearing on the debate over whether the mind is extended. But I don't spend much time discussing their bearing on this thesis. This is because, as I briefly argue in chapter 1, I think these ideas can, and should, be treated independently of the 'extended mind' issue. And while I do make use of some of the key ideas developed by Jackson and Braddon-Mitchell (2007) in my account of s-representations, my account doesn't assume the truth of their theory as a whole.

Part I

Chapter 1: Functionalism and the Extended Mind

Abstract:

This chapter focuses on Andy Clark's argument for the extended mind thesis (EM henceforth), and the 'common-sense' brand of functionalism on which it relies. Mark Sprevak claims to have developed a *reductio ad absurdum* of both Clark's argument and the functionalist presuppositions on which it relies. I argue that this *reductio* only works if we accept an optional assumption about mental kinds, an assumption that is rejected by the very functionalists on which Clark's arguments draws. So long as this assumption is rejected, Sprevak's *reductio* fails. And so too does Clark's argument for EM. With this result established, I will explore its wider implications for the status of both common-sense functionalism and the wider debate about EM. I will argue that things look rather better for functionalism than they do for EM. This is convenient for my purposes, as I will be gleaning ideas from the functionalist tradition in later chapters.

§1: The Mark of the Mental and EM as Extended Functionalism

It is widely acknowledged by writers in the EM literature that the EM debate must be settled by appeal to a mark of the mental.[1] If this is right, then the advocate of EM must find a set of conditions sufficient for mental status, and then show that some extra-cranial states of affairs satisfy these conditions. Meanwhile, the opponent of EM must find a set of conditions necessary for mental status and show that in all putative cases of mental extension, at least one of these conditions is not met. If this view of the debate is right, then those, like me, who think that such analyses of concepts like 'mental' cannot be given will hold out little hope for a resolution to the EM debate any time soon.[2] But I will bypass this issue here by concerning myself with one of the few arguments for EM in the recent literature that does not rely on some mark of the mental or another, an argumentative strategy

[1] This is a point of agreement among both friends of EM – for example, Rowlands (2009, 2010a) and Wheeler (2010a) – and foes of EM like Adams and Aizawa (2001, 2010) and Rupert (2010b). Although some prefer the term 'mark of the cognitive' (I will discuss the mental/cognitive distinction in more detail below).

[2] I should say, more precisely, that I suspect there can be no non-trivial, counterexample-proof, finite analyses of concepts like 'mental' (see Chalmers and Jackson 2001 for more on this issue). My reason for pessimism on this score can best be described as the standard pessimistic meta-induction from such failed attempts at conceptual analysis as that sparked by Gettier (1963). I will not argue in detail for this pessimism, since I will not be relying on it in what follows. Note also that one could be sceptical about the prospects of finding a mark of the mental while being optimistic about the prospects of other conceptual analyses. One might think, for example, that there is something about the concept 'mental' that makes it particularly resistant to analysis.

developed and refined by Andy Clark. Before examining his argument, however, we need to get clear on what the EM thesis, as formulated by Clark, amounts to.

Clark has recently started referring to EM as a form of 'extended functionalism'.[3] By doing so, he implies that his argument for EM doesn't just assume a form of functionalism. Rather, the claim seems to be that EM *just is* a form of functionalism. Clark is not alone in thinking about EM in this way; a growing number of theorists endorse the idea that EM is a functionalist thesis.[4] I suggest that we capture this idea by defining EM as follows:

EM: At least sometimes, in the actual world, the realisers of functional roles definitive of certain mental items include extra-neural components or processes.[5]

More needs to be said about this definition before we can proceed. First, we need to deal with issues of scope. As will become clear in what follows, the scope of EM, as defined above, is vague. This vagueness in scope means that the definition could be accepted by most advocates of the thesis. But at the same time, we can tailor the definition's scope so that it provides a good fit with a particular version of EM. Let's see how the definition can be tailored to fit with Clark's theoretical commitments.

To what class of 'mental items' is EM supposed to apply? EM can come in a variety of strengths. It can apply to conscious states, intentional states, proto-intentional cognitive states, or some combination of these options. Clark's version of EM applies to intentional states like beliefs and proto-intentional representational states (those which are inaccessible to consciousness), but it does not apply to conscious states (or the conscious aspects of intentional states, if such aspects exist).[6] I will be focussing on his argument for the claim that EM applies to intentional states (specifically, non-occurrent beliefs). I will also be understanding the 'items' part of 'mental items' such that it refers not only to mental states like beliefs but also

[3] See (Clark; 2008b,p.37 & §3) and (Clark; 2010a, p.449).
[4] See, for example Michael Wheeler (2010a, 2010b). For a commentary on the link between EM and functionalism, see Wheeler (2010b). See also Di Paolo (2009) for an attempt to develop an anti-functionalist, 'autopoetic' version of the extended cognition thesis.
[5]There are good exegetical reasons for thinking that Clark has something like this definition in mind when he talks of EM. Due to limitations of space, I cannot go into these reasons here. But see (2008a, xviii) for what I take to be a more evocative (but less precise) functionalist definition of EM.
[6] See Clark and Chalmers (1998) for original statement of the view and its scope. See Clark (2009) for an argument against the view that EM is true of conscious states. For arguments in favour of EM about conscious states/processes, see Hurley and Noë (2003), Noë and Thompson (2004a, 2004b), Rowlands (2010a, 2010b).

to mental processes, like believing, remembering and so on. Clark takes EM to be true of mental states and mental processes.[7] Finally the reference to 'extra-neural components or processes' is ambiguous between that which is bodily but extra-neural and that which is extra-neural *and* extra-bodily. In other words, EM, as defined above, could be understood as the 'embodied' claim that the mind extends past the brain into the body or as the 'extended' claim that the mind extends past the body and into the world. Clark claims that the mind is both embodied and extended.[8]

The next major task is to clarify the notion of 'functional role' in play in the above definition. Different versions of the extended mind thesis can be formulated (and argued for) in terms of different forms of functionalism. As we will see below, Clark states explicitly that his arguments for EM rely on common-sense functionalism as formulated by Jackson and Braddon-Mitchell (see Clark 2008a; pp.88-9). So I will cash out his version of EM in terms of this view. On this brand of functionalism (as on most brands), to be a mental item is just to be that which plays (or is disposed to play) a certain set of functional roles. So to be a belief is just to play a certain functional role within a given functional architecture – to be caused (typically) by perceptions of the appropriate kind, to interact with other beliefs in the appropriate way, to interact with desires in a way that typically generates actions of the appropriate kind, and so on. So the mental kind 'belief' is identified with a certain functional role. With this identification in place, Jackson and Braddon-Mitchell claim, we can say that the realiser of the kind 'belief' is just that which stands in the causal relations definitive of that kind. Normally, it is assumed that the realiser in question is just some region of the brain, but this is precisely the assumption that EM denies.[9]

---

[7] There are some, like Wilson (2010, p.183) and Rowlands (2010a, pp.63-65), who dissent from this analysis. They claim that EM applies only to mental processes. I will not go into their reasons for doing so, the issue has no direct bearing on the dialectic presented here.

[8] In the literature, there has been some tension between writers in the 'embodied' camp – e.g. Lackoff and Johnson (1999), Shapiro (2004), Noë (2004) – and those in the 'extended camp', like Clark (2008a). For a discussion of this tension, see Clark (2008b) and Rowlands (2010a, p.102-6). The tension is normally traced back to difference between the 'liberal' functionalism presupposed by writers like Clark and the less liberal variety adopted by those in the 'embodied' tradition. I will start by focusing on Clark's position, and its relation to the brand of functionalism on which it rests.

[9] Jackson and Braddon-Mitchell in fact go further than this. They claim that a given token belief that *p* is to be identified with a certain functional role. Put more generally, they think the content of a

The claim distinctive to *common-sense* functionalism of the sort advocated by Jackson and Braddon-Mitchell is that we should look to common-sense folk psychology in order to determine which functional roles are definitive of which mental states. This view stands in contrast with *empirical* functionalism (sometimes called psychofunctionalism), according to which we should look to our best scientific theories in order to determine which functional roles are definitive of which mental states. [10] The idea, roughly, is that ordinary folk rely on a tacit theory about the minds of others in order to make sense of their behaviour on an everyday basis. The job of the theoretician is to make explicit the most central theoretical commitments of this folk psychological theory – the folk platitudes – and to build out of these an explication of the folk roles definitive of mental states. But how are these platitudes to be identified? And, perhaps more importantly, why are they to be trusted? Jackson and Braddon-Mitchell's answer to this is that the folk platitudes are predictively and explanatorily potent, while at the same time being theoretically modest (2007, pp.268-272).[11] We can best see what they mean by considering an example.

One folk psychological platitude might be that if a person believes that action A will bring about state of affairs S, and desires that state of affairs S is brought about, then the person will typically be disposed to perform action A. This folk platitude tells us about the functional roles typically exhibited by beliefs, desires and actions, in that it tells us how they interact causally. It is explanatorily/predictively potent, because it is the kind of assumption that allows us to predict and explain otherwise baffling human behaviour. For example, when I see people crouching in a mountain stream, passing mud through sieves, I can explain this behaviour by attributing to them the belief that there's gold in the stream and the desire to find it. Once I have made an attribution of this kind, their behaviour is neatly explained.

---

given mental state is determined by its functional role, and nothing besides. As far as I can see, Clark's argument for EM does not rely on this additional claim.

[10] In this respect, his arguments for EM differ from that put forward, for example, by Wheeler, which seems to rest on a (nuanced) version of empirical functionalism (see Wheeler; 2010b, §5). Robert Rupert, a critic of EM, makes direct appeal to empirical functionalism, or psychofunctionalism, in his critiques (2004, p.423; 2009, pp.91-4; 2010, p.345)

[11] What follows is a brief sketch of the motivations for common-sense functionalism, and should not be taken as an attempt to give a full-scale defence of the position – for one of these, see Jackson and Braddon-Mitchell (2007). I will return briefly to the debate between empirical and common-sense functionalism in §4.

But when I make this attribution, I only commit myself to the claim that there is *something* in the person's mind playing the causal role typical of beliefs/desires of the sort attributed. I make no claim about what it is in the head/body of the person that plays this functional role, or how it is that anything can perform a role of this kind. It is in this sense that the attribution is theoretically modest: it is non-committal about how this functional role is actually implemented (or *realised* in the functionalist terminology). This modesty makes it the case that folk psychology is peculiarly unlikely to be proved wrong by future scientific research (2007, pp.269-70).

To see why this might be the case, consider a pharmacological analogy. Imagine we live in a society where doctors have discovered a number of substances that cure pain – certain types of tree-bark, sap gleaned from poppy plants and so on. They have no idea how these substances work, but they have extremely strong evidence for the claim that they are in fact efficacious when ingested. The doctors have made a non-trivial discovery which gives them predictive power – they know that a person in extreme pain will feel better having ingested one of these substances. It also gives explanatory power. They can explain the fact that the person addicted to one of these substances fails to notice when they injure themselves quite badly. They explain this fact in terms of the substance's proclivity to kill pain. Such an explanation is not a full one, since we do not know *how* the substance involved kills pain. But it is nonetheless a non-trivial explanation. It allows us to predict, for example, that once cured of addiction, the addict will regain the ability to notice injuries. So their discovery is predictively and explanatorily potent.

Notice that the explanation is also theoretically modest. Since they make no claims about *how* these substances cure pain, the doctors avoid making risky explanatory claims of a certain sort. They can therefore be pretty confident that future research will not falsify their discoveries. Of course, it could turn out that our observational evidence was not as strong as it seemed, and that some of the substances they thought of as painkillers do not in fact kill pain. But we can imagine a society in which the relevant observational evidence is extremely strong, even

though the mechanism by which the drug acts is understood poorly or not at all (in fact, there are many drugs of which this is true in our society).

So we can have a very good understanding of the causal role characteristic of a given drug while having no understanding at all about how the drug produces its effect. And we can be extremely confident in our prediction that future research will not falsify a given claim about the effects of some drug, even if we have next to no theory (or no confidence in the best available theory) of how it produces this effect. Platitudes about folk roles are supposed to be analogous to this case. They are neutral on questions of realisation, and so do not go beyond the observational evidence. And the observational evidence in support of them is supposed to be extremely strong. So when trying to identify the platitudes of folk psychology we must look only to those folk claims about functional role that are extremely well supported by observational evidence, but which are also explanatorily/predictively potent. Consider the claim that perceiving an event typically leads to the belief that the event in question occurred. This claim is not exceptionless (think of depicted events, for example), but then neither is the claim that aspirin kills pain. Generalisations about functional role behaviour can be governed by *ceteris paribus* clauses while still being predictively and explanatorily useful.

The next issue to address is the distinction between EM and content externalism. The traditional definition of the extended mind thesis as 'vehicle externalism' was supposed to mark out this distinction clearly.[12] Content externalism, famously motivated by Hillary Putnam's (1975/1985) twin earth thought experiment, is the view that the contents of at least some mental states are individuated (or, in some stronger versions of content externalism, constituted) by external, worldly features. By contrast, the standard line goes, extended mind is about the vehicles of mental states – the carriers of mental content.[13] My functionalist definition gives a way of preserving what is right about this distinction while bringing it into the functionalist fold. To see how, let's consider the contrast between EM and the long-armed functionalism designed to accommodate content

[12] See Hurley (1998) and Rowlands (2006) for seminal characterisations of the EM thesis as vehicle externalism.
[13]See Dennett and Kinsbourne (1992) for a seminal and entertaining treatment of the vehicle/content distinction.

externalism.[14] According to long-armed functionalism we must, when specifying the functional role definitive of some mental item, include clauses that specify causal interactions with distal (environmental) objects. So, for example, we might specify perceptual states in terms of the worldly objects that are disposed to cause them, rather than specifying them in terms of the retinal or neural inputs that are disposed to cause them. So according to long-armed functionalism, the specification of the *roles* definitive of mental items must contain essential reference to beyond-the-skin factors. By contrast, EM is a thesis about the *realisers* of mental items; it is the claim that the functional roles definitive of mental items are *realised* in part by extra-neural goings on. One might at this point ask how Clark moves from common-sense functionalist claims about folk functional roles to the EM claim, which is a claim about the realisers of these roles. To answer this question, we have to look at Clark's argument for EM.

§2: Clark's Argument for EM

Clark's best-known argument for EM is the Otto and Inga argument, originally developed in a paper co-authored with Chalmers (1998), but repeated and refined in Clark's later work. It is on this argument that I will focus initially. The aim of Clark's argument is to show that there are situations in which some mental item (non-occurrent belief in the case I consider) is realised partly by extra-neural resources. The argument starts with Inga, who constitutes a benchmark case – a case in which non-occurrent beliefs are uncontroversially realised. Inga believes that the Museum of Modern Art (MoMA) is on 53rd street. When she wants to go to MoMA, she consults her memory, retrieves her belief about its location, and goes on her way. A more controversial case (Otto) is then introduced. Otto has Alzheimer's disease, so cannot form and retrieve non-occurrent beliefs in the normal way. But he has a way around this problem. Instead of storing information in his brain, he stores it in his notebook. When he is told that 'MoMA is on 53rd street' he writes this sentence down in his book. When he wants to go to MoMA, he consults his notebook, discovers its location, and goes on his way. The aim of the

[14] See Jackson and Braddon-Mitchell (2007, p.119) for a defence of this kind of functionalism See Block (1990, p.58) for introduction of the term 'long-armed functionalism'.

argument is to show that Otto, like Inga, is capable of forming and retrieving non-occurrent beliefs (even though he does so in an unorthodox way).

The argument rests on the claim that, at the appropriate level of abstraction, the procedure performed by Otto is functionally analogous to the procedure performed by Inga. This is where common-sense functionalism comes in. The idea is to appeal to folk psychological intuitions in order to establish the claim that there is no intuitively significant functional difference between the procedure performed by Otto and that performed by Inga. And if there is no significant functional difference between Inga and Otto, then the common-sense functionalist has no justification for claiming that Inga has non-occurrent beliefs while Otto lacks such beliefs. For the common-sense functionalist, forming/retrieving beliefs is just a matter of realising certain functional roles, so if Otto and Inga are functionally analogous in the relevant respects, then they are both forming/retrieving beliefs.

And once it is established that Otto is capable of forming/retrieving beliefs when he uses his notebook in certain ways, it is a relatively short extra step to the claim that these beliefs are partly realised by the relevant parts of his notebook (or processes that constitutively involve these notebook-parts). Some object to this extra step, but it seems to me an easy one to make.[15] If Otto really does believe that MoMA is on 53rd street once he has written this sentence in his book, it would seem unmotivated to claim that the sentence written in his book forms no part of the realisation base for this belief. But my argument does not rest on this claim, so I will not argue for it. I will be focussing on the argument to the claim that Inga and Otto are functionally analogous in all relevant respects.

The argument for this claim is basically that we can get from Inga's case to Otto's without subtraction of any functional role properties that are essential to the mental item in question (non-occurrent beliefs). Clark claims that there are no relevant functional differences between Otto and Inga just in case the following conditions are satisfied by Otto's notebook-involving procedure:

[15] Shapiro (2008), for instance, raises the problem of distinguishing between a realiser of some functional role and something that merely contributes causally to the realisation of said role.

> 1) That the resource be reliably available and typically invoked (Otto always carries the notebook and won't answer that he "doesn't know" until after he has consulted it).
>
> 2) That any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.
>
> 3) That information contained in the resource should be easily accessible as and when required.
>
> (Clark, 2010b, p.46 conditions quoted verbatim)

Clark stipulates that Otto's notebook-use satisfies these conditions. And, he argues, when these conditions are satisfied, Otto's procedure is sufficiently similar to Inga's to justify the claim that Otto is, like Inga, capable of belief storage/retrieval.

To do so, he essentially lays down a challenge to his opponents. The challenge is to find some intuitively significant difference between the functional roles instantiated by Inga and those instantiated by Otto. Let's call such a difference a *Deep Difference*. If, after a suitably long period of time and effort, nobody can point to a genuinely *Deep Difference* between Inga and Otto, we can (defeasibly) conclude that there are no such *Deep Differences* to point to. But how do we test such a difference in order to determine if it is genuinely deep? Clark's procedure for testing is to create a thought experiment designed to test the putative *Deep Difference* against our folk intuitions. It is here that the commitment to common-sense functionalism comes in. Clark justifies such appeal to folk intuitions by invoking Jackson and Braddon-Mitchell's common-sense functionalist claim that 'normal agents command a rich (albeit largely implicit) theory of the coarse functional roles distinctive of various familiar mental states' (Clark; 2008a, p.88). He writes that it is this 'coarse or common-sense functional role that […] displays what is essential to a given mental state' (2008a, p.89).

Let's look at an example of this procedure in action, putting to the test the recency/priming/generation difference raised by Adams and Aizawa (2001).[16] Let's focus on recency: if I am asked to remember items on a list read out to me by someone, I am more likely to remember those items on the list that were read out last (*ceteris paribus*). Not so with Otto (assuming he has a neat filing system in his

[16] See also Rupert (2004).

notebook that stops old entries from becoming swamped under by newer ones). This is a difference between the functional profile of Otto's putative memory system and what we take to be paradigm cases of memory systems. Otto's notebook system also fails to exhibit other quirky features characteristic of human memory (priming and generation, for example). Assuming that Inga exhibits these quirks of memory, we have here a functional difference between Otto and Inga.

To show that this difference is shallow, Clark imagines a Martian, or just an unusual human, who doesn't exhibit the recency effect, or similar memory effects of this type, but who is just like normal humans in all other psychological respects (2008a, p.93). Would we claim that such people/aliens don't remember? Plausibly not. As Clark puts it 'to insist that some alien mode of storage and retrieval was not cognitive *just because* it failed to exhibit features such as recency, priming and crosstalk would be simultaneously to scale new heights of anthropocentrism and neurocentrism' (2008a, p.93 emphasis added). According to Clark, if we met such a creature, the appropriate thing to think would be that their memory systems were slightly unusual. And if we think this is the right thing to say about the alien case, what motivation, other than neural chauvinism, do we have for applying a different standard to Otto?

Notice how this thought experiment works. The original claim is that Inga exhibits a certain functional feature lacked by Otto (the recency feature, for example). Clark shows that exhibiting functional features like recency/priming/generation is not necessary for instantiating the functional role definitive of belief storage/retrieval. He does so by imagining a creature that lacks this feature, but is like us in other respects. He then claims that we would not deny that such a creature is capable of belief storage/retrieval *just because* it lacks the functional features in question. This is just to say that such functional features are not individually necessary for belief storage/retrieval. The idea then seems to be that, if we can get from Inga to Otto by subtracting from Inga without the subtraction of any functional feature that is individually necessary for realising belief storage/retrieval, then we are justified in claiming that Otto instantiates belief storage/retrieval, just so long as Inga does. So Clark's challenge to his opponents

can be put as follows: show me some *Deep Difference* between Otto and Inga, where a *Deep Difference* is defined as follows:

A) Inga exhibits functional feature F_i but Otto doesn't.
B) Exhibiting F_i is individually necessary for realising mental item *M*.

Each time an opponent raises a putatively significant functional difference between Otto and Inga, Clark constructs a bespoke Martian thought experiment to show that this difference is merely shallow. Some illustrative examples are as follows:

F_i = the property of not inspecting one's memories by perceptual means. Counterexample to (individual) necessity: Terminator (Clark and Chalmers; 1998, p.16)

F_i = the property of involving non-derived as opposed to derived content.[17] Counterexample to individual necessity: Martian 'endowed with an extra biological routine that allowed them to store *bit-mapped images* of important chunks of visually encountered text.' (2010c, p.88) Clark claims that 'If we accept the Martian memory into the cognitive fold, surely only skin-and-skull prejudice stops us extending the same courtesy to Otto' (2010c, p.89).[18]

F_i = the property of being a persisting and integrated part of the relevant cognitive system.[19] Counterexample to (individual necessity): Metamorpho/Metamento (Clark; 2010a, p.457).

This list is far from exhaustive, but it gives a sense of Clark's argumentative methodology. Until someone finds some *Deep Difference* that is immune to counterexamples of this kind, there is a presumption in favour of the claim that Otto and Inga are functionally alike in the relevant respects. I do not go into the details of Clark's particular counterexamples because such details are not relevant to my argument. I will be taking issue with the overall argumentative strategy employed

[17] See Adams and Aizawa (2001; 2010a) for more on this distinction. The rough idea is that linguistic representations derive their meaning from convention, or from individual stipulations. By contrast, the representational status of mental states like beliefs and desires is not supposed to be derivative in this sense
[18] See Clark (2003), Clark (2005) for more details.
[19] See Rupert (2004, 2009).

by Clark, not the particular thought experiments he deploys. But first let's look at Sprevak's (2009) *reductio* of Clark's argument.

Sprevak's *reductio* works by applying Clark's own argumentative strategy to the three conditions which, by Clark's stipulation, must be satisfied if Otto's notebook-use is going to count as a genuine case of belief storage/retrieval. To see how this works, let's consider Clark's first condition:

1) Condition: the notebook must be 'reliably available and typically invoked'. Counterexample: Imagine a Martian whose putative 'non-occurrent beliefs' are neither reliably available nor typically invoked. He sometimes (maybe even often) manages to recall facts learned at an earlier time, but his ability is not very reliable. Sometimes, after a bout of heavy drinking (say), his 'memory' just packs up and stops working temporarily. When his memory-recall isn't working well, he will often say he 'doesn't know' when asked some question, without first consulting his memory (which, he detects, is playing up again). Does this Martian lack the capacity to remember? Is the Martian entirely lacking in non-occurrent beliefs?[20]

This procedure is then repeated on the other two conditions. To undermine the second condition (the automatic endorsement condition) Sprevak invites us to imagine a creature who 'redundantly' performs a quick plausibility check on any putative belief they retrieve, and thereby fails the automatic endorsement condition (2009, pp. 514-5). Such a creature, he argues, would still count as a believer. And to undermine condition three, he claims we would not deny that someone has beliefs just because s/he finds those beliefs 'difficult to access' (2009, p. 515).

Let's suppose (as seems plausible) that Sprevak's counterexamples are successful. Let's also suppose, for the sake of argument, that we agree with Clark that Otto realises belief storage/retrieval. One could justifiably ask at this point how far we should generalise from the Otto case. Does anyone who reads anything in a notebook (or any artefact bearing written symbols for that matter) also count as retrieving non-occurrent beliefs? Clark's three conditions were supposed to block

[20] I construct my own version of the counterexample, rather than quoting laboriously from Sprevak. For original version, see (2009, p.514). Nothing I say rests on the exact details of the counterexample.

absurd consequences of this kind. The idea, for example, is that most symbol-bearing artefacts we use are not reliably available and typically invoked, so most reading we do falls short of belief retrieval by the lights of Clark's conditions.

But if Sprevak's counterexamples are successful (and they seem every bit as compelling as Clark's own Martian cases), then Clark is not entitled to rely on his three conditions when attempting to exclude obviously absurd cases of putative cognitive extension. This is because the conditions fail to pick out genuinely *Deep Differences* between Otto and a given absurd case of cognitive extension. So Sprevak can then confront Clark with the following challenge: show me some *Deep Difference* between Otto and Blotto (where 'Blotto' is just some obviously absurd case of mental extension). If Sprevak's counterexamples are successful, none of Clark's three conditions can succeed at this job. And until some more successful conditions are formulated, we can defeasibly conclude that Clark's argument generalises to create absurd results.

§3: Diagnosing the Problem

Something has clearly gone badly wrong here. Sprevak's suggestion is that the problem lies with common-sense functionalism, and its attempt to systematise our intuitions:

> The correct lesson might be that our intuitions about mental systems cannot be systematised without doing serious damage to our concept of mentality. Functionalism aims to provide an answer to what makes certain systems mental. Perhaps such an answer cannot be given. (2009, p.522)

I argue that an alternative diagnosis is available. To see this, let's review Clark's basic strategy. There are many functional differences between the procedure performed by Inga and that performed by Otto. Let's use $F_{1,\ldots,n}$ to denote the full set of these differences. Each time Clark's opponent tries to claim that some difference on this list is a *Deep Difference*, Clark creates a Martian-style thought experiment in order to support his argument for the claim that the functional property lacked by Otto, F_1 say, is not individually necessary for realising the mental item in question. Sprevak repeats this procedure, but instead of using it to move from Inga to Otto, he uses it to move from Otto to Blotto. But this argumentative strategy relies on a critical assumption that I will attempt to undermine.

To see what the assumption is, suppose Clark successfully undermines the individual necessity of F_1 for realising mental item *M*, and then does the same for F_2, and F_3. Even if he does this successfully, he has not ruled out the possibility that realising some combination of these three conditions *is* necessary for realising *M*. To see why this possibility might undermine Clark's argument for EM, let's look again at the list of functional features $F_{1,\ldots,n}$ which are realised by Inga but not by Otto. Clark might show that no feature on this list is individually necessary for realising *M*, while failing to show that there is no combination of features on the list whose realisation is necessary for realising *M*.

Suppose I put it to you that the drum stick I am holding in my hand is a pen, and defy you to convince me otherwise. I challenge you to point out a single feature lacked by my drumstick, which is essential to penhood. You point out that I can't write with my drumstick. But then I reply that something can be a pen even if it's not possible to write with it. A pen that's broken, out of ink, or locked in a secure safe to which nobody has the key is still a pen, despite the fact that we can't write with it. You then object that it doesn't even look like a pen, and I respond again that this feature is also inessential, concocting a range of clever thought experiments to prove my point. This goes on for some time and eventually, whether out of exhaustion or boredom, you give up trying to convince me. Suppose I lay down this challenge to lots of clever people, and each time wear them down in a similar fashion. Does the fact that I can do so give me a strong (though defeasible) justification for the claim that my drum stick is a pen? Plausibly, it does not. And the possibility is open that in the relevant respect, Otto is to Inga as my drumstick is to some paradigm example of a pen.

And Jackson and Braddon-Mitchell, on whose common-sense functionalism Clark explicitly relies, use exactly this analogy to describe mental states:

> There is a list of features that we regard as paradigmatic of pens. Something that satisfies every single one is by definition a pen. If it uses ink, is small enough to fit in the hand, is used to write with , is called a 'pen', is barrel shaped, and has a nib, then it is a pen. That follows from our concept of a pen. But nothing in that list is sacrosanct. Any single feature may be absent and yet the object still be a pen […] what matters is that enough of the list or near enough is satisfied, and what counts as enough may itself be a vague matter (2007, p.54)

The key point is that a pen can lack any of the properties enumerated by Jackson and Braddon-Mitchell, while still counting as a genuine pen. But, the idea goes, if enough of the relevant properties are absent, we reach a point where the right thing to say is that the thing in question is not a pen.

When applied to the mental, Jackson and Braddon-Mitchell's idea is that each functional role property by which a given mental item is defined should be qualified with a *ceteris paribus* clause. Take beliefs, for example. Beliefs are typically action-guiding, but we could imagine lots of beliefs that are not. This point is illustrated by David Lewis' resolute deceiver, who is 'disposed come what may to behave as if his mental states were other than they really are' (1994, p.418). Such a person may be a slightly atypical believer, but is a believer nonetheless. Lewis claims this is the case while denying that 'anything goes' when it comes to mental state attribution (1994, p.418).[21] Jackson and Braddon-Mitchell, whose brand of common-sense functionalism is heavily indebted to Lewis, take a similar line.

More, however, needs to be said about how this proposal is supposed to work and how it might be deployed to block Clark's and Sprevak's arguments. Jackson and Braddon-Mitchell give less guidance than would be desirable about how their proposal is supposed to work, so I suggest we turn to Lewis, whose remarks on this matter are instructive. In discussing cases similar in kind to those of Otto and Inga, Lewis suggests that we posit a kind of ambiguity about exactly which folk roles are definitive of which mental items. He claims that this ambiguity is not concocted *ad hoc* by the common-sense functionalist, but is rather 'a commonplace kind of ambiguity – a kind that may arise whenever we have tacit relativity and criteria of selection that fail to choose a definite *relatum*' (1980a, p. 221). I propose that 'typicality' is the relevant criterion of selection we should use when determining which folk roles are definitive of which mental item. There are many different respects in which a given mental item might be typical of its kind, so there are many different typicality criteria against which a given mental item might be judged.

To see how this proposal works, let's take belief again. One axis of typicality along which a belief might be assessed is its action-guiding role. As the deceiver

[21] See also Lewis (1980a) for a seminal statement of this view.

case shows, a state can count as a belief despite being atypical on this axis. But there are many other axes of typicality against which a given putative belief might be assessed. Below is an illustrative but inexhaustive list:

1. Evidence-sensitivity: typically, the credence we attach to a given belief is sensitive to the evidence we have available for that belief. But there are *bona fide* beliefs that are atypical in this respect – irrational beliefs being an obvious example.[22]
2. Involuntariness: beliefs are typically involuntary in a meaningful sense. I typically cannot believe any old thing at will, but a skilful self-deceiver might be able to do so.[23]
3. Internal role: beliefs typically interact with other beliefs (and other mental states), often creating new beliefs (and other mental states). The paradigm example of this is inference. Beliefs typically combine in certain ways with other beliefs to generate new beliefs. But it is possible for beliefs to be compartmentalised to an atypical extent while still counting as beliefs. Until Clark stipulates otherwise, we must assume that Otto's beliefs are thus compartmentalised (Clark has not specified a mechanism by which the sentence written by Otto about MoMA will interact with other sentences written in his notebook).
4. Casual role: Otto's 'beliefs' are atypical with respect to causal role. For example, we can render him incapable of accessing them by poking out his eyes.

Beliefs can be atypical in certain respects, but typical in others. If a putative belief is atypical in one respect, but typical in all others, it might make sense to say that it is still a belief. But if it is atypical in just about every respect, we might justifiably wonder if it is a belief in any meaningful sense. Ambiguity arises when we try to determine what should be said about the cases in the middle. If some putative belief

[22] Some people, like Hohwy and Rajan (2012), question the doxastic status of extremely irrational beliefs (delusions). But this fact should call us to question the claim here made. First, there are many who think it is wrong to deny the doxastic status of beliefs (e.g. Bortolotti 2009). Second, one could deny the doxastic status of delusions while still maintaining that states which are irrational in a less extreme sense can still count as beliefs. Third, many of the arguments against the claim that delusions are beliefs rest on the fact that delusions are not only epistemically atypical, but atypical in other respects too (for discussion, see Bortolotti 2010). In the case I am imagining, the beliefs are epistemically atypical, but in all other respects typical.

[23] See Sterelny (2004) for discussion of a similar point in the relation to Otto's notebook.

is belief-typical on certain axes of typicality, but atypical along others, it might be difficult to say one way or another whether it is suitably close to paradigm cases of belief to count as a mental state of the same kind. The typicality axes referred to here are derived directly from the folk platitudes – the folk platitudes dictate that mental item *M* typically exhibits a certain set of functional features. But the folk platitudes do not tell us what we should make of putative mental items that are typical in some respects, but atypical in others.

If something like this story is right, then there is something seriously wrong with Clark and Sprevak's methodology. They challenge their opponents to identify some respect in which Otto or Blotto are atypical. Once the opponent identifies some relevant respect, Clark and Sprevak proceed to show that a person (or an alien) could be atypical in the respect identified while still instantiating the relevant mental item. They then repeat this procedure for each respect identified by some opponent. But this isn't good enough. One has to show that the total sum of respects in which Otto or Blotto are functionally atypical (with respect to some mental item) is outweighed by the total sum of the respects in which they are typical. Until this is done, the question of whether Otto or Blotto realise the relevant mental item must remain open. And from this, it follows that common-sense functionalism of the kind outlined here has not been shown to imply either EM or the absurd conclusions derived by Sprevak.

§4: Drawing the Line

So far so good for the common-sense functionalist, but one might start to worry that the line drawn between the mental and the non-mental is starting to look a little arbitrary. Lewis assures us that his story does not imply that 'anything goes' (1994, p.418), and Jackson and Braddon-Mitchell (2007, p.55) repeat this assurance. This seems right in the sense that while we might struggle to draw the line in certain borderline cases, we can certainly draw the line between that which definitely is a belief and that which definitely is not. That which scores highly on all relevant typicality axes is clearly a belief. That which scores highly on none of them is clearly not. So we have a method for making meaningful distinctions in at least some cases. But is the common-sense functionalist's project in any way undermined by the fact that the folk concepts, when explicated, deliver no clear means of

distinguishing between beliefs and non-beliefs in borderline cases? To answer this question, we need to remind ourselves of the common-sense functionalist's theoretical goals, and then ask whether any of these goals are undermined by the existence of borderline cases on which the folk concepts give us no clear guidance.

The aims of (at least most) common-sense functionalists are twofold: the vindication of folk psychology, and the reduction of folk psychological postulates (like beliefs) to the physical. I will start with the latter claim, since it is easier to deal with in this context. The basic strategy of reduction is to claim that states like belief just are complex sets of functional roles, and then to show that these functional roles are realised by complex organisations of physical matter. [24] The existence of borderline cases seem unlikely to threaten this reductive project, unless there is some suggestion that there is something going on in the borderline cases that goes beyond what can be accounted for in terms of physically realised functional roles. And it is far from obvious that borderline cases give us special reasons for thinking that this is the case. The former aim, the vindication of folk psychology, is trickier to deal with, since there are two importantly different threats to folk psychology which must be diffused if the vindication is to be successful. I will take these in turn to see if either threat is made more pressing by the existence of borderline cases of the sort discussed above.

The first is the strong eliminativist's challenge, where strong eliminativism is just the view that folk psychological postulates, like beliefs, do not exist. Like phlogiston and daemons, they will one day be superseded by scientific concepts and consigned to the dustbin of history. Recall that Jackson and Braddon-Mitchell argued for the existence of folk psychological postulates on basis that folk psychological attributions were predictively/explanatorily potent, and theoretically modest. We explained the behaviour of the gold-sifter, for example, by attributing to him a belief and a desire. This attribution gave us predictive and explanatory power, and did so while making only a very modest empirical commitment. This modesty was ensured by the fact that we said nothing about what it was that realised the relevant beliefs and desires; we only said that there must be some aspect of the

[24] Although Jackson is perhaps best known for his arguments to the effect that such reductions cannot be performed, he is now a paid up physicalist. See Jackson and Braddon-Mitchell (2007, chapter 8) for a partial explanation of this change.

sifter's cognitive system that is realising the functional role definitive of the attributed mental states. We can see the gold-sifter case as a paradigm case: a case in which it is, for the common-sense functionalist, clearly appropriate to make the relevant attributions. The beliefs attributed, let's assume, are typical in all respects, and the attribution passes the modesty and potency test.

Cases like this are the ones that common-sense functionalists use to combat the eliminativist. It is a case where a belief-attribution is so predictively/explanatorily powerful, and yet so empirically modest, that we can say with some confidence, even prior to any empirical inquiry, that the attribution is justified. If the common-sense functionalist's claims about such cases are true, then we have an existence proof against the eliminativist: a case where a belief is clearly instantiated. And if such a case exists, it seems to matter little that there are other cases where things are far less clear. The existence of borderline cases where it is unclear whether or not a belief is being instantiated does not cause problems for the view that beliefs are *at least sometimes* instantiated. We might wonder whether a virus is really a living thing, but doing so should not cause us to doubt the claim that at least some things are alive. So it looks like the existence of borderline cases cause no threat to this part of the common-sense functionalist's story: all the work is being done, for the common-sense functionalist, by the paradigm cases, and so borderline cases are orthogonal to the dialectic.

The second criticism against which the folk psychological postulates must be defended is a charge of uninformativeness. One might accept that strong eliminativism (as characterised above) is wrongheaded, while still claiming that the folk psychological platitudes in terms of which common-sense functionalists explicate the folk concepts are unacceptably vague and uninformative. And one might further think that we should look to the best scientific theories of the mind in order to find more precise and informative functional definitions for the relevant states. To adopt this position is to adopt a form of empirical functionalism. I do not want to get into the standard arguments between these two brands of functionalism, but rather want to ask whether the borderline cases discussed above have a bearing on this debate. And on the face of it they might seem to. The borderline cases are cases where the platitudes of folk psychology give us no guidance on the question of

whether an attribution of a given mental state should be made. But if it turned out that scientific theorising can give us informative guidance on those borderline cases about which folk psychology was silent, this might speak as a consideration in favour of empirical functionalism.

And if we look at the criteria against which folk psychological attributions were, according to Jackson and Braddon-Mitchell, to be assessed, it seems entirely conceivable that such a situation might obtain. Belief attributions, in paradigm cases, were taken to be justified because they were modest and explanatorily/predictively potent. In other words, they were treated as one would treat any scientific postulate, and shown to perform well when so treated. But if we apply these criteria when testing paradigm cases, why would we not do the same when testing borderline cases? We might think that these conditions should not be applied so demandingly in the borderline cases, since we are not using such cases as the basis for being confident, in advance of empirical inquiry, that strong eliminativism is false.[25] But I see no reason for denying that the attributions in the borderline cases should, like attributions in the paradigm cases, be tested against the criteria of predictive/explanatory potency and modesty. All this must mean is that an attribution is justified if it explains/predicts a lot, while not requiring excessive theoretical commitments.

Let's see how this might work in the Otto case, starting with predictive/explanatory potency. Is there a predictive/explanatory advantage to be gained by attributing to Otto the ability to form and retrieve new beliefs by writing sentences in his notebook? To answer this question in the affirmative, it would not be enough merely to establish that we can explain and predict Otto's behaviour by attributing this ability to him. We would also need to show that this is the best available means of predicting/explaining his behaviour. One obvious competing option, considered by Clark and Chalmers in their original paper, was 'to explain Otto's action in terms of his occurrent desire to go to the museum, his standing belief that the Museum is on the location written in the notebook, and the accessible fact that the Museum is on 53rd street' (1998, p.13). On this story, Otto doesn't believe that MoMA is on 53rd street just because he has this sentence written in his

[25] It was the paradigm cases which are supposed to act as the basis for such confidence.

book. Rather, he believes that he might be able to determine the location of MoMA by looking in his notebook.

Is this story more or less predictive/explanatory than the EM story? I'll look at ways of answering questions of this sort in the next section, arguing that neither of these stories has the edge when it comes to predictive/explanatory potency. But first I should emphasise what is at stake in answering this question. If it turns out that there *is* a determinate answer to questions of this kind, then common-sense functionalism is in trouble. Common-sense functionalism, at least of the kind I have been discussing, has nothing informative to say about cases like Otto, other than to label them borderline cases. But if scientific inquiry can say something informative about whether mental state attributions are worthwhile in such cases, then it looks like scientific inquiry is more informative than folk psychology when it comes to distinguishing justified mental state attributions from unjustified ones. And if this is right, it looks like we have a consideration in favour of replacing common-sense functionalism with a version of empirical functionalism.

Common-sense functionalists might object to this characterisation of the dialectic. They might argue that they are quite happy to leave borderline cases to the scientists, and that doing so in no way undermines their position. For example, they might be quite happy to leave scientists to discover whether or not chimps, rats or bees have beliefs, while claiming only that we can know in advance of inquiry that paradigm believers, humans definitely have *some* beliefs. Their agnosticism about certain animals doesn't seem to undermine their claim to have a grasp of what it is to be a believer, so why should a similar agnosticism about Otto cause them problems? This response fails because the cases are different. We might be agnostic about whether some animal has beliefs because we don't know enough about that animal. The common-sense functionalist can lack this knowledge while being perfectly clear about what it is to be a believer. But we cannot so easily put our agnosticism about Otto down to similar ignorance. It is plausible to think that we are unclear about the Otto case not because we lack the relevant knowledge about Otto, but because the relevant mental state concept, as explicated by the common-

sense functionalist, delivers vague results when applied to Otto.[26] And given that it is the mental state concept generated by common-sense functionalism that puts us in this predicament, it had better be the case (for the common-sense functionalist) that rival means of carving up the mental terrain lead to predicaments similar in kind.

§5: Is EM an Inference to the Best Explanation?

By asking whether the Otto case and similar cases can be resolved on empirical grounds, we are asking, in effect, whether EM might be an inference to the best explanation. To do so is to move back to a battleground on which the EM debate is often fought.[27] While the scientific fruitfulness or otherwise of EM has been hotly contested, little consensus has emerged. So there is hope for the common-sense functionalist yet. And as it happens, Sprevak, in another influential paper (2010), has developed a powerful argument for the claim that the debate will not admit of empirical resolution. If the arguments presented so far are on the right lines, this puts Sprevak in the curious position of lending indirect support to the very functionalist position for which he tried to produce a *reductio*; if Sprevak is right, empirical investigation is just as uninformative as the common-sense functionalist's folk platitudes when it comes to resolving the EM debate, so empirical functionalism gains no decisive advantage.

Sprevak's basic argument is that, in order to show that EM is an inference to the best explanation, we would have to show that it is a better hypothesis than all available rivals. But, he argues, there is an alternative rival internalist hypothesis with scientific standing equivalent to that of EM. This position is called the hypothesis of embedded cognition (HEMC) and is formulated by Robert Rupert as follows:

> [C]ognitive 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 (2004, p.393).

[26]I am here relying on the claim that we know all the relevant empirical facts about Otto's notebook-use. Those who deny this claim must point to some fact about the Otto case about which we lack knowledge, where it is our lack of knowledge about this fact that explains our agnosticism.
[27]Adams and Aizawa (2007), Rupert (2004, p.425; 2009), and Hurley (2010, pp.106-7), Wilson (2010, p.173) all claim that this is the terrain on which EM debates should be fought. See next footnote for Clark's statement of dissent on this issue.

Sprevak argues that this hypothesis, conjoined with the claim that EM is false, has explanatory credentials equal to those of EM (2010). Applied to Otto, we can say that cognitive processes realised by Otto depend in all sorts of interesting ways on his notebook, but that the notebook is an external prop on which cognition depends, rather than a constitutive part of the cognitive process. While Rupert (2004) argues that such alternative glosses of putative EM cases perform better than the EM story, Sprevak argues that there is no way of telling which of the two glosses perform better. I will not repeat Sprevak's argument in detail here, although it is worth briefly noting that it is an argument that has convinced Clark.[28] Rather, I will apply my own version of the argument to what I take to be some illustrative cases. My argument is similar to Sprevak's, but not identical. While Sprevak's is pitched at the general level, my argument focuses on two of the cases that seem to be the most clear illustrations of EM's explanatory power, and shows that the results can be accommodated within a framework that denies EM.

Connectionism, pattern-recognition and off-loading

The first idea I want to discuss was born out of attempts to get connectionist networks to perform tasks like logical reasoning (Bechtel; 1994) and arithmetical reasoning (Rumelhart, McClelland, and the PDP Research Group; 1986). Connectionist networks are typically very bad at these sorts of tasks, but very good at tasks like pattern recognition (for example, tasks like facial recognition or the recognition of hand-written words). By contrast, classical symbol systems like computers and calculators are very good at arithmetic and logical inference, but very bad at pattern recognition tasks. The inability of connectionist systems to perform tasks like mathematical and logical inference are often seen as weaknesses which preclude the possibility creating connectionist models of higher level cognitive abilities (which involve these types of inference) in humans. Although humans are significantly less good than classical symbol systems like computers at these forms of inference, they still out-perform connectionist systems by a considerable margin. Connectionist systems just aren't designed to perform these

[28] Clark writes as follows: 'I am now fairly convinced (for a good argument here, see Sprevak, in press) that there will be no straightforward empirical resolution to the questions concerning cognitive extension' (2010a, p.460). If the main argument presented here is on the right lines, this leaves Clark's case for EM in a pretty bad way. If EM cannot be motivated on empirical or conceptual grounds, it's not clear what other options are left.

types of inference. But if this is true, then it seems that orthodox theories in cognitive science (henceforth GOFAI for 'good old-fashioned artificial intelligence' models), according to which cognitive systems somehow realise and process representations analogous to those realised and processed in classical symbol systems, seem to have the edge over cognitive scientific theories inspired by connectionist principles or post-connectionist principles (e.g. those of dynamic systems theory).[29] At the very least, it would seem to follow that GOFAI offers a superior explanation for these kinds of high-level reasoning abilities.

But this situation is potentially changed, for example, by Rumelhart et al.'s innovative connectionist model of arithmetical reasoning. This model first exploits the fact that simple sums, like 2 x 3 = 6, can be performed by pattern recognition and completion systems that are easily implemented in connectionist systems. Once we have a model that can complete simple operations like these by pattern-recognition, we then face the problem of explaining how such a system can complete more complicated arithmetical operations like 343 x 822. More complicated operations like these cannot be performed by pattern recognition, because the number of patterns a system would need to recognise in order to multiply any two three digit numbers would be enormous. This means that complex arithmetic would very quickly become computationally intractable for connectionist systems. But this problem is neatly bypassed by Rumelhart et al.'s model.

To see how, we can first think about how most numerate humans would go about performing such operations. Most people would solve the problem by writing out the sum, and then following the rules of long multiplication. Following these rules amounts to breaking the sum up into a number of much easier sums. For example, the first step in this procedure would be multiplying both numbers in the 'ones' column (i.e. 2 and 3). As we have already seen, a step of this sort can easily be completed by connectionist systems. We would then write down, in the appropriate column, the answer to this simple multiplication, before moving on to the next step, the nature of which is dictated by the rules of long multiplication. This step would be a similarly simple multiplication, meaning that it would also be a multiplication of which a connectionist system is capable. The procedure can carry

[29] The term GOFAI was first coined by John Haugeland (1985)

on in this way until the multiplication is completed. This means that, so long as a connectionist system is coupled to what Rowlands calls 'the capacity to manipulate mathematical structures in the environment' (2010a, p.44), even complex multiplications become relatively straightforward.[30]

This coupling of connectionist systems with external resources considerably enhances the explanatory power of connectionist systems. But, importantly, this coupling would increase the explanatory power of connectionist systems *whether or not we count the manipulation of external symbols as parts of the overall cognitive system.* It doesn't matter whether or not we count the external stuff as literally part of the cognitive system. All that matters is that we appreciate the crucial role it plays in enabling the connectionist system to work its magic. This suggests that, by looking at ways in which cognitive tasks can be offloaded onto the environment, we can develop new and non-standard models of cognitive process, models which would have seemed completely unsuited to the task at hand if they had not be been coupled with offloading activity. Looking for instances of cognitive offloading can be a useful heuristic in developing new theories about how cognitive processes work. And doing so serves this role whether or not we want to say that it is literally the case that external processes count as part of the overall cognitive process. This extra claim adds very little to our understanding of how the target phenomena actually work.

Alva Noë, change-blindness and virtual content:

Noë (2004) claims that instead of having to build detailed inner representational models of the world in order to perceive it we can, to use Rodney Brooks' famous phrase, simply 'use the world as its own best model' (Brooks; 1991, p.139). That is, we can simply access information in the environment as and when we need it, by virtue of our possession of the exploratory skills that allow us to do so. As Rowlands puts it:

> The world, by providing a stable and relatively permanent structure that can be probed and explored at will by the visual modality, obviates the need for

[30] Bechtel (1994) develops a connectionist system that uses similar principles to break down the process of natural deduction into a series of small steps which cause no problem for a connectionist system.

> at least certain sorts of visual representations as these were traditionally understood. (2010a, p. 89).

This is another example of offloading of cognitive tasks onto the environment, although this way of putting it might be misleading. It is not necessarily the case that work that was previously done by the brain was offloaded onto the environment; rather, it is more likely the case that work which *we previously thought* must be done by the brain, turns out to be something that can be accomplished by appropriate and timely explorations of the environment. I will examine Noë's 'virtual content' thesis in much more detail in later chapters, but for now this brief summary will suffice to illustrate my point. If Noë's idea is on the right lines, then it makes significant changes to the way we conceptualise the process of perception, *whether or not we take the exploratory activity to be a constitutive part of the process of perception.* If Noë is right, then we no longer have to explain how it is that the brain builds and updates detailed internal perceptual models of the environment. Rather, we have to explain how it is that perceivers manage to access environmental information as and when they need it. And we can accept this change in explanandum even if we do not think that the mechanisms responsible for the exploratory activity on which Noë's theory relies literally count as part of the system that realises perception.

The Upshot

In both these cases, the idea that cognitive work can be 'off-loaded' onto the environment changes our conception of the nature of the cognitive tasks faced by the brain. It thereby frees us up to propose new models of cognitive processes, models which would not have seemed capable of getting the job done without some kind of coupling with, or offloading onto, external resources. This again suggests remaining on the lookout for cases of cognitive offloading which might serve as a useful heuristic for generating new hypotheses about cognitive processes. The aim of the heuristic would be to keep modellers of cognition on the lookout for ways in which the cognitive load faced by cognitive systems can be made tractable by appropriate use of the environmental resources. The thesis can play this heuristic role regardless of whether we think it is literally true that processes occurring outside the skull or the skin count as properly cognitive parts of cognitive processes.

As well as being very difficult to establish conclusively, this extra claim adds little empirical content to the off-loading idea.

The idea that we treat the extended mind idea as an invitation to consider ways in which cognitive loads might be offloaded is supposed to be similar in spirit Richard Dawkins' formulation of the 'extended phenotype' idea (1982). The 'extended phenotype' thesis was the idea that genes might reach beyond the organisms in which they are realised and find expression in features of the environment. It might, for example, make sense to talk about beavers having a gene for dam structure. In recommendation of his idea, Dawkins writes that 'I have found that the viewpoint represented by the label 'extended phenotype' has made me see animals and their behaviour differently, and I think I understand them better for it. The extended phenotype may not constitute a testable hypothesis in itself, but it so far changes the way we see animals and plants that it may cause us to think of testable hypotheses that we would otherwise never have dreamed of.' (1982, p. 2). I suggest that something similar is true of the extended mind thesis. Clark at times seems to adopt pretty much this position, inviting us to

> [P]ractise the art of flipping among different perspectives [on the EM debate], treating each as a lens apt to draw attention to certain features, regularities and contributions while making it harder to spot others or to give them their problem solving due (2008, p.138).

He adds that when it comes to determining which hypothesis best fits the available evidence, we should 'stop worrying and enjoy the ride' (2008, p.138). It seems to me baffling that he should write such things in the very book devoted to establishing the truth of EM by appeal to common-sense functionalism. By contrast, this quietist attitude to the EM debate sits far better with the position developed above, according to which Clark's arguments from common-sense functionalism deliver indeterminate results on the question of EM.

§6: Conclusions

We have seen that, *contra* Clark, common-sense functionalism of the sort advocated by Jackson and Braddon-Mitchell does not imply EM. And neither does it generate absurd results, as Sprevak (2009) claimed. These results can be blocked once we allow the possibility that mental concepts, and the functional analyses by which

these concepts are explicated, can legitimately be seen as vague in the sense described in §3. Once this possibility is granted, which it is by Jackson, Braddon-Mitchell and other common-sense functionalist, Clark's and Sprevak's arguments fail to generate their intended conclusions. And once this possibility is granted, the common-sense functionalists can consistently treat the EM question as one with no determinate answer. What's more, if Sprevak (2010) and Clark (2010a) are right to think that the EM question receives no straightforward scientific resolution, the common-sense functionalist can remain non-committal on the EM question without giving a dialectical advantage to the empirical functionalist. If neither side's preferred methodology provides a determinate answer to the EM question, a draw is all that can be claimed.

Chapter 2: A Theory of Content for S-Representation

Abstract:

The task of this chapter is quite ambitious. It is to create a reductive analysis of a particular kind of mental representation: s-representations. I aim to do so by employing functionalist machinery broadly similar to that used by the common-sense functionalists discussed in the previous chapter. My aim is to persuade the reader that s-representations have the key properties we ordinarily think of representations as having and that these representational properties can be reductively analysed.

§1: Introduction

I borrow the term 's-representation' from William Ramsey (2007), who in turn took it from Robert Cummins (1989, 1991).[31] According to both Ramsey and Cummins, s-representations are supposed to be distinctive in that their representational properties are closely tied (in ways we will discuss below) to isomorphisms that obtain between them and their representata. They are, in this respect, supposed to be analogous to ordinary maps. Let's take an example. There are certain isomorphisms (topological similarities) between a standard map of London and London itself. The nature of the isomorphism relations that obtain between the city and the map plausibly play a significant role in determining the map's representational properties (its accuracy conditions, for example). Whereas the representational properties of statements written in a natural language, for example, are not obviously isomorphism-dependent in a similar way. S-representations are supposed to be more like maps than linguistic statements in this respect. It is the aim of this chapter to cash out something like this basic idea and to say exactly what conditions the representational properties of s-representations are grounded in. As we will see, however, my analysis will depart significantly from previous attempts to explicate the notion of s-representation. But before moving to explication, I should

[31] The 'S' in s-representation stood for 'simulation' in Cummins' original treatment, but is used by Ramsey as ambiguous between 'simulation' and 'structural'. Ramsey uses this ambiguity to acknowledge his debt to both Cummins (1989) and Chris Swoyer (1991).

first say a little about what work the notion of s-representation, once explicated, is ultimately supposed to do.

The ultimate hope is that appeal to s-representations can do significant explanatory work; we can explain all sorts of otherwise puzzling psychological explananda by making appeal to s-representations. It will be the job of the next chapter to show that this hope is well founded. But for now, I should say something quite general about the kinds of explanatory demands I will and will not be making of s-representations. I will not claim that conscious states like perceptions or imaginings are examples of s-representations. Nor will I claim that intentional states like beliefs, desires and so on, which are accessible to consciousness, should be understood as instances of s-representation. I will do nothing to rule out such claims, but the claims I will be making about the role of s-representations in cognition is less ambitious.

The claim will be that we can explain certain basic cognitive phenomena by positing the existence of s-representations whose operations are inaccessible to consciousness and are not subject to volitional control. It is possible that more sophisticated, high-level cognitive capacities are explainable in s-representational terms, but I will not be arguing for the claim that they are. I will rather restrict myself to discussion of cognitive phenomena of the most basic kind. So my aim is not to provide a reductive analysis of complex mental states like beliefs and perceptions, but only to provide a reductive analysis of what I take to be one of the most basic kinds of mental representation: s-representation. Before embarking on this project, it is important first to get clear on the challenges such an analysis must meet.

§2: Challenges for a Reductive Analysis of Mental Representation

Ramsey distinguishes between two challenges that any analysis of any kind of mental representation must meet. The first is the 'job description challenge', which he describes as an account of 'what it is to *function as* a representation in a physical system' (2007, p.29). He distinguishes this from the 'theory of content challenge', which is the task of presenting the 'the set of physical or causal conditions that ground the content of a representation – the conditions that determine how a state

or structure comes to have intentional content in the first place' (2007, p.29). There is an aspect of this distinction I take to be right, and an aspect I take to be a pernicious legacy carried over from Cummins (1996), into Ramsey's account of s-representation. By exploring these two aspects, I hope not only to provide a clear list of challenges that my analysis must meet, but also to explain how my analysis will diverge from that given by Cummins (1996) in attempting to meet these challenges. In exploring the pernicious aspect, I use Ramsey's distinction as a foil for illustrating the root disagreement between myself and Cummins. So it matters little whether my treatment does justice the distinction as he intended it (which is lucky, since I fear that it might not). I want to stress my divergence from Cummins because his is the main rival theory of content for s-representation. I go to less significant lengths to distinguish my view from Ramsey's because, as he acknowledges, he does not attempt to give a comprehensive theory of content for s-representation (2007, p.92).[32]

Let's start with the aspect of Ramsey's distinction that I suspect of being a pernicious relic of Cummins' (1996) analysis. Ramsey (2007, p.29) talks about the challenge of showing how a system '*functions as* a representation in a physical system' (the job description challenge) as distinct from the challenge of giving 'the set of physical or causal conditions that ground the content of a representation' (the theory of content challenge). It looks the theory of content challenge is just the challenge of giving a reductive account of a representation's content. A representation says something about that which it represents, and one puzzle is to say what it is by virtue of which it can do so. The task is just that of analysing that which is representational in terms of that which is not representational. I'm happy with this description of the theory of content challenge, but I am unhappy with the way in which Ramsey seems to treat the job description challenge as distinct from it. His distinction, on one reading, presupposes the claim that the 'set of physical or causal conditions that ground the content of a representation' is something other than a set of facts about how the putative representation '*functions as* a

[32] Another prominent position similar to my own is that developed by Rick Grush (1997; 2003; 2007). I explore the similarities and differences between my theory and his in the next chapter. The upshot of my analysis will be that Grush's emulators count as s-representations, but that his account of what it is that makes his emulators genuine cases of content-bearing representation is inadequate and incomplete.

representation in some physical system'. To presume this is to claim that a representation's content is grounded in something other than its functional role. I'm not sure if Ramsey intends to make this presupposition, but it is clearly endorsed by Cummins (on whose notion of s-representation Ramsey builds). So, to avoid the perils of misinterpretation, I will focus in on Cummins, who couldn't endorse this presupposition more clearly. Cummins claims that, on any theory of representation, '*[t]he content of a representation must be independent of its use or functional role'* (1996, p.86). So, for Cummins at least, the story about how a system functions as a representation in a physical system is and must be separate from a story about what grounds the representational content of said system. I dissent.

According to Cummins, we can only account for representational error if we acknowledge the independence of a representation's content from its 'use' or 'functional role'. His basic view is that we get representational error when there is a mismatch between a representation's content and the manner in which the content is applied. But, the idea goes, if we make content dependent on use, we cannot make coherent sense of the idea of content *mis*applied. His positive proposal is that we should first distinguish between the content of a representation and its target. To reuse the map example, the map's target would be London, and its content would be the properties it assigns to London. As this example suggests, Cummins is using the term 'content' in a slightly unorthodox way. He is using the term to mean 'predicative content' – content that predicates properties of its target. This use of the term is narrower than is standard. Most would say that there is a meaningful sense of content on which the map of London has the city of London as its content. I will use the term 'singular content' for this kind of content. To have singular content is just to be about some target in the way that the map of London is about London, and the singular content of a representation is that which it is about. Whereas the predicative content of a representation is the properties that it assigns to its target. With this distinction in place we can see that Cummins' claim is actually weaker than I made it sound in the previous paragraph. His claim is that *predicative* content must be entirely independent of functional role.

This is still a claim from which I dissent, but let's see how Cummins deploys the idea. An s-representation's singular content is, for Cummins, determined by

something like function (or, more accurately, 'proper function' – more on which later). And its predicative content is determined by spatial isomorphisms. So, to use the map example, the map has London as its target because of the manner in which it is used (or is supposed to be used), but its predicative content is determined entirely by the spatial isomorphisms that obtain between the elements of the map and the elements of the city. With this framework in place, Cummins describes representational error as a mismatch between singular and predicative content – the inappropriate assignment of properties to a target (1996). I will not give a detailed critique of Cummins' positive proposal. Nor will I give a detailed critique or his argument to the effect that we cannot account for representational error if we do not treat predicative content as independent of use. Rather, I will endeavour to develop an existence proof against the latter claim. I do so by developing an account of s-representation that ties content, both predicative and singular, to functional role (and nothing other than functional role) but which still makes room for error. In other words, my analysis of s-representation will be a functionalist one. And will draw inspiration from Jackson and Braddon-Mitchell's (2007) functionalist analysis of mental states like beliefs and desires.

This brand of functionalism does not make appeal to anything like the teleological notion of 'proper function' found in teleosemantic accounts of representational content.[33] A system's proper function is, roughly, the function for which something is designed or evolved, or the function it is in some other sense *supposed* to perform. A system's proper function is supposed to be different from the causal role it is currently playing. My heart's proper function is to pump blood. And this is true even if I am having a heart-attack, and my heart is currently doing no such thing. The heart is supposed to pump blood, irrespective of whether or not it is doing so now. Proper function is normally introduced into theories of mental content in order to deal with representational error. The idea, roughly, is that a misrepresentation occurs when a representational system fails to perform its proper function (an erroneous representation is, roughly, a malfunctioning one). In what follows, I will try to account for s-representational error without making appeal to teleological notions like proper function. There is little point in my enumerating the

[33] See Millikan (1989, 1993, 2005), Macdonald and Papineau (2006) for more on such approaches.

virtues of this approach, since the proof of the pudding will be in the eating. But there is one point I should emphasise.

As mentioned, above, the main motivation for bringing proper function into an account of mental representation is the hope that the notion will help account for representational error. Indeed, there is a widespread sense in the teleosemantic camp that representational error cannot be accounted for without appeal to something like proper function. Given that I will be trying to do without such a notion, one would expect the problem of error to be the hardest problem for my account to deal with. And given this, it will come as no surprise that the account I offer is organised around this problem. But it is worth noting that, by refusing to make appeal to proper function, I am not just creating work for myself without motivation. The appeal to proper function brings with it all sorts of problems of its own.[34] And by making do without the notion, I am sidestepping these problems. I will not, for example, have to say anything about Swampman cases of the sort introduced by Davidson (1987). And, as I will argue in §7, my account deals more easily with the 'causal role of content problem', than those theories (like that offered by Cummins, and like most teleosemantic theories) according to which the content of a representation is entirely or partly independent of facts about its causal role.

Now to the aspect of Ramsey's two-challenge distinction I take to be useful. Suppose I want a story about what it is by virtue of which my map of London represents what it does. To provide such a story is to provide a theory of content for a certain sort of representation. Suppose I start with the eminently plausible assumption that the map is indeed a representation of London and does indeed predicate properties of London. I then proceed to work out exactly what it is that gives the representation the content that it possesses. To do this, I need a story about what grounds its singular content – the fact that it represents London rather than some other city (New York, say). I also need a story about predicative content – about what it is for the map to ascribe properties, accurately or inaccurately, to London. And I need to say what it is that gives the map the particular accuracy conditions that it has. I will say more about such questions of content in due course, but for now I want to focus on what was assumed (plausibly) before these questions

[34]See Macdonald and Papineau (2006) for a collection of recent exchanges on these problems.

got off the ground. We assumed from the beginning that the map was indeed a representational device, and then we thought about what it was that grounded its various representational properties. On one interpretation, the job description challenge is the challenge of justifying the assumption that some entity (my map, for example) really is a representation. By contrast, the theory of content challenge is that of giving a reductive story about what it is that grounds the entity's various representational properties.

I want to argue that the distinction between these two challenges, when construed in this way, is useful when it comes to thinking about mental representations. It is useful because it serves as a prophylactic against assuming prematurely that a given entity is a genuine representation. To see this, imagine that I start by assuming that the firing pin in a pistol is a representational device that represents the state of the pistol's trigger, and endeavour to create a theory of content for said firing pin. I might do so by appeal to causal covariance. The state of the firing pin represents the state of the trigger because it covaries causally with the state of the trigger: when the trigger is pulled, it causes the release of the firing pin, and when the trigger is not pulled it does not. So I might say that the release of the pin has the content 'trigger being pulled now' and its non-release has the content 'trigger not being pulled now'. Someone might at this point offer a counterexample to this analysis. Suppose that the gun is badly manufactured, meaning that the firing pin would also be released if the gun were dropped from a small height. In a situation where this happens, one might ask, does it follow that the firing pin's release now represents the state of being 'dropped now from a tall height' with which it now covaries causally? And, the point is pressed, if the answer to this is 'yes', then how can the firing pin misrepresent? It seems to represent any state by which it is caused, so it doesn't look capable of misrepresenting.

I might try to answer this kind of question by appeal to some notion of proper function: the firing pin is supposed to covary with the state of the trigger, but it's not supposed to covary with the gun's being dropped. Since the latter causal covariation is an instance of malfunction, it is also a case of misrepresentation. On this new theory, the firing pin's release represents all and only those states with which it is supposed to covary. An objector might continue to ask how exactly we

work out what a gun is 'supposed' to do in this context, prompting a whole host of standard objections and replies.[35] But an onlooker to this debate might legitimately ask what reason we have in the first place for thinking that the gun's firing pin is a genuinely representational device. After all, we could just describe it as a reliable causal mediator between the pulling of a trigger and the firing of a bullet. This, Ramsey claims, is intuitively a more natural way of describing the device, and it obviates the need for a theory of content for firing pins (2007, pp.136-8). This is just to say that the firing pin fails the job description challenge on the prophylactic reading of that challenge; it's not initially plausible to think of the firing pin as performing a genuinely representational function, so we should never have started engaging in the imagined project of building a theory of content for it in the first place.

So on this reading, the job description challenge is little more than an initial intuitive test. If some entity's functional role is most naturally described in non-representational terms, don't bother trying to devise a theory of content for it. If the entity in question looks very much like a representation (and hence does pass this test), then we have good (though defeasible) reason for thinking there must be some way of accounting for its representational properties (which is the purview of the theory of content). Ramsey argues that many notions of mental representation in the literature fail this intuitive test. Indeed, the firing pin example was designed by Ramsey (2007, pp.136-8) as an illustration of the manner in which Dretske's (1988) notion of representation falls short of the mark. I agree with Ramsey's assessment of Dretske's view, but nothing in what follows hangs on this assessment, so I will not rehearse the arguments for it here. For now, the only point I want to argue for is procedural. Before launching into an attempt to provide a theory of content for some entity, think first about whether it is really plausible to think of such said entity as a genuinely representational device. Once this assumption is made, all sorts of ingenious moves can be made in an attempt to devise a theory of content for said entity, but if that assumption is unjustified, then such ingenuity is misspent.

Of course, we can imagine a case in which something initially looks very much like a representation, but where repeated attempts to provide a theory of

[35] See Macdonald and Papineau (2006) for a sense of how this debate about how to cash out 'design' might progress.

content for it seem to have failed. In such cases, we might want to revisit our initial intuitive assessment – the fact that an account of its representational properties seems elusive might tempt us to think that maybe it's not a genuine case of representation after all. If, after many attempts, we can't make theoretical sense of the idea that some entity or class of entities have truth/accuracy conditions, that may in some cases be good reason for thinking that the entity or class of entities don't have such conditions after all.[36] So our initial intuitive assessment isn't infallible, but it still seems very much like a good place to start. And so, when I introduce s-representations, my first task will be to show that they pass this intuitive test of representational status.

To summarise, the challenges my analysis of s-representation should meet are as follows. First is the job description challenge, which is that of showing that s-representations are plausibly treated as representations in the first place. Then there is the theory of content challenge, which is that of showing how s-representations' representational properties can be accounted for in non-representational terms. This is the major challenge that any theory of representation must face. And as we will see, there is a set of set of standard problems that any theory of content should solve. My main strategy for defending the theory of content I propose will be to show that it has the resources necessary for dealing with these standard problems. Unlike Cummins, I will do so by appeal to functional properties alone. And unlike writers indebted to the teleosemantic tradition, I will do so without appealing to the notion of teleology in building the account.[37] One final thing to note is that on the account of s-representations I will offer, predicative and singular content do not come apart in the way they do on Cummins' theory. It is plausible to think that linguistic representations can have singular content without predicative content. The word 'Hillary', for example, might be thought to have a target (or a range of targets) without predicating any properties of that target. But s-representations, on

[36] Though in some cases, such failure would clearly not warrant this conclusion. Even if many attempts to give an analysis of pictorial representation ended in utter failure, it would be strange to conclude from this failure that pictures don't represent after all.

[37] In this respect, and in others besides, I deviate from Dan Ryder (2004) who also attempts to give an isomorphism-based theory of mental content. Ryder also appeals to the notion of proper function in his theory of representation, so that is another respect in which my account differs from his.

the account I will be offering, cannot have predicative content without singular content (or vice versa). The two come as a package, or not at all.[38]

§3: Three Types of Representation

Ramsey prepares the ground for his introduction of s-representation by making a threefold distinction between different types of non-mental representation.[39] The distinction is as follows:

1. Icons: these are 'connected to their object by virtue of some sort of structural similarity or isomorphism between the representation and its object' (2007, p.21)
2. Indices: these signs 'designate things or conditions by virtue of some sort of causal or law-like relation between the two. (2007, p.21)
3. Symbols: 'symbols are connected to their objects entirely by convention' (2007, p.21)

Some examples will serve to illustrate the differences between these sign-types. Maps, pictures and diagrams would be examples of icons. Take maps as an example. They represent (at least in part) by virtue of the spatial isomorphisms that obtain between them and the areas they represent.[40] Examples of indices might be smoke indicating fire. Fire causes smoke, so smoke can be used as a relatively reliable indication of the presence of fire. Linguistic tokens are examples of symbols – the word 'dog' refers to a particular type of mammal by virtue of certain linguistic conventions.[41] S-representations are supposed to be mental representations that represent in the way that icons represent. That is they are representations that represent at least in part by virtue of certain isomorphisms between them and the objects that they represent.

I will retain this basic idea, while developing it in a distinctive direction. But it is worth pointing out that things are not as simple, even in the non-mental case, as

[38] Jackson and Braddon-Mitchell (2007), following Lewis (1994), claim that the content of mental states like beliefs is similarly non-decomposable.
[39] Ramsey attributes this threefold distinction to C.S. Pierce (1931-58), while acknowledging that he is no Pierce scholar. I avoid the task of interpreting Pierce's notoriously complex semiotics, but it is worth noting that that Pierce distinguishes a grand total of sixty six classes of signs (1998, p.481). Thanks to Josh Black for pointing this out.
[40] Or at least, on many theories, such isomorphisms play a crucial role in grounding the representational status of such representations.
[41] Note that this claim does not have to rely on the notion that such conventions were arrived at by explicit agreement. See David Lewis (1969) for a classical treatment of this issue.

Ramsey's distinction might suggest. Take the idea that language represents by virtue of convention. There is an obvious sense in which this is true. The word 'dog', for example only represents dogs because there is a convention in place according to which it does so. But language, on many stories, represents not merely by virtue of convention, but by being structurally isomorphic with the world it represents. Swoyer, for example, claims that it is the parallels between the structure of the world and the logical structure of language that allows us to use language to make inferences about the world, and to describe relations that obtain in the world (1991, p.492-3). If this is right, much of linguistic representation exploits isomorphism as well as convention. Similarly, it seems plausible to think that maps represent not merely by virtue of isomorphism, but by virtue of certain conventions (the conventions of cartography). It may very well be that there is a tight link between the choice of cartographical conventions and the demand for isomorphism, but we should recognise that the picture is quite a bit more complex than Ramsey's distinction at first suggests.

But it is possible to maintain that while language, for example, does exploit isomorphism in all sorts of complex ways, its representational status isn't tied inextricably to isomorphism. I do not intend to spend time defending this claim in the case of language, or to depend on it in what follows. Rather, I want to stipulate that, s-representational status *is* tied inextricably to isomorphism. Isomorphism of a special kind (which I specify below) is partly constitutive of s-representation, or so I will claim. So I want to agree with Ramsey that there is an important sense in which the notion of s-representation is tied to isomorphism. I also follow Ramsey in thinking that the best way of understanding s-representations is to think of them by analogy with non-mental representations. If we want to argue that s-representations intuitively look like genuine representations (and hence pass the job description challenge), a good way to do so is to demonstrate that they are closely analogous to something whose representational status is uncontroversial. Non-mental representations serve as useful analogues for this kind of argument. Everyone (I would hope) agrees that maps, for example are genuine representations, and everyone has at least an intuitive sense of how maps are used (we all know that we use them to find our way around, for example). So if we can show that s-representations are relevantly similar to things like maps, we have *prima facie*

grounds for thinking they are genuine representers. As we will see, I think s-representations should be thought of by analogy not with maps, but rather scientific models, or simulations (which, I will argue, also represent by virtue of isomorphisms of a certain type). But the methodological point still stands.

§4: S-Representations Introduced

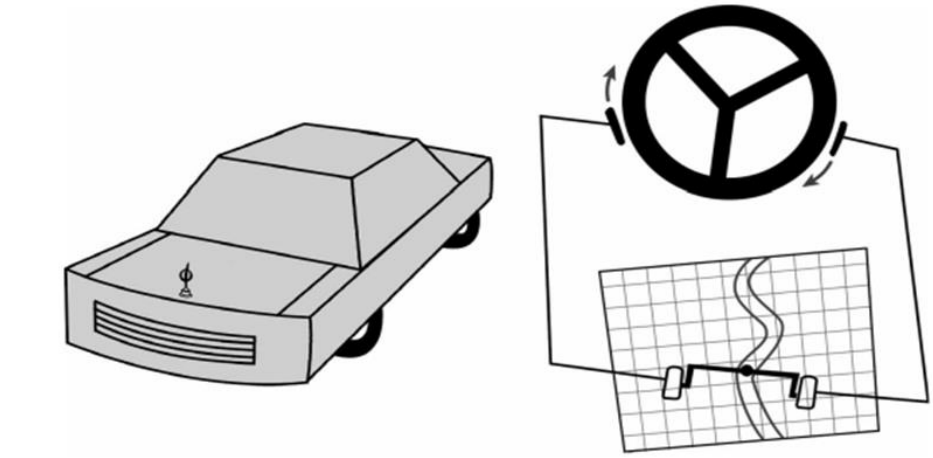
With these preliminaries out of the way, it's time to introduce s-representations. The easiest way of doing so is to give an example of s-representation. I will now introduce a simple, toy case, developed by Cummins (1996) and Ramsey (2007), but I hope to generate philosophically interesting results by reflecting on it. And in the next chapter, I will argue that these results generalise to biologically plausible instantiations of s-representation. First, though, let's start with a toy case.

Imagine a driverless car which is able to navigate through an s-shaped tunnel without hitting the tunnel's walls. It does so by exploiting a rudder and groove system (see fig.1 below).[42] The car's steering system is hooked up to a rudder. And as the car moves through the s-shaped tunnel, the rudder moves through a similarly s-shaped groove. If left to its own devices, the rudder would just move forward in a straight line. But as it moves through the s-shaped groove, the groove's walls push it to the right, to the left, and to the right again, such that the rudder follows an s-shaped path through the groove. The rudder is hooked up to the car's steering system in such a way that when the rudder moves to the right, the car steers to the right. And when the rudder moves to the left, the car also moves to the left. The rudder, which is forced by the groove to describe an s-shaped path, in turn causes the car to describe an s-shaped path.

We can see that a steering system of this kind, if calibrated correctly, could enable the car to navigate through the s-shaped tunnel without hitting the tunnel's walls. In order to get the right calibration, we would have to set up the rudder-groove mechanism in such a way that the rudder interacts with the groove in much the same way as the car would interact with the tunnel, were the steering mechanism absent. To do this, we could maybe start by making sure that the s-shaped groove is the same shape as the s-shaped tunnel, and making sure that the

[42] Diagram taken from Ramsey (2007, p.199)

Fig. 1



size of the rudder relative to the groove is similar to the size of the car relative to the tunnel. If we did this, and ensured that the rudder moved through the groove at the same speed as that at which the car moves through the tunnel, we would be almost there. The last step would be to ensure that the rudder starts its journey through the groove just before the car starts its journey through the tunnel. This would allow the rudder to adjust the car's steering just before the car crashed into a given tunnel wall.

In this imagined system, the rudder interacts causally with the groove in much the same way as the car would interact with the tunnel, in the absence of this predictive steering system; if the car lacked a rudder and groove system, it would bounce through the tunnel in the manner of a bumper car – being shunted from one side to the other. But because the rudder's position relative to the groove is always just ahead of the car' s position relative to the curve, it is always the rudder that gets shunted back and forth in this manner; the rudder interacts causally with the groove so that the car doesn't have to interact causally with the walls of the tunnel. Cummins (1996) and Ramsey (2007) consider this to be a genuine case of representation. The idea is that the car is using the rudder and groove system as a representation of its terrain and, by doing so, it is able to navigate successfully. Ramsey expresses his intuition about the case as follows:

> [I]n this case, internal structures can still serve as surrogates (as stand-ins), even if we drop the involvement of an inferring and learning mind. This is because the surrogative problem solving can be automated. A mindless organism can still take advantage of the structural isomorphism between

> internal structures and the world, and in so doing, employ elements of those internal structures as representations-qua-stand-ins. (2007, p.200)

There are two important ideas I want to draw from this. First is the intuitive idea that the structural isomorphism between the rudder-groove system and the putatively represented target allows the rudder-groove system to act as a 'surrogate' for the represented target. This idea, which Ramsey borrows from Swoyer (1991), is one that I will try to accommodate in my analysis of s-representation. Although the notion of surrogacy may sound teleological, I will be sure to cash out the notion of surrogacy without smuggling anything teleological into the picture. The second point is Ramsey's insistence that the rudder-groove system can represent despite the fact that there is no 'inferring and learning mind' interpreting the system and drawing conclusions from it. A natural thought about maps and scientific models is that they can only represent because there are minded individuals who can interpret them as representing some target. The basic idea behind s-representation is to find a mental representation that represents in much the way that maps or models do, but to claim that they can do so despite the fact that they are not interpreted by some outside observer, or homunculus within the system.

There is a philosophical tradition that denies the possibility of representation, even mental representation, without such interpreters. A natural worry for such a view is that it leads to an explanatory regress. Mental representations are the very entities that theorists posit in order to explain complex cognitive capacities like interpretation. And if we need to posit interpreters in our explanation of cognitive capacities like interpretation, then we quickly find ourselves faced with an explanatory regress. Each mental representation needs an interpreter, but it's difficult to see how we can explain the interpreter's ability to interpret without positing yet more mental representations. 'Homuncular' functionalism is developed in part to explain how the regress can be avoided.[43] The idea that s-representations are genuine representations assumes the claim that you *can* have mental representation without anything close to being an interpreter. In what follows, I will be assuming that something like this claim is true. There is a tenuous sense in which you could describe the car in which the rudder-groove mechanism is embedded as using the rudder-groove mechanism for guidance (and hence interpreting it), but

[43] For a seminal defence of homuncular functionalism, see Lycan (1982)

this is a bit of a stretch.[44] I will argue that the rudder-groove system is only a representation because it is used in a particular way (because it is used as a surrogate), but I will not try to make the leap from this claim to the claim that the mechanism is interpreted. The hope is that the rudder-groove system described above seems sufficiently map-like or model-like in the way it is used to give us at least a defeasible confidence that it might be a genuine case of representation (even in the absence of an interpreter).

Unlike the firing pin discussed earlier, it doesn't take a perverse act of intellectual contrivance to view the rudder-groove system as a representational device. According to Ramsey, it takes contrivance *not* to see the rudder-groove system as representational (2007, p.200). But that is a stronger claim than I need. I need only the claim that the system looks sufficiently close to being representational to justify the endeavour of attempting to develop a theory of content for it. If such a theory of content cannot be found after significant effort, we might want to roll back on the claim that it is a genuinely representational system. But for now, the fact that it seems to be acting as a surrogate in something like the way that a map or a model does is enough to get the ball rolling. There is little I can say to the sceptic who maintains that representation is always interpreter-dependent and who therefore insists that s-representations are no representations at all. But then it would be absurdly ambitious to expect the story here told to convince everyone.

§4.1: Tweaking the Rudder-Groove Case

Before moving on to develop a theory of content for s-representation, I want to tweak the toy example just a little, in order to bring it closer into line with the manner in which s-representations might actually be implemented in human and animal brains. The tweak is not intended to make the rudder-grooves system genuinely biologically plausible, but is rather intended to show how an s-representational system might arise from a set of interactions with sensory input. Imagine that instead of moving through an s-shaped curve, the car's rudder now

[44] Grush (1997, p.6) claims that his emulator representations are user-relative, but seems to use 'use' in such an undemanding way that the car would count as a user of the rudder-groove system. I don't see this as a particularly helpful way of framing things, but the disagreement here looks to be terminological rather than substantive: if the car counts as a user, then I agree that the rudder-groove's representational status is user relative. In the next chapter, I will discuss Grush's emulation theory in greater detail.

moves over a rectangular, putty-like landscape that is (initially) flat. The rudder takes the path of least resistance through this landscape, so will start by moving in a straight line through it. But the car's front corners have now been fitted with pressure-sensitive pads. Each of these pads is hooked up to a rudder-shunter. This means that when the car hits into a wall, and the pressure pad is pushed, a mechanism will give a shunt to the rudder. When the car collides with the left-hand tunnel wall, this causes the rudder to be shunted to the right. When the collision is with the right-hand tunnel wall, the rudder is shunted to the left. Suppose that the shunts received by the rudder are similar in magnitude to the shunts received by the pressure pads. This means that when the car crashes hard into the left-hand tunnel wall, the rudder receives a hard shunt to the right. As the car bumps its way through the tunnel, the combination of shunts to the rudder causes the rudder to follow an s-shaped path through the putty-like landscape. And because the rudder is still hooked up to the steering system, these shunts still ultimately lead to steering adjustments.

So far, the system described sounds just like a slightly over-complicated bumper car, being shunted this way and that. The rudder system isn't guiding the car to follow an s-shaped path through the tunnel. Rather, the car's pressure pad system is causing the rudder to describe an s-shaped path through the putty landscape. But imagine that each time the rudder moves in an s-shaped path over the putty landscape, it makes a small, lasting indentation. Imagine also that each time the car finishes its journey through the tunnel, it is taken back to the start, has its rudder set back to the initial position, and is set off again. Over many repetitions, this indentation slowly turns into a groove. As the groove deepens, the rudder (which takes the path of least resistance) becomes more and more inclined to move in an s-shaped path through the landscape. As this happens, the car collides less and less violently with the tunnel walls, and the shunts applied to the rudder become correspondingly less violent. Eventually, the groove comes to be just like the one described in the original case. At this point, the car's pressure sensors become redundant, since so long as the tunnel through which the car is moving remains the same shape, there will be no more collisions.

The resulting car may seem identical to the original one, but it is importantly different. To see why, imagine what would happen to this new groove-landscape if we were slowly to change the shape of the tunnel through which it travelled from an s-shape to some other shape. The car would start to collide with the tunnel-walls again, and these collisions would cause shunts to the rudder. These shunts, if sufficiently violent, would force the rudder to follow a slightly different path. And in doing so, they would lead to a situation where the rudder gradually alters the putty-landscape through which it travels. And if we imagine also that the putty is in fact a slow-moving slime which eventually fills in any indentation that doesn't have a rudder moving over through it on a regular basis, a change in the rudder's path through the landscape would eventually result in the landscape reforming itself to match the new shape of the tunnel through which the car travels. In short, we now have a car capable, over a long enough time-frame, of creating s-representational models of a whole range of different tunnels. Notice also that as the matching between the rudder-groove mechanism and the outside environment improves, there is a decrease in the amount of 'input' the system receives (where input takes the form of feedback from the pressure sensors). I will try to render this basic idea more precise later on, and argue for its significance, but for now I want only to draw attention to it before moving on to develop a theory of content for s-representation.

§5: What it Takes to be an S-Representation

I want to claim that the s-representational status of the rudder-groove mechanism is secured entirely by facts about the functional role it plays in the system in which it is embedded. And, on the story I will be telling, to perform a certain functional role is just to be disposed to exhibit certain causal dynamics under certain conditions. I will try to make this basic idea compatible with the idea that an s-representational mechanism represents by being isomorphic with, and acting as a surrogate for, that which it represents. In this section, I'll formulate the conditions a system must satisfy if it is to count as an s-representation. In later sections, I will show how my account of s-representation deals with s-representational content (singular and predicative) and how it deals with standard problems, like the problem of misrepresentation.

Let's start with isomorphism. There are two different types of isomorphism I could appeal to in an attempt to ground the representational status of the rudder-groove mechanism. The first is a kind of spatial isomorphism. Roughly speaking, the shape of the groove is similar to the shape of the tunnel. There is an isomorphism in the spatial relations in which elements of the tunnel stand and the spatial relations that obtain in the s-shaped groove. This is the kind of isomorphism Cummins appeals to when accounting for the predicative content of the rudder-groove mechanism.[45] Spatial isomorphism of this kind is independent of functional role. It doesn't matter how the s-shaped groove is used, the spatial isomorphism between it and the tunnel still obtains. So if predicative content is determined only by spatial isomorphism, the predicative content of the groove is entirely independent of the manner in which it is used. Cummins dramatizes this point by claiming that turning the groove upside down 'does not change its representational content, any more than rotating a map changes its representational content' (1996, p.99).

In contrast to Cummins, I want to account for s-representation by appeal to a type of isomorphism that does not come apart from functional role. This is isomorphism of functional organisation. To get an intuitive idea of what this kind of isomorphism amounts to, imagine that I buy identical computers, each with identical software installed. Imagine that I then turn both computers on simultaneously. By doing so, I set in motion two functionally identical processes. The causal chain initiated in one computer is identical to that initiated in the other. This is functional identity. Functional isomorphism is just functional similarity rather than functional identity. There must be similarity between one process (or causal chain) and another if they are to be called functionally isomorphic. That is a first approximation of the basic idea, at least.

Functional isomorphism of this kind is normally that by virtue of which a certain class of models represent. For example, suppose I model the behaviour of an aeroplane tail in a wind tunnel (an example borrowed from Swoyer (1991)). The interaction between the wind and the tail is, if the model is a good one, isomorphic to the interaction that would occur between the tail and ambient air, were the tail

[45] It is actually a card-slit mechanism in his example, but this difference is immaterial.

attached to an aeroplane that was flying through the air. The interactions might be similar in that, in both cases, similar patterns of turbulence are created, and similar forces, stresses and so are exerted on the tail. This host of functional isomorphisms allows the engineers who created the experiment to determine whether tails relevantly similar to the one used in the model are fit for purpose, for example.

And notice that it is no accident that these functional isomorphisms obtain on a particular occasion. It is because the computers are put together in a certain way that they are disposed to behave similarly under relevantly similar circumstances. And something similar goes for the aeroplane tail case as well. The relevant systems are disposed to behave in an isomorphic manner in relevantly similar circumstances. The important point here is that the relevant isomorphisms are counterfactual-supporting. It's not just that the two computers happen to behave similarly under similar conditions – they are also disposed to do so. Something is fragile because it is disposed to smash under certain conditions, not merely because there is some actual situation in which it happens to smash. It is difficult to say exactly what it is by virtue of which a given entity has a given disposition, and I will not attempt to answer such questions here (although see §11 for a brief discussion of the relation between a disposition and its categorical basis). So I will not say anything about what two systems must have in common in order to be disposed to behave in a functionally isomorphic manner. But I will claim that to count as an s-representing system, a system must be disposed to behave in a manner that is functionally isomorphic with the behaviour of that which it represents. And I will give a more precise account of what this claim amounts to in due course.

But first I should make some general comments about the notion of disposition I will be employing. The notion I will be using is one that tolerates exceptions. Such usage is not unusual.[46] The weather in Britain is disposed to be bad even though it is, occasionally, quite pleasant. And a computer is disposed to behave in a certain way under a certain set of conditions even though it occasionally freezes for no apparent reason. Similarly, I will be claiming, s-representations count

[46] See Fara (2005) for an account of this sort, where exception-tolerant dispositions are called 'habituals'. As noted in the last chapter, Lewis (1980) and Jackson and Braddon-Mitchell (2007) analyse mental states in terms of the functional roles they typically play. Such an analysis also makes use of the idea of dispositions that tolerate exceptions – that was the point of the *ceteris paribus* clauses with which their analyses were qualified.

as representations if their behaviour is disposed to be isomorphic with that of their target, even if there are occasions when this isomorphism breaks down temporarily. This toleration of temporary breakdowns will play a crucial role in my account of misrepresentation. The second general point is that there is nothing teleological about the notion of disposition I will be employing. Two systems can be disposed to behave in an isomorphic manner even if this is not their goal. Indeed, two systems can be isomorphic with respect to functional dispositions even if neither have any goal, or design, in any sense at all. Imagine two identical solar systems at opposite ends of the universe. They were not designed to be the way they are – they are just the product of cosmic fluke. Yet they are still disposed to behave in a similar manner under similar circumstances.

I now need to say something about the notion of surrogacy I will be employing in my account of s-representation. As the computer example above shows, the disposition to behave in a manner that is functionally isomorphic with some state of affairs it not sufficient for representing that state of affairs. Two computers at either end of the world might be functionally isomorphic with each other, but that does not, in and of itself, make either computer a representation of the other. But if we look again at the aeroplane tail in a wind tunnel, we can get a sense of what must be added to the picture in order to get model-like representation. In this case, the modelling system's disposition to behave in a manner that is functionally isomorphic to the state of affairs of which it is a model is exploited in what Ramsey calls a process of 'surrogative problem solving' (2007, p. 96). Scientists use the model as a surrogate for a non-actual (but possible) set of circumstances, in order to draw conclusions about how a real aeroplane tail would behave in those circumstances. Obviously, there are no scientists in the s-representation story, so we have to find some way of automating this surrogacy relation.

I will argue that an s-representation acts as a surrogate for the non-actual state of affairs with which it is isomorphic when it is set up in such a way as to prevent that non-actual state of affairs from becoming actual. The basic idea will be that the s-representational mechanism is disposed to anticipate, and thereby prevent, a certain class of interactions. And to be so disposed is to act as a surrogate for those

anticipated events. All this will be explained below, but it is worth mentioning two things here. First, the surrogacy condition, like the isomorphism condition, will be cashed out as a disposition manifested by s-representational systems. Second, the notion of surrogacy in play will be quite a specific one. It is quite possible for something to act as a surrogate for some state of affairs without preventing that state of affairs from becoming actual. But on my use of 'surrogacy', to act as a surrogate for some state of affairs will be to anticipate, and thereby prevent that state of affairs. While this usage is narrower than the common usage, I hope the surrogacy relations I go on to describe are still recognisably surrogative.

All this is best illustrated by the rudder and groove case.[47] Here is a first approximation of the functional isomorphism by virtue of which it represents:

> *The rudder-groove mechanism is disposed to generate behaviour similar to that which the car-tunnel system would be disposed to generate if the car did not have the rudder-groove mechanism embedded within it.*

The basic idea is that the rudder interacts with the groove in much the same way that the car would interact with the tunnel if the car lacked a steering system. The rudder bashes its way through the groove in much the same way as the car would, in the absence of the groove. And, by doing so, it prevents the occurrence of any actual interactions between the car and the tunnel.

To get a more precise view of how this works, let's focus on one single interaction generated by the rudder-groove mechanism. Let's imagine that the car is just setting off into the tunnel and the rudder is, correspondingly, just setting off into the (fully formed) groove. If the car lacked a rudder-groove steering mechanism, or any other steering mechanism for that matter, it would keep moving forward until it eventually collided with the tunnel wall. But this is not what happens. Instead, before this collision occurs, the rudder collides with the wall of the groove. And this collision causes the car to steer away from the tunnel wall, thereby preventing an interaction between the car and the tunnel wall. Let's call the actual interaction that occurs between the rudder and groove 'I_A' and refer to as 'I_C' the 'counterfactual'

[47] Unless specified otherwise, I will be referring to my augmented version of the rudder-groove mechanism in what follows. It is from this case that I want to draw general results (although I do not deny that most of the results I draw could also have been drawn from the more simple case).

interaction that would have occurred between the car and tunnel, had I_A not occurred. Prior to I_A, the causal dynamics of the rudder groove system were recognisably similar to those of the car-tunnel system: in each case we had something moving into a similar landscape with a similar trajectory. Then I_A occurred, changing the causal dynamics of the rudder-groove system (i.e. changing the trajectory with which the rudder moved through the groove). Let's use the term 'state-changing interaction' to refer to an interaction that changes causal dynamics of a system. We can then say that the state-changing interaction that actually occurred between components of the rudder-groove system prevented a state-changing interaction between components of the car-tunnel system.

Now the prevented interaction (that which would have occurred between the car and the tunnel wall) may not have been identical with that which did in fact occur between the rudder and the groove. While the rudder is nudged on to a new trajectory, the car might have simply crumpled to a halt. On the story I'm telling, this doesn't matter (for reasons that will become clear in what follows). There is still at least a *minimal similarity* between I_A and I_c: we start with two systems with similar causal dynamics, and in both systems, these initial dynamics would lead to a state-changing interaction.[48] What's more, it is the occurrence of I_A in the rudder-groove system that prevents the occurrence of I_c in the car-tunnel system. Let's put what we have so far into a set of conditions:

(1) There is a minimal similarity between I_A and I_C.
(2) If I_A had not occurred, I_C would have. But because I_A did occur, I_C did not.

I want to generalise to the idea that a system disposed to generate interactions satisfying these two conditions is a representational system. And it is a representational system because it is disposed to (1) generate interactions that anticipate interactions which would occur in some represented system and is (2) is hooked up to a control system in such a way as to prevent the anticipated interactions from taking place. To get this result, I will need to formulate a third criterion (a disposition criterion), but I first need to define I_A and I_C at a suitable level of generality.

[48] Although in the case of I_c, this interaction is in fact never actualised.

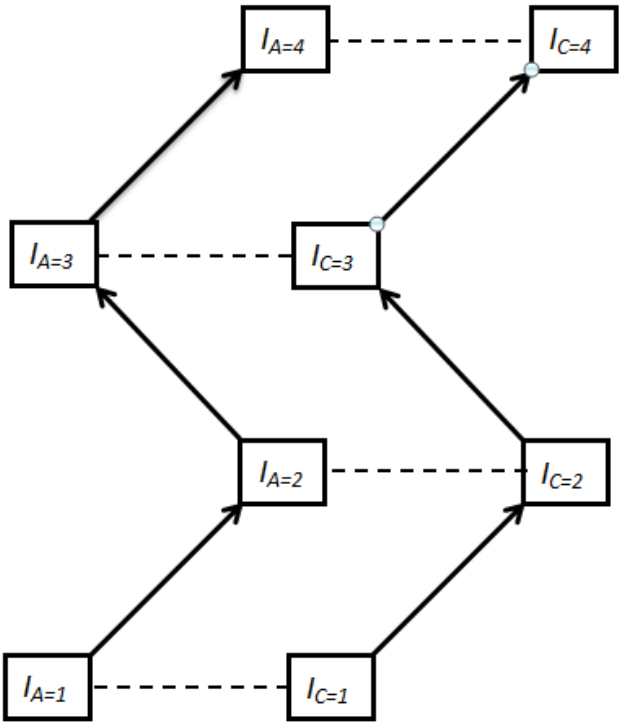
Above I used 'I_A' to refer to an interaction between the rudder and the groove, and used 'I_C' to refer to an interaction that would have obtained between the car and tunnel. But more generally, we can think of I_A as a state-changing interaction that occurs between components of a control system, where a control system is just a system disposed to control the behaviour of some controlled mechanism.[49] In our example case, the rudder-groove system is a control system and the car is the controlled mechanism, because the rudder-groove system is disposed to control the car's behaviour. The particular I_A we considered was a particular state-changing interaction between components of the rudder-groove control system. We can then think of I_C as a state-changing interaction between the controlled mechanism (the car in our case) and its environment (the tunnel in our case). In our case, this state-changing interaction was a collision between car and tunnel-wall that altered the car's trajectory relative to the tunnel.

Now to the disposition condition. Satisfying conditions (1) and (2) is not sufficient for s-representation. In order to count as an s-representation, a system must be *disposed* to generate interactions that satisfy these two conditions. We can put this requirement as follows:

> (3) Control system S is disposed to generate interactions (I_As) satisfying conditions (1) and (2) for some counterfactual interaction (I_C)

When a control system is so disposed, it counts as an s-representation, or so I claim. Let's first see what it means to say that the rudder-groove mechanism is so disposed. So far, I have made sense of the idea that the first interaction between the rudder and the groove satisfies conditions (1) and (2). But this is not the only such interaction that satisfies these conditions. As the rudder moves through the groove, it is constantly bumping against the groove wall, and thereby preventing the car from bumping against the tunnel wall. In other words, the system is constantly generating interactions that satisfy conditions (1) and (2). We can depict this situation as follows:

[49] And by 'control' here, I mean nothing more than 'have a causal impact upon'. Obviously an s-representing mechanism might have causal impact on all sorts of things, but its impact on all these things will not be such as to satisfy conditions (1) and (2). So it may 'control' lots of things, but there will be special class of things that it controls in a special way, where this specialness is captured by the two conditions. When I talk in what follows of 'action guiding' I again mean nothing more than 'controlling' as here defined.



On the left-hand side is a crude depiction of the state-changing interactions through which an s-representational system like the rudder-groove mechanism passes. The boxes represent individual state-transitions which satisfy conditions (1) and (2), and the dotted line depicts the correspondence between an interaction in one domain and an interaction in another. $I_{A=1}$ is the first of such transitions, but it is followed by others. If this diagram were entirely accurate, it would mean that only four of all the state-transitions underwent by the rudder-groove mechanism satisfied conditions (1) and (2). On the assumption that the rudder groove mechanism moves through many hundreds of different states as the rudder passes all the way through the groove, this would constitute a pretty small proportion of transitions satisfying (1) and (2). But imagine that the rudder-groove system was perfectly calibrated. In this scenario, every state-changing interaction satisfies the two conditions. This system would be perfectly disposed to satisfy conditions (1) and (2).

But imagine that there were some imperfections in the rudder-groove mechanism. For example, imagine that there was the odd kink in the groove. Such kinks might throw the rudder a little off course, and might thereby even cause a minor collision between the car and the tunnel wall. The state-transitions that occur as the rudder passes through these kinks would no longer satisfy (1) and (2). But it would be a strange to say that the rudder-groove system as a whole is no longer

disposed to generate interactions that satisfy those conditions just because a couple of such kinks existed. The right thing to say is that the system as a whole is disposed to generate behaviour (causal interactions) that satisfy these conditions, but that it is not perfectly disposed to do so. At any rate, the notion of disposition employed in (3) is such that a system can satisfy it even if it does not manifest the disposition perfectly. This is not to say that anything goes, however. If we introduced enough kinks into the s-shaped groove, there would come a point at which it would no longer make sense to say that the system has the disposition picked out by (3). And when this point comes, on the account I am developing, it will at this point no longer make sense to say that the system counts as an s-representation. I have nothing interesting to say about exactly where we should draw the line for satisfying (3), but nor do I need to. All I need is the claim that there is a meaningful distinction between cases that definitely do pass the condition and those that definitely do not. And the existence of borderline cases does not threaten this distinction.

But why, one might ask, should we think of a system that passes condition (3) as a representational system? One way of answering this question is to show that we can think of such systems as representational while giving a decent answer to the standard questions a theory of representation must face. I will attempt to do that in what follows. But first it's worth exploring the intuitive idea behind the proposal. The intuitive idea is that a system satisfying (3) is disposed to produce behaviour that is functionally isomorphic with the behaviour of some target system and, by doing so, is also disposed to act as a surrogate for that system. The notion of functional isomorphism here invoked is much stronger than the weak 'minimal similarity' invoked in condition (1). The individual interactions generated by the rudder-groove system are only similar in a very weak sense to those that would occur in the putatively s-represented system. But the behavioural dispositions of the system as a whole are isomorphic in a much deeper sense with those of the system it putatively s-represents. It doesn't just produce a one-off interaction that is similar to some interaction in the s-represented system; it consistently produces interactions that exhibit such similarity. And, what's more, the system doesn't just happen to produce such interactions: it is disposed to do so.

To see what all this adds to the picture, let's consider first an interaction that satisfies conditions (1) and (2), when (3) is not satisfied. Suppose we have a car moving into an s-shaped tunnel. It is equipped with a rudder and putty landscape mechanism, but the putty landscape is yet to be formed into the appropriate s-shaped groove (it is just the flat landscape with which we started in §4.1). So it looks like the car is going to bash into the tunnel-wall. Fortuitously though, a loose car-component that has been rolling around in the vicinity of the putty landscape happens to knock into the rudder at just the right moment, and with just the right force to cause the car to steer away from the tunnel-wall just in time. Imagine a situation of this sort in which conditions (1) and (2) happen to be satisfied. The interaction that occurs is appropriately similar to that which would have occurred in its absence, and the fact that it does occur prevents the relevant car-tunnel interaction. It would be strange to say that this situation was one in which the imminent car-tunnel collision was *represented* and thereby avoided; the more natural thing to think is just that the collision was avoided by happy accident.

Now suppose that we have the same fortuitous set up, but increase the number of happy accidents. The loose part rolls around and, by sheer fluke, again and again bounces into the rudder at just the right moment to prevent collisions between the car and the tunnel. Imagine that this fluke chain of events creates a huge number of interactions satisfying (1) an (2). Still, the car does not pass condition (3), because it is not *disposed* to generate interactions satisfying (1) and (2). It happens, by happy accident, to do so multiple times in a row, but this is insufficient for disposition. The sense of disposition I am using requires not just some behaviour happens to be produced on a given occasion, but that it would do again, under relevantly similar circumstances. I will not attempt the hard task of providing an analysis of exactly what these relevant circumstances are, but the intuitive difference between this lucky chain of events and that produced by the rudder-groove mechanism is easy to see. It is not just a happy accident that the rudder-groove mechanism produces the behaviour that it does; it is set up in a manner that ensures that it would produce similar dispositions in similar

circumstances.[50] Again, while it would seem strange to think that the chain of lucky events imagined above would get us anything like representation, the claim seems less strange in the case where the chain of events is a manifestation of a counterfactual-supporting disposition of a system

I finally want to say something about the sense I have given to the claim that an s-representation must be disposed to act as a surrogate for that which it represents. The notion is a little odd on first inspection. It is normal to think of scientific models as acting as surrogates for that which they represent, but they don't need to prevent the events they anticipate in order to represent them. If the model aeroplane tail acts as the engineers had hoped, for example, the model may in fact make it the case that the interactions it represents become actual. Or they may not have any impact on what comes to pass, if the engineers were just running the model out of interest, to see what might happen in a particular case. But what I needed was a surrogacy notion that could be automated. And the notion of preventative surrogacy fits this bill. The basic idea is that instead of using input from the world in order to guide its behaviour, the s-representational mechanism, by modelling the relevant causal dynamics, anticipates what this input would be. And it is this anticipated input, rather than input from the world, that acts to guide action. As a result, the s-representing system acts to avoid a situation in which the mechanism whose action it is guiding receives input. And the guided mechanism only receives input from its environment when the s-representing system's disposition to satisfy conditions (1) and (2) is not manifested.

§6: Singular and Predicative Content

So far I have given a rough sense of what it takes for a system to count as an s-representation: it must be disposed to generate interactions that satisfy conditions (1) and (2). Now we need a story about how we should assign content to an s-representation. While it is the dispositions of an s-representational system as a whole by virtue of which individual token interactions count as content bearing items, there is a separate story to be told about what content should be assigned to a given token interaction. In this section, I will give an outline of the view of s-

[50] Notice that to say this does not require appeal to facts about what it was designed to do. The rudder-groove mechanism might itself have been brought into the world by some cosmic fluke and it would still manifest the same dispositions.

representational content I will be defending, and in subsequent sections I will say something about why the view I'm offering is an attractive one.

Let's return to the rudder-groove case to see how the content story works. The best way to get a handle on the content of s-representations is to look at the manner in which they guide action in a good case. The rudder-groove system guided the action of the controlled system in which it was embedded (the car) by anticipating interactions between that system and its environment, and thereby preventing those interactions from taking place. Let's again take as an example the first interaction (I_A) that takes place between the rudder and groove. This interaction is similar to that which would have occurred between the car and the tunnel wall in the absence of the rudder. Condition (1) captures this fact. And the occurrence of the interaction between the rudder and groove prevents occurrence of the interaction it anticipates, as condition (2) secures. If the rudder-groove mechanism's action-guiding behaviour is to be given a representational gloss, we need to find a means of cashing out this behaviour in terms of the content it conveys.

I want to do so by distinguishing two aspects of its content – a 'representer-to-world' aspect and a 'world-to-representer' aspect.[51] I can best illustrate the point of this distinction by using a concrete case. The representer-to-world aspect of I_A's content is the interaction that it anticipates. There is some aspect of the world (not the actual world, but the world as it would be in a specified set of circumstances) that the interaction (or state-change) tokened by the system anticipates. We can say that the system assigns the property of imminence to a certain interaction between the controlled system (the car) and that system's environment (the tunnel). Put (somewhat artificially) in linguistic form, this content can be expressed as follows: 'state-changing interaction I_C between controlled mechanism and its environment is imminent', where I_C is some state-changing interaction that is minimally similar to I_A. To see why the representer-to-world aspect of the interaction's content is worthy of the name, notice that it says something descriptive about the way the represented world is – it assigns the property of imminence to a certain interaction (the interaction that would occur in its absence). This content is accurate just in case the

[51] See Elizabeth Anscombe (1957) for a classical treatment of this distinction. I use the term 'representer' instead of the more standard 'mind' because I do not want to suggest that s-representing systems necessarily count as minded systems.

anticipated interaction is in fact imminent, and is inaccurate just in case there is no anticipated interaction to which it corresponds. So in order for this aspect of the content to be accurate, it must be that the state of the content-bearing item (I_A) matches some aspect of the world in an appropriate way.

Now let's turn to the world-to-representer aspect of I_A's content. This can be expressed as follows: 'perform action A to avoid I_C'. I_A does not just anticipate an interaction; it prescribes a course of action that will lead to the avoidance of the interaction anticipated. It does so, in the rudder groove case, by causing the car to make a turn necessary to avoid collision with the tunnel (I_C), but I have couched the prescriptive content of I_A in general terms so that we can generalise beyond the specific case. This aspect of the content is world-to-representer because the world must be changed in order for the content to be satisfied. In our example, the car's trajectory relative to the tunnel must change if the car if collision I_C is to be avoided. In general terms, the content is satisfied in cases where the prescribed action (A) results in the avoidance of the relevant collision (I_C).

The first point to note about this way of cashing out the content of a given I_A is that although I have distinguished two components of its content, it should be noticed that both aspects of content attach to the very same representational token. This makes I_A what Ruth Millikan has called a 'Pushmi-Pullyu' representation: a representation with double direction of fit (1996). To illustrate this concept, Millikan uses the example of a parent saying to a child 'we do not eat peas with our fingers' (1996, p.187). This sentence both describes the way in which 'we' do or do not behave, but it is also prescriptive. It prescribes that the child refrains from eating peas in a certain manner. So the sentence has truth conditions (it is true by if it accurately describes the eating habits of the relevant group of people) and satisfaction conditions (it is satisfied if the child changes his/her behaviour to conform with the stated convention). There are questions that could be asked of this illustration, but I will not go into these. My claim that s-representations have double direction of fit does not rest on the felicity of Millikan's example. I_A has double direction of fit because it has both accuracy conditions (which are dependent on the representation's matching in the right way with the world) and satisfaction

conditions (which are dependent on the representation's making it the case that the world changes in the requisite manner).

It is also worth noting that on this story, singular and predicative content come together or not at all. While the word 'Hillary' can, one might think, represent a certain individual without predicating any properties him/her, I_A only has content when it is representing some interaction *as* imminent and prescribing a course of action that will lead to its avoidance. It is also worth noting that I_A is quite limited in the kind of content it can express. It represents one class of thing: state-changing interactions between the system whose behaviour it controls and that system's environment. And it can only assign the property of 'imminence' to the relevant interaction and prescribe behaviour that leads to its avoidance. This may seem like a fairly massive restriction on the representational powers of s-representations. But as I hope to show in the next chapters, s-representation can still do significant explanatory work in spite of these restrictions.

One last thing to note is that the content expressed by s-representational interactions on this story is essentially dependent on the context in which it is tokened. If I am at a food canteen and do not know the names of the different foods I am being asked to choose between, I might point to one tray of food-stuff and say 'I'd like that one please'. The meaning of the indexical 'that', in this situation, is crucially dependent on facts about me (the way in which I am pointing) and the situation I find myself in. Something similar is true of s-representational content. In the above example, the interaction I_A was claimed to have anticipated I_C, where I_C was an interaction minimally similar to I_A. And I_C was an interaction that would be avoided if I_A caused the system it was controlling to perform the necessary action A (a wheel-readjustment in our case). So in order to see what properties are assigned to I_c we must know certain facts about I_A (with which I_C is relevantly similar). Notice also that the notion of 'imminence' requires similar reference. I_C is imminent in the sense that it would happen shortly after the time at which I_A in fact occurred. I do not see these self-referential aspects as problematic because, as the canteen example illustrates, they are far from unprecedented.

One might additionally worry that I have left the notion of imminence quite vague, but then I see no way of avoiding this. Imagine that the car moves through

the tunnel at quite a pace. The rudder's passage through the groove would have to be similarly quick if it is to manifest the disposition described by (3). Similarly, if the car was extremely sluggish, the rudder would have to exhibit similar sluggishness. But the notion of imminent necessary in the quick case would be very different from that which would apply in the slow case. So it looks like the notion of imminence must be applied differently in different cases. It may be possible that we can still say something systematic about what standards of imminence must apply in general, but I will have to leave this as a question open for future inquiry. I don't see why imprecision on this count would fatally undermine the account I am giving.

§7: Making Room for Misrepresentation

Probably the most significant challenge facing any theory of s-representation is the problem of misrepresentation. And if the account I have given can make room for misrepresentation, that is a major consideration in its favour. To see how we can make room for s-representational misrepresentation, let's quickly recap on some relevant points. A system counts as an s-representational system when it has the requisite dispositions – those described by (3). That is, it counts as a representation when it is disposed to generate interactions satisfying (1) the minimal similarity condition and (2) the surrogacy condition. And, as mentioned in the last section, although content attaches to the individual interactions generated by an s-representational system, they count as bearers of content by virtue of being produced by a system disposed to create interactions satisfying (1) and (2).

The significance of this point is that an individual token interaction can carry representational content even if it does not satisfy (1) and (2). It can do so just so long as it is the product of a system typically disposed to produce interactions satisfying these conditions. To see how this might work, let's take a bad case of s-representation – a case of s-representational misrepresentation – to contrast with the good case discussed in the previous section. Let's use the case of the kinked groove example discussed earlier. In this situation, recall, the rudder moves into a kinked section of the s-shaped groove, where this kink in the groove does not correspond to any kink in the tunnel through which the controlled mechanism (the car) is driving. Suppose that this kink generates a jolt to the rudder. Let's use the term $I_{A=K}$ to refer to this jolt. Suppose that $I_{A=K}$ fails to satisfy conditions (1) and (2) for some $I_{C=K}$. In

other words, the jolt does not successfully anticipate, and thereby prevent, an interaction between the controlled system (the car) and its environment (the tunnel wall). Notice that the jolt is *ex hypothesi*, still an interaction produced by a system with the overall disposition described by (3). And so, on my account, the jolt still has content, even though it does not itself satisfy conditions (1) and (2).

And the content we ascribe to the jolt will be of the same form as that which is ascribed to interactions in good cases. Accordingly, the descriptive (the representation-to-world) aspect will be as follows: 'state-changing interaction $I_{C=K}$ between controlled mechanism and its environment is imminent', where $I_{C=K}$ is a state-changing interaction that is minimally similar to $I_{A=K}$. And the prescriptive (representation-to-world) aspect will be 'perform action A to avoid $I_{C=K}$', where action A is just the swerving behaviour produced in the car by the rudder's interaction with the kink. The content here has exactly the same form as it did in the good case, but the interaction described as imminent and the avoidance behaviour prescribed have both changed. And given that $I_{C=K}$ is not in fact imminent, and is therefore not in fact avoided by action A, the representation's descriptive content is not accurate, and its prescriptive content is not satisfied. This is just to say that it is a case of misrepresentation.

Room here is made for misrepresentation by the fact that there is a mismatch between the interactions that the s-representation is disposed to generate and the interaction that it does in fact generate on a particular occasion. A given interaction can be aberrant in that it can fail to count as a manifestation of the system's overall disposition. And s-representational misrepresentation occurs in cases of such aberration, since it occurs in cases where an interaction fails to manifest the disposition by virtue of which the system as a whole counts as an s-representational one. Such deviant interactions bear content by virtue of the fact that they are generated by a system with certain dispositions (dispositions from which they themselves deviate). But they misrepresent because the contents they express are neither accurate nor satisfied.

The main point can be put as follows. The status of a given s-representational interaction as an item that bears predicative content is secured by the dispositions of the system of which it is part. But the actual content borne by a given token s-

representational interaction is determined by features of that interaction itself. This was the egocentric point made at the end of the last section. $I_{A=K}$ represents an interaction ($I_{C=K}$) to which *it* is minimally similar and whose imminence is to be cashed out relative to the time at which *it* occurs. This yields a situation in which an s-representational item's accuracy conditions are different from the conditions by virtue of which it counts as a bearer of predicative s-representational content. And this in turn opens up the possibility that something might satisfy the latter conditions, while failing to satisfy the former. That is, it opens the possibility that a given interaction misrepresents while still counting as a bearer of predicative content.

It might be objected that I have not said enough to ensure a principled distinction between cases of s-representational misrepresentation and cases of failure to s-represent at all. There is something right in this objection. Suppose we start with a rudder-groove system whose disposition, described by (3), is perfect (perfect in the sense that the disposition is manifested without exception). Suppose we then start adding kinks to this system, one by one. I cannot say when exactly the system goes from satisfying (3) imperfectly to not satisfying (3) at all. But I do not need to in order to make room for misrepresentation. All I need is the claim that we could add *some* kinks (or at least one kink) like the one described above without thereby creating a system that fails to satisfy (3). If this claim is granted, then I will have made space for at least one instance of s-representational misrepresentation.

Another worry one might have is that there is a kind of circularity involved in the account I am offering here. Individual interactions produced by an s-representational system get their content-bearing status from the fact that the system by which they are produced has the disposition to create interactions of a certain sort. So the interactions' representational status depends on the system's dispositions. But the system's disposition is nothing over and above the interactions it is disposed to generate. This is not circularity but rather mutual dependence. What follows from this mutual dependence is that one cannot have a content-bearing s-representational interaction outside of a system of s-representation and one can't have an s-representing system unless you have many interactions bearing s-representational content. But this kind of mutual dependence or non-

decomposability is not unusual. Mutual dependencies of this sort are standard in functionalist theories of content and are not as ontologically strange as they might sound.[52] Imagine a pointillist portrait composed only of small dots. The picture would not be a picture if you took away all the dots, and an individual dot on its own would not predicate any properties of a depicted person. The picture's status as a representation relies on at least some subset of the dots of which it is composed. And, conversely, an individual dot can only predicate a property (a twinkle in the eye, for example) of the person depicted by virtue of being part of a larger configuration of dots.

§7.1 A Disjunction Problem for S-Representation?

One might be unconvinced by my claims about the kinked groove case discussed in the last section. Why say that this kinked rudder-groove system is imperfectly disposed to satisfy my conditions (1) and (2) for some smooth s-shaped tunnel? Why not rather say that it is perfectly disposed to satisfy those conditions for some kinked tunnel? If this description of the case is available, then I have failed to secure the claim that the case is one of genuine misrepresentation. I need the kink case to be one in which the system represents a smooth curve as kinked, and the proposed alternative gloss threatens that view of things. If the system could equally well be treated as perfectly representing some kinked groove, then we have not secured the claim that it is misrepresenting a smooth one. Can we rule out the alternative gloss? I think we can. If the interactions generated by the rudder-groove system are to satisfy conditions (1) and (2) they must be functionally similar to those that *would in fact occur* in their absence, but which *do not* occur because they were generated. And when the rudder passes over the kink in the groove, the resulting interaction will not satisfy these criteria. The kink shunts the rudder around in a way that does not mimic the manner in which the tunnel wall would, in the absence of the rudder-groove system, have shunted the car around. So when the rudder passes over the kink, the generated interaction does not satisfy (1) and (2). The system as a whole is

[52] See Lewis (1994) and Jackson and Braddon-Mitchell (2007) for theories of mental representation according to which individual mental representations (beliefs for example) cannot exist in isolation from systems of mental states. This kind of non-decomposability is similar in kind (though different in detail) from that described here.

still disposed to produce interactions satisfying those two conditions – it's just that this disposition is imperfect.

§8: A Further Requirement for Representational Error

There is another problem of misrepresentation that arises due to the particularities of the representational theory on offer here. This is the problem of passing what Mark Bickhard (1999) calls the '*system-detectable error'* criterion. Bickhard formulates the criterion as follows:

> [R]epresentation must be capable of being in error in such a way that that condition of being in error is detectable by the agent or system that is doing the representing (1999, p.436)

Applied to s-representational systems, this criterion requires that the s-representational system be capable of detecting the fact that it is in a state of error. This criterion, according to Bickhard, must be satisfied by any system whose representational status is not interpreter-dependent (1999, pp. 436-7). Since I argued in §4 that s-representations are not interpreter-dependent, Bickhard would claim that his criterion must be satisfied by my theory. I will not go into the arguments for and against thinking that Bickhard's criterion must be satisfied by any non-interpreter-dependent account of representation. I will simply show that my theory of s-representation satisfies Bickhard's criterion, and is therefore immune to criticism on Bickhardian grounds. But it is worth pointing out that while it might be contentious to claim that this 'system detectable error' criterion is necessary for counting as a representation as a certain type, it is far less contentious that satisfying this criterion is likely to be necessary for any type of mental representation that does serious explanatory work. If a representational system could not detect when it was in error, it could not update itself to deal with changing environmental factors. And it would, therefore, be a highly inflexible source of action-guidance.

But there is scope for seeing how s-representational systems could update themselves in response to detected error. When s-representation is representing accurately, the interaction between the mechanism controlled by the s-representational system and that controlled mechanism's environment is decreased. In our example, when some interaction between the rudder and groove system satisfies (1) and (2), this interaction prevents the car from crashing into the tunnel.

In at least some cases of erroneous s-representation the relevant interaction between the controlled system and its environment that is *not* prevented will actually occur shortly after the tokening of the misrepresentation. In other words, something like a kink in the rudder will sometimes cause a collision between car and wall. [53] Such un-prevented interactions provide the s-representational system with environmental feedback of a sort. And, in the elaborated version of the rudder-groove system I introduced in §4.1, this feedback is used to update the s-representational rudder-groove system. The rudder's passage through a kink in the groove would result in feedback if the kink was sufficiently substantial and was of a certain sort. And, if the kinked system was made out of the imagined putty, we could have cases where the kink would be smoothed out after a suitable number of trials.

So my augmented s-representational system doesn't just pass Bickhard's criterion, which required the system to have some sensitivity to cases of representational error. It would also satisfy a more demanding criterion, which requires the system to be capable of updating itself in situations of representational error, in order to reduce the repetition of such errors in the future. I will not argue that passing such a demanding criterion is necessary for s-representation. But anyone who thinks it is can rest assured that the s-representational system I imagined satisfies it. This is no small thing since if Bickhard's (1999) arguments are on the right lines, many theories of representation on the market fail to satisfy even his less stringent requirement. And it's not just my imagined s-representational system that satisfies the updating requirement, which is plausibly necessary for a certain kind of adaptive flexibility in a representational device. The updating requirement is also satisfied by the biologically plausible cases of s-representation I will discuss in the next chapter (as we will later see). The basic idea will be that an s-representing system only receives input in situations in which the predictions of the s-representational model are erroneous – situations which, I will argue, are cases of

[53] Note that in some cases the un-prevented interaction will not in fact occur subsequently. Suppose the system wrongly anticipates an interaction that was not going to occur in its absence and adjusts its steering accordingly. Imagine in this case that no collision was going to occur and no collision in fact occurs. The system fails to anticipate an interaction and also fails to prevent an interaction. The latter is true because you cannot prevent what would not otherwise have occurred. Suppose that, in a moment of paranoia, I think there is a murderer in my house and lock myself in the bathroom. I do not thereby prevent a murder, because there was no murder to prevent. But all this does not change the fact that in some cases, erroneous s-representation will lead to feedback, and that is all I need here.

what has been referred to in recent cognitive scientific literature as cases of 'prediction error'.[54]

§9: The Problem of Excessive Liberality

Accounts that explain a representation's ability to represent some target by appeal to isomorphism typically run into a host of standard problems. One of these problems is what I will call the *problem of excessive liberality*.[55] The basic problem is that one thing can be isomorphic with lots of others, but it does not thereby represent those things with which it is isomorphic.[56] Two computers at opposite sides of the world could, as it happens, be disposed to exhibit functionally isomorphic behaviour, but they would not thereby count as representations of each other. So I have to say what it is by virtue of which one system comes to s-represent another, aside from functional isomorphism, or dispositions to produce functionally isomorphic behaviour. And what I say has to rule out counterexamples of the kind just described. Fortunately, the account I have given is fully capable of doing this. Functional isomorphism between one domain and another is not sufficient for s-representation. For a system to s-represent, it must be disposed to generate interactions passing the minimal similarity condition (1) *and* the surrogacy condition (2). Functional isomorphism alone is not sufficient for s-representation, so the liberality problem loses its immediate bite.

One might still worry that the account is vulnerable to some more complicated sufficiency counterexample, and is therefore still too liberal. I cannot entirely rule out this possibility, but there is one feature of the s-representation notion as I have cashed it out that renders it particularly illiberal in a specifiable sense. This feature is egocentricity. Recall that I_A was defined as an interaction that occurred between components of the putative s-representation and I_C was defined as an interaction that occurred between the mechanism whose behaviour the s-representation controlled and the environment of that mechanism. By defining s-representation in terms of relations obtaining between interactions defined in this

[54] See Clark (2013) and Friston (2010) for applications of the 'prediction error' concept.
[55] See Sprevak (2011) for an enumeration of some of these problems. The problems discussed in this and the next chapter are two of the most significant challenges raised by Sprevak.
[56] For a classic formulation of this problem, see Fodor (1981). For a discussion of the problem in the context of s-representation, see Ramsey (2007, p.93).

way, I created an account on which the s-representation's target must include as a potentially interacting component the very system whose behaviour the s-representation is controlling. The car's s-representational system, in our example, s-represents the manner in which the car itself would (in certain situations) interact with its environment. This means that an s-representation cannot represent any old thing; it must represent the behaviour (under certain counterfactual circumstances) of the very system whose behaviour it is controlling. And this is not just an arbitrary stipulation. Rather, it falls out of the very manner in which an s-representation guides behaviour. The system guides behaviour in that it anticipates, and thereby prevents, interactions between the mechanism it is controlling and that mechanism's environment. And it is difficult to see how it could do so if the mechanism it was controlling was not part of the modelled domain. The system, in other words, plays its guiding role by modelling the behaviour of the mechanism it is controlling relative to that mechanism's environment. Only by modelling such behaviour can it perform its action-guiding role.

§10: The Symmetry Problem

The symmetry problem can be stated as follows: isomorphism is a symmetrical relation, whereas representation (typically at least) is not. If I have a map of London, and the map is isomorphic to the city, then it follows (by symmetry of the isomorphism relation) that London is also isomorphic to the map. But the representation relation between the map and London is not correspondingly symmetrical. The map represents the city and not the other way round. This can be seen as just a version of the *problem of excessive liberality*. There is not a representation relation for every isomorphism relation, so the representation relation must be more demanding than the isomorphism relation.

Given that the problem is similar in this way, we should expect it to receive a similar answer. And this is what we find. Again, conditions (1) and (2) tell us what it is for an interaction to act as a surrogate for an interaction minimally similar to itself. And the relation of 'being a surrogate for' is not a symmetrical relation. The rudder-groove interactions are disposed to acts as a surrogate for the car-tunnel interaction; the occurrence of rudder-groove interactions is disposed to prevent the car-tunnel interactions from occurring. In a good case of s-representation, an

interaction generated by the s-representing system causes the avoidance of the represented interaction. This causal relation is not symmetrical. Put simply, the rudder-groove interactions are disposed to occur in the car-tunnel interaction's stead (and not the other way around). This gives us the requisite asymmetry.

§11: The Causal Role of Content Problem

Suppose someone explains my act of going over to the fridge and opening it in by suggesting that I believe there is beer there. In doing so they are, on the standard story, explaining my action by hypothesising that there is a representational state (a belief) causing me to act in a certain way. And on the natural way of cashing out this story, the belief causes this particular action because of the particular content it has. If I had believed there was celery in the fridge, I may not have gone over and opened it. If I thought the beer was in the cupboard, I would have opened the cupboard and not the fridge. So the belief played the causal role *qua* representational state with the content 'there is beer in the fridge'. A problem facing any theory of mental content is to explain this tight explanatory link between a mental representation's content and its causal role.

As I mentioned earlier in passing, this problem is more difficult for some theories of mental representation than it is for others. If, for example, you have a theory of mental content that appeals to factors that go beyond causal role, then you create a gap between content and causal role – a gap which must later be bridged. The problem, as described by Jackson and Braddon-Mitchell, is that of seeing 'how one could get the needed intimate connection to behaviour by 'addition' of items that themselves are only very distantly connected to behaviour.' (2007, p.216) We would therefore expect this problem to be particularly pressing for a theory like that offered by Cummins, according to which '*[t]he content of a representation must be independent of its use or functional role'* (1996, p.86). This independence creates a gap between functional role and content, a gap which must later be bridged. I don't want to claim that this problem is fatal for accounts according to which content is fixed by something other than causal role. I only want to point out that the distance between causal role and content is less large on the current theory than it is on theories that endeavour to make content in large part independent of facts about causal role. But there is still some work for me to do on this question.

On the account offered above, the content-bearing status of individual interactions derives from the functional dispositions of the system by which they are produced. But this raises a standard problem. On most views, a disposition is not the kind of thing that can be causally efficacious. It's not the glass' fragility that causes it to break. On a standard story, it is the categorical basis of the glass' disposition that does the relevant causing, where its categorical basis includes things like the strength of bonds between the molecules of which it is composed, and so on.[57] But there is a close link between the disposition of the glass and its categorical basis, although the exact nature of this connection is up for grabs. On some stories, the glass' fragility is a second order property that supervenes on a categorical basis; facts about the glass' categorical basis settle without remainder facts about its fragility or otherwise (but not vice-versa). On other stories, dispositions are type-identical with their categorical bases; to be fragile just is to have a categorical basis of a certain type (and vice-versa).[58] On still other views, every property is both categorical and dispositional, and the dispositional aspect of a property instance is separable from its categorical aspect in thought only, but not in fact.[59]

Settling this issue is beyond the scope of this chapter (and this thesis). Different ways of cashing out the relation between a disposition and its categorical basis will yield different ways of accounting for the causal role of those representations that represent by virtue of their functional dispositions. All I can say here is that I see no reason for thinking that my account is in a position any worse than other functionalist theories in regards to this question.[60] And this is not too bad a place to be in, since the gap between the contentful and the causally efficacious is here much smaller than it is on theories according to which content is in some strong sense independent of functional role.

[57] See Choi and Fara (2014) for more on this issue.
[58] Jackson and Braddon-Mitchell defend the latter view (2007, chapter 6).
[59] See John Heil (2004, pp.197-201) for a view of this sort.
[60] See Jackson and Braddon-Mitchell (2007, chapters 6 and 15) and Jackson (1995) for one means of solving this problem.

§12: Internalism/Externalism

It is natural to ask whether the theory of content here developed is internalist or externalist. If, on the theory developed here, the content of a given s-representation was determined entirely by facts about its intrinsic properties, the theory would be internalist. If, on the other hand, s-representational content constitutively depended not only on such intrinsic properties, but also on facts about how the s-representational system is related to the environment in which it is causally embedded, then the theory would be externalist. With this distinction in place, it looks pretty clear that my theory of s-representation is externalist in character. In order for the s-representation relation to obtain, an s-representing system has to be embedded in a mechanism that is related to its environment in such a way that isomorphism relations obtain, and surrogacy relations obtain. Take the car with the rudder-groove system embedded in it. If we took this car out of the tunnel through which it is navigating and just set it loose on flat ground, the relevant isomorphism and surrogacy relations would no longer obtain. The rudder-groove mechanism would no longer be anticipating, and thereby preventing, certain interactions. And it would therefore no longer meet the conditions for s-representation. This would be true even if the intrinsic properties of the rudder-groove mechanism had remained unchanged.

§13: Conclusion

I have developed a theory of content for s-representations, and shown how it can answer some of the challenges that face theories of this kind. I do not pretend to have solved all the problems that a theory of content must solve, but I hope to have given a passable solution to some of the most serious ones. I hope, in doing so, to have shown that the theory of content sketched here is a viable enterprise. The ultimate aim, in doing so, was to show that s-representations are *bona fide* representations. In some cases, the best way to show that some putative representation is a genuine representation is to build a theory of content for it that can answer the problems such a theory must face. This is because the task of building a successful theory of content for some putative representation is basically the task of showing that the putative representation has all the properties that we ordinarily take representations to have. In showing how s-representations can assign

properties (accurately or inaccurately) to their target, and can avoid standard problems afflicting accounts of what it is to make such assignments, I hope to have shown that s-representations have some of the key properties we associate with genuine representations.

Chapter 3: The Significance of S-Representations

Abstract:

I start this chapter by trying to establish the theoretical significance of s-representations. I do so by showing that theoretically significant explanatory posits, which have already been developed in the relevant empirical literature, are instances of s-representations. In other words, I show that people are already using s-representations to explain key mental phenomena, and using them in a promising manner. This would obviously be a good result for me, given that I devoted the whole of the last chapter building a philosophical theory for s-representations. I go on to distinguish the theory of s-representation I developed in the previous chapter from a similar theory, developed by Clark and Grush (1999). Finally, I argue that while the theory I have developed is representationalist, it avoids some of the problems that have motivated writers in the 'enactivist' tradition to reject representationalism.

§1: Emulators and Predictive Coders as S-Representations

The aim of this section is to show that emulator representations described by Rick Grush and predictive coding mechanisms described by Andy Clark both count as instances of s-representations. If I can do this, then I can show that s-representations are theoretically significant, just so long as predictive coding mechanisms and emulator representations are theoretically significant. Let's start by giving a brief recap on what it takes to count as an s-representation. The three conditions that must be met by a system if it is to be an s-representation are below. When taken out of the context in which they were developed, they can seem a little bewildering. So I will paraphrase them below. It is best to read condition (3) before reading conditions (1) and (2):

(1) There is a minimal similarity between I_A and I_C.
(2) If I_A had not occurred, I_C would have. But because I_A did occur, I_C did not.
(3) Control system S is disposed to generate interactions (I_As) satisfying conditions (1) and (2) for some counterfactual interaction (I_C)

Recall that 'I_A' refers to some interaction generated by a putative s-representing system and 'I_C' was an interaction that would, under certain conditions, occur between the mechanism controlled by the s-representing system and that mechanism's environment. The basic idea is that to count as an s-representer, a

system must be (3) disposed to generate a certain class of interactions (I_As). The interactions (I_As) in question must (1) be functionally similar to interactions (I_Cs) that would take place between the mechanism being controlled by the s-representing system and that mechanism's environment. And it must be the case that (2) the interactions (I_As) which the system is disposed to generate prevent the occurrence of those interactions (I_Cs) with which they are functionally similar. So an s-representation is a control system that works by mimicking the causal dynamics that would occur in its absence and, by doing so, thereby prevents those causal dynamics from actually occurring.

I will start by showing that emulators, as described by Clark and Grush, satisfy the functional similarity criterion. Recall that in order for two interactions to be functionally similar in a manner that passes this criterion, both interactions must have similar starting states and, in both cases, these starting states need to lead to a trajectory-changing interaction. One way of putting this is to say that for both interactions, there had to be a similar mapping from causal input to causal output. This is just to say that a similar initial state leads to a similar resultant state. Using this terminology, Clark and Grush describe emulators as follows:

> An emulator is just a mechanism (circuitry, software routine, whatever) that takes as *input* information about starting (or current) state of a system (e.g. biomass, temperature, etc.) and about the control commands that are being issued (e.g. increase heat by 2 degrees) The emulator then gives as *output* a prediction of the next state of the system. This prediction takes the form of a set of values for the future feedback that the new state of the system should yield. (1999, p.6)

Let's illustrate what this means with a simple example introduced by Clark and Grush. We are asked to imagine a chemical plant that needs to 'control an on-going reaction by adding chemical to a mix but where waiting for feedback cues to prompt the process is impossible since, by the time the cues are received it would be too late for the chemical infusions to work' (1999, p.6). This is the kind of problem an emulator is designed to solve: cases in which achieving some desired result seems to require reacting to feedback that has not yet arrived. It solves the problem by emulating the causal dynamics of the emulated system (the bio-reactor in our case), predicting the feedback that the system would produce, and (crucially) providing the control system with this feedback ahead of time.

Notice that to emulate causal dynamics in this way, the emulator has to have a starting state similar to that of the bio-reactor. It then needs to move into a state which is similar to that which the bio-reactor would move into, in the absence of feedback from the emulator. In other words, the emulator needs to satisfy the isomorphism condition (1). Next, notice that if the emulator is to do its job, it must also prevent the bio-reactor from going into the state that it would have gone into in the emulator's absence. If the emulator does its job, the control system will never actually receive the anticipated feedback from the reactor. It will make the necessary adjustments in advance to avoid the situation in which this feedback arrives too late to be acted upon. So the emulator anticipates feedback in order to avoid the situation in which actual feedback is received. Notice that for the control system to receive feedback from the reactor is just for its behaviour to be changed by a change in the reactor's state (maybe an increase in reactor-temperature will cause the control mechanism to turn on some cooling mechanism). In our case, instead of being caused to do so by an actual change in reaction temperature, the control system is caused to do so by *anticipated* change reaction temperature. So by anticipating and thereby preventing feedback, the emulator is anticipating and thereby preventing a state-changing interaction between the control mechanism (which is the mechanism whose behaviour the emulator in turn controls) and that control system's environment (where 'environment' here just means that which the control system interacts with causally – the bio-reactor, in other words). This is just to say that condition (2) is satisfied. And presumably the emulator doesn't just perform this predictive and preventative role on just one occasion. Rather, it is *disposed* to do this over and again. In other words, it passes condition (3).

So it looks like emulators of the sort described by Clark and Grush (1999) satisfy my conditions on s-representation, and hence count as s-representations.[61] While Clark and Grush diverge a little in their later writing, they seem to retain their commitment to something like this emulation framework. Let's start by looking at Grush's stand-alone work. Even in his most recent (and as yet unpublished) work, he seems to have the same view of emulation, claiming that the 'emulator's job *just is* to mimic the operation of the [emulated] process' (manuscript;

[61] I cannot defend the claim that each of the emulators they describe satisfies all my conditions, since this would be too laborious a task. So if pushed, I would concede that I have only shown that at least one of what they take to be a paradigm case of emulation satisfies my conditions.

chapter 1, p.17). He doesn't emphasise the preventative role of emulator representations, but the example he uses to illustrate what he means by 'emulation' seems to satisfy the surrogacy condition (2). The example is basically identical to the bio-reactor case, but instead of a control system controlling a chemical reaction, he imagines a thermostat system controlling temperature. He imagines that this thermostat has to overcome a significant feedback delay problem, and uses an emulator mechanism to overcome the delay. I won't go into the details of the case, because they are easy enough to imagine. All that really changes is that the control system is controlling temperature rather than controlling a chemical reaction.[62] So if what I have said about the bio-reactor was right, the same should hold for Grush's new paradigm case. I am willing to admit that not all Grush's emulators are s-representations, since he sometimes wants to put emulators to all sorts of uses, some of which might not be such as to pass my condition (2). But I maintain that at least his paradigm case still passes my criterion.

Now let's turn to Clark. The predictive coding mechanisms he describes approvingly in recent work rely on generative models which, as the name suggests, are also model-like representational systems. He describes them as follows:

> A generative model […] aims to capture the statistical structure of some set of observed input by tracking (one might say, by schematically recapitulating) the causal matrix responsible for that structure. (2013, p.2)

This described recapitulation of causal matrices is what secures the similarity between predictive coders and Grushian emulators.[63] But this is not the feature on which Clark's analysis focuses. He lays particular emphasis the explanatory importance of 'prediction error' minimisation. Prediction error is the mismatch between a model's predictions and the input it actually receives, or, in Clark's words, the 'divergence from the expected signal' (2013, p.3). The idea is that a modelling system generates predictions about input that will be received, and that

[62] Indeed, as Clark and Grush suggest, it could be that the bio-reactor is controlled by means of temperature control, so it could be that the temperature controller described above just was a thermostat (1999, p.6).

[63] Indeed, as Clark mentions, Grush's emulation theory is 'highly congruent' with the story Clark develops, and differs mainly in placing less emphasis on certain technical details that I won't go into here (see 2013, p.24 n. 44). Grush (2007) also considers the mechanisms Clark writes about to be kinds of emulators. He also considers the same to be true of Ramsey's s-representations (2008).

'prediction error' is the divergence of the input actually received from the input predicted.

Prediction error is interesting for a number of reasons. First, the predictive coders Clark describes are only sensitive to prediction error; they are not sensitive to any input that is already predicted. This means that predictive coding mechanisms are sensitive only to 'news-worthy' information, and that this selective sensitivity allows them to save 'bandwidth'. The coders don't have to record expected input, because they can 'assume' that no news is good news (no news means their predictions were correct). So while they ignore predicted input, they use prediction error to fine-tune themselves; the prediction error acts to refine the models in order to reduce future prediction error. I hope this picture is intuitively familiar from the last chapter. Our rudder-putty-landscape mechanism only received input when its predictions had gone awry, and the effect of input, when it was received, was used to update the putty landscape to better reflect the features of the car's environment.

The second interesting feature of prediction error is that it acts as a kind of currency of exchange between hierarchical layers of predictive coding mechanisms. To see how this works, first imagine a set of predictive coders on the 'ground' level. These coders generate predictions about sensory input received from the environment, and do so by modelling the causal dynamics of that which the sensors sense. At the next level up, we will have more predictive coders whose job it is to make predictions about the dynamics of the coders on the level below. They do so by modelling the causal dynamics of these coders. These higher level predictors themselves only receive input when the predictions they make about the lower level coders go wrong – and these are cases of what you might think of as higher order prediction error. So prediction error moves between levels of the hierarchy.

The final point of interest about these coders is that, like the rudder-groove system we saw in the last chapter, they are disposed to create behaviour that reduces input to the system – so to reduce prediction error. Our rudder-groove mechanism produced behaviour that minimised collisions between the car and its environment, and hence minimised input to the system. Something similar is supposed to be true of predictive coders. And because there is an organised hierarchy of coders, each layer influences the behaviour of the lower layer so as to minimise prediction error.

So there is reciprocal influence between the layers. When the lower layers mis-predict, the resultant prediction error moves up the chain of command, and the higher-level layers prescribe those actions that tend to minimise future cases of prediction error. And the idea is that as you increase the number of layers, you increase the degree of cognitive sophistication. The lower layers are capable of detecting simple regularities in the input they receive, and predicting the continuation of these patterns. And higher-level layers tune to patternings in the simple first-layer patterns, and so on up the chain. The ultimate idea is that the first level layers might tune to simple regularities, like edges and lines, while the higher level layers might tune to more complex environmental features. Note that these coders satisfy my conditions. They mimic the functional profile whatever they make predictions about, and they do so in a way that minimises input being received any state-changing input from that which they model. So they are disposed to anticipate, and thereby prevent, input.

Clark's most ambitious claim is that, when we put all this together, what we get is a unifying perspective on the mind:

> Perception, cognition and action – if this unifying perspective proves correct – work closely together to minimize sensory prediction errors by selectively sampling, and actively sculpting, the stimulus array. They thus conspire to move a creature through time and space in ways that fulfil an ever-changing and deeply inter-animating set of (sub-personal) expectations. (2013, p.6)

If something like this picture turns out to be correct, we have come a long way from our simple toy rudder-groove case, but in both cases the basic principles are the same. In both cases we have a model that predicts the behaviour of a system it controls, a system that predicts what input that system would predict in the absence of the modelling controller. And in both systems, the modelling system guides the behaviour of the system it controls in order to prevent a situation in which this input is actually received. And in both cases, input is only received for the environment when the model fails to predict accurately.

As we have just seen, if the headiest enthusiasts are to be believed, predictive coders (which, I have given reason for thinking, are s-representers) might form the basis for

something of a unified field theory for the mind.[64] On (slightly) more modest proposals, they might provide enlightening explanations of a whole range of phenomena, from binocular rivalry to delusions and thought insertion phenomena.[65] And, if Grush's reviews of the relevant scientific literature are to be believed, emulation theory holds similarly significant explanatory promise (see Grush 1997, 2003, 2004, 2007 for these reviews). I am not qualified to add anything to these claims, or to pass judgement on their plausibility (although in a later chapter, I will explore some implications that some of Grush's claims would have if true). My only aim was to tie the notion of s-representation to explanatory posits in which there is significant empirical interest. If I have succeeded in doing so, then I can claim to have provided a theory of content for an empirically significant explanatory posit. And that would be a happy result.

§2: Clark and Grush's Decouplability Criterion

As we have seen, I claim that emulators and predictive coding systems just are s-representations. And my claims about the ubiquity and significance of s-representations in cognitive systems piggy-back on the arguments advanced for the claim that emulators and predictive coders are ubiquitous and significant. Given this, one might wonder how my account is supposed to differ from that offered by Grush and by Clark. I don't disagree with them about the nature of the systems they describe. Nor do I disagree with them about the representational status of these systems. But I disagree with them on the question of what it is by virtue of which these systems get to count as representations (what it is that grounds their representational status). The theory of content for s-representations I detailed in the last chapter is distinct from anything devised by either Grush or Clark. In this section, I give a sense of what the relevant differences amount to. I'll start by targeting a definition of representation leaned on in an influential paper co-authored by Clark and Grush. By showing the inadequacies of this definition, I hope to kill two birds with one stone by distancing my position from both Clark and Grush. In the next section I will argue that as well offering an account that differs in substance from mine, Clark and Grush also differ from me in their theoretical aims. I aimed to

[64] See Friston (2010) for a suggestion that this might be the case, and for a useful review of the predictive coding literature.
[65] See Clark (2013) for a review of the evidence in favour of such hypotheses.

provide a theory of content for s-representation that deals with the standard problems that any such theory must tackle. Clark and Grush seem to show little interest in such a project, or so I will argue.

But first to the differences in substance. Clark and Grush offer the following definition of representation:

> [O]ur suggestion is that a creature uses full-blooded representations if and only if it is possible to identify within the system specific states and/or processes whose functional role is to act as *de-couplable surrogates* for specifiable (usually extra-neural) states of affairs. Motor emulation circuitry, we think, provides a clear, minimal evolutionarily plausible case in which these conditions may be met. (Clark and Grush; 1999, p.8)

I will argue that this definition is both unnecessary and insufficient for representational status. It is first worth noting that this definition is presented as a definition of representation in general. Clark and Grush want to define representation in general, and then show that s-representations (or emulator representations as they call them) fit this definition. To engage with this definition, we first have to find out what is meant by 'decouplability'. Clark and Grush do not define the term in the paper I quoted from. They come closest to defining it when they talk of 'the basic strategy of using inner states to stand in for (in this case temporarily) absent states of affairs' (1999, p.10), but this was not offered as a strict definition. The best place to look for an explicit definition of 'decouplability' is Clark's stand-alone work. It seems to be Clark who introduced the term. Clark describes decouplability as 'the capacity to use inner states to guide behaviour in the absence of the environmental feature [for which it is a surrogate]' (1997, p.144).[66] I will assume that when Clark and Grush say a system must function as a decouplable surrogate, they mean that the system must guide behaviour in the absence of the environmental feature represented. In doing so, I will be following such critics as Wheeler (2005) and Chemero (2009). I want to show, *contra* Clark and Grush, that the decouplability condition is not sufficient for s-representation.

[66] Grush seems to have something similar in mind when he writes that 'the ability to use an entity as an off-line stand-in depends crucially on its not being causally linked to, and its not carrying information about, the entity it represents' (1997, p.7). See also (2003, p.83) for use of the decoupling idea.

But first it is worth noting that the condition is going to need tweaking if it is to be even initially plausible. When Clark talks about guiding behaviour 'in the absence of' some environmental feature, he surely means 'in the absence of *feedback from*' that feature. Suppose I memorise the layout of my house so thoroughly that I am able to walk through it with my eyes closed, without feeling my way around. This seems like a paradigm case of decouplability. I am using my memory to guide my action in the absence of any perceptual information about the house. But the house is still *present*. If it was not, I would not be able to walk around it. The point is surely that any feedback from the house (in the form of perceptual information) is absent. From now on, I will assume that Clark and Grush mean 'absence of feedback' when they talk about the absence of some feature represented. With this minor tweak in place, it is easy to see that s-representations typically satisfy this decouplability constraint; an s-representation makes predictions that are used to guide some control system's behaviour, such that the control system doesn't have to rely on actual environmental feedback. So s-representations allow the control system to work in the absence of actual feedback. But although this condition may be typically satisfied by s-representations, it is not sufficient for representational status. We can imagine decouplable systems that do not count as representers. This point is not a new one. It has been argued convincingly by both Chemero (2009) and Wheeler (2005).

Chemero argues that decouplability can be achieved 'just by causal connection and momentum' (2009, p.57). First he asks us to imagine a bird tracking a fox as it moves behind a rock. Before the fox goes behind the rock, the bird is rotating its head to track the fox's movement. Suppose that, when the fox goes behind the rock, the bird's head just carries on with the same momentum it had before. Also suppose that the fox moves at the same rate while behind the rock. When the fox comes out from behind the rock, the bird will still be tracking its movement accurately. In this scenario, the bird has used its own state (the momentum of its head-rotation) to guide its behaviour in the absence of feedback from the thing it is supposed to be tracking. So, according to Clark and Grush's decouplability condition, it looks like this should be a case of representation. But it seems counterintuitive to say that the bird is representing the occluded fox by simply continuing to move its head at the same speed during the occlusion phase.

Wheeler makes a similar argument, claiming that Rutkowska (1994) has argued convincingly that such cases of tracking during occlusion phases is 'seemingly a matter of serendipitous mechanical ballistics, not of representation-driven anticipatory search' (2005, pp.215-6). To count this kind of case as a case of genuine representations would be to cheapen the notion considerably. Yet it seems to meet the decouplability constraint, suggesting that the decouplability constraint is insufficient for representation. These considerations have been taken by Chemero to be reasons for rejecting Grush's emulator notion of representation. I consider this rejection to be premature. Chemero and Wheeler are right to think that the definition of representation Clark and Grush are working with is inadequate. But Grush's emulators are not the problem. The problem is the definition of representation with which he is working.

Before moving on, it is worth adding that the decouplability condition is not only insufficient for representation. It is also unnecessary. To illustrate this, let's return to the rudder-groove system discussed in the last chapter. Let's start by imagining a perfectly decoupled version of this system: the car navigates through the tunnel without receiving any feedback at al from the tunnel. Suppose we are happy to call this a case of genuine representation. Now imagine that we add to this system a sensing mechanism that allows it to receive feedback from the tunnel wall (let's imagine that it sends out sonic pulses and then uses echolocation to determine its distance to the wall). Suppose also that, in this new setup, the rudder-groove mechanism can only guide behaviour successfully when it operates in tandem with the feedback gleaned by echolocation, but that it otherwise works in the same way. Imagine, say, that the sensory information must be somehow collated with the rudder-groove's anticipatory feedback, such that successful steering happens only when both information sources are in agreement. In this scenario, the rudder-groove mechanism still has the functional features that tempted us to think of it as representational in the first place. If we were tempted before to say that the rudder-groove mechanism looked to be behaving like a model, nothing in the alteration I describe should change this fact. Yet once this alteration is in place, the rudder-groove mechanism is no longer decouplable in the relevant sense. It is no longer guiding the car's behaviour in the absence of feedback from the environment with which the car is interacting.

Note that it is important to my account that decouplability is not necessary for s-representation. If decouplability *was* necessary for s-representation, I would be obliged to show that my conditions on s-representation implied the decouplability condition (on pain of admitting that my conditions are insufficient for s-representation). But, because I have shown that decouplability is not necessary for s-representation, I do not have to discharge this obligation.

§3: A Difference in Theoretical Ambition

One might worry that the analysis of the previous section was a little unfair on Clark and Grush. Though their use of the decouplability criterion is misleading (if not confused), a sympathetic reading of the paper in which they formulate it suggests that they do have a good understanding of the features by virtue of which s-representations represent. Their descriptions of emulators are clearly designed to pump the intuition that the emulators act like models in recapitulating the causal dynamics of that which they represent. Indeed, it was these kinds of descriptions that I relied on in §1 when I argued that emulator representations just were s-representations. Given this fact, one might argue, there really isn't much distance at all between my theory of s-representation and the theories they put forward. I want to ward off this kind of criticism by showing that there is a difference in theoretical ambition between my account and those offered by Clark and Grush. I gave a theory of content for s-representation and attempted to show that s-representations have all the properties we normally associate with representations. I argued that my account solves the problem of misrepresentation, the problem of asymmetry, the liberality problem, and so on. Clark and Grush don't seem to be engaged in this kind of project, or so I will argue.

My argument to this effect focuses on Grush. While I am fairly confident that Clark does not undertake the aforementioned project in any detail, he is a prolific writer and I may have overlooked some part of his oeuvre where such issues are dealt with. So I will focus instead on Grush, who sometimes comes very close to dealing with concerns of this nature. And the first thing I should do is roll back on my claim that Grush has never addressed these concerns in any detail. He has done so, but, to the best of my knowledge, only in his PhD dissertation (1995, pp. 152-174). And while in this work he endorses an instrumentalist view of representation

based largely on the work of Cummins (from whose view I distinguished my own in the last chapter), he distances himself from this kind of view in his published work. So, while his original view was a version of 'interpretational semantics' (1995, p.153) akin to that offered by Dennett and Cummins (see Grush; 1995, p.155), he later rejects views of this kind (1997, p.17). But he doesn't then replace his old analysis with a new account of the conditions by virtue of which emulator representations exhibit the properties we normally take representations to exhibit. He has a brief paragraph addressing the objection that emulators don't exhibit 'well-known properties of representational content such as failures of substitution' (1997, p.18). But the attempt to show briefly that emulators exhibit one example of a property ('substitution') associated with content is not the same as a systematic attempt to accommodate the whole range of such properties.

Elsewhere, Grush briefly addresses the problem of misrepresentation, but what he says doesn't seem to grapple with the real source of the puzzle. To demonstrate the possibility of emulator misrepresentation he asserts the possibility of a situation in which '[T]he emulator might not quite have [input-output] the mapping down. The image of the arm might in some conditions move farther than the real arm given the same motor command' (2003, p.83). The situation Grush envisages here is one in which an emulator, which is supposed to emulate the movement of a represented arm (given certain motor commands) fails to emulate accurately by emulating an arm-movement that goes farther than the represented arm would have done, given the relevant motor commands. I agree that this would be a case of s-representational misrepresentation, but the task of meeting the challenge of misrepresentation is not that of pointing to a case in which such misrepresentation occurs. You can't just point to a case where misrepresentation intuitively seems to occur: you need also to solve all the standard problems that come along when you attempt to account for misrepresentation. The challenge is that of providing a set of conditions by virtue of which a system counts as a misrepresentation, such that the system can satisfy those conditions while still counting as a case of misrepresentation. So, if an emulator is defined as a system that mimics the input-output mapping of that which it represents, we have to explain how a system can still count as an emulator while failing to produce the relevant mapping. In the last chapter, I showed that this could be done by appeal to

imperfect (exception-tolerating) dispositions. This is not the only available option, but the challenge of misrepresentation must be met by appeal to an option of this sort. Merely pointing to a case in which it is intuitive to think that misrepresentation is occurring does not engage properly with the challenge.

In a more recent paper, Grush distinguishes his theoretical project from that of giving a distinctive theory of content. He writes as follows:

> My concern is with the spatial content of experience. But since the word 'content' often is used to indicate what some word is *about*, it won't quite suit my purposes. As I will explain later, there can be states, in particular experiential states, that *carry information about* space and spatial relations […] while not having any spatial significance for the subject. (2007, p.2)

He goes on to say that he will be concerned with 'purport', both behavioural and spatial, instead of content. So Grush here seems to be saying that aboutness, or contentfulness, is not sufficient for behavioural and spatial 'significance' or 'purport'. In the rest of his paper, he goes on to spell out what must be added to contentful perception in order to get purport of the right kind. We need not worry about this part of his story. Let's accept that content isn't sufficient for purport of the right kind, and accept also that giving an account of purport is a worthwhile project. The fact that content is insufficient for purport does mean that we can simply take content for granted and divert our efforts to working out what must be added to content in order to get an account of the kinds of phenomena in which we are interested. Getting an account of how a state comes to have content is still a key task, and not one that can be passed over in silence.

This point is not supposed to be especially hostile to Grush. I am attempting to show that his theoretical enterprise is incomplete rather than misguided. His account is incomplete in that it lacks a theory of content. And I hope my previous chapter will go some way towards rectifying this incompleteness. The idea is that once we put my theory of content together with the emulation theory developed by Grush and others, we have a theory of content for a notion of representation capable of doing interesting empirical work.

§4: S-Representations and Enactivism

One question to ask at this point is whether s-representation theory (or emulation theory for that matter) constitutes a departure from 'enactivism'. The answer to this question depends very much on what is meant by 'enactivism'. I won't attempt to define this contested term. I will rather look at different theses developed by people in the 'enactivist' tradition and ask which of these the s-representation story is compatible with. The first thesis I consider is anti-representationalism. It should go without saying that the s-representation story developed here is incompatible with anti-representationalism (a view according to which the role of mental representations in the explanation of mental phenomena is far more limited than most theorists think it is). If the s-representation story here told is accurate, even very simple behaviours are explainable in representational terms. And this result does not sit well with anti-representationalism.

Dan Hutto and Eric Myin argue for thoroughgoing scepticism about mental content. They do so by arguing that the mental 'representations' posited by philosophers of mind are representations in name only. Let's see if their arguments cause problems for the claim that s-representations are genuine representations. Hutto and Myin claim that teleosemantic theories of mental content (and their predecessors) are all fatally flawed. Hutto and Myin suggest many theorists have failed to provide successful theories of content, despite the sustained effort of many talented theorists over the course of decades. And they claim that this is an indication that there is something wrong with the project. They quote approvingly from Godfrey-Smith, someone who was 'initially optimistic about the prospects of teleosemantics', to suggest that there is a growing sense of discontent even within the teleosemantic camp (2013, p.80). The quote goes as follows: "There is a growing suspicion that we have been looking for the wrong kind of theory in some big sense. Naturalistic treatments of semantic properties have somehow lost proper contact with the phenomena' (2006, p.42). Quoting someone who used to think *X* but now thinks *Y* is not an argument for *Y*. But let's suppose Hutto and Myin are right in

their assessment of teleosemantics and predecessor theories.[67] Is this analysis going to extend to the attempt to provide a theory of content for s-representations?

In one obvious sense, this analysis will not extend to the s-representation account I have been defending. The account is a very new attempt to provide a naturalistically acceptable theory of mental content. It has barely yet been formulated completely, and so has not yet had the opportunity to be tested to destruction. I have also made attempts to mark off my approach from the standard teleosemantic approaches on which Hutto and Myin focus.[68] In this respect at least, it is not vulnerable to the 'tested to destruction' line of attack developed by Hutto and Myin. But there is little I can do to engage with scepticism of the Hutto/Myin stripe, other than to present the theory of representation developed in the last chapter and challenge the sceptic to find fault with it. I tried to show that s-representations have the properties necessary for counting as content-bearing representation, and the onus is now on the sceptic to prove me wrong.

§5: Representationalism and Anti-Computationalism

As we have seen, if 'enactivism' amounts to anti-representationalism, then the account I have been developing is plainly incompatible with enactivism. But I now want to examine the relation of my s-representation story to another theme that runs through much of what might be called the enactivist literature: anti-computationalism. Computationalism, roughly, is the view that the mind has a functional architecture similar to that of a computer. I use the term 'computationalist' to refer to any theory that endorses the 'language of thought' hypothesis. According to this hypothesis, mental representations are fundamentally language-like. On views of this sort, mental representations that assign properties to objects in the world are built out of representational atoms, or symbols, which are combined according to syntactical rules of composition in order to create sentences in 'mentalese'. So the brain might have a symbol for clouds and a symbol for the

[67] My assessment of Hutto and Myin's arguments does not depend in anyway on the details of the teleosemanticist positions they critique. This is because I assume for the sake of argument that the claims they make about such positions are correct, and try to determine what follows from this assumption. I pursue a similar strategy in the next section too. Because my argument doesn't depend on such points of detail, I do not need to describe the teleosemantic theories that Hutto and Myin critique.

[68] I did so by being sure to avoid reliance on the teleological notion of proper function.

property of fluffiness and combine these to form the sentence 'clouds are fluffy'. Once cognizers have enough symbols of this sort, they can perform logical operations on them in much the same way that we do when we make inferences using sentences in a natural language.[69] This very brief sketch may not do complete justice to the subtleties of the varying views I am labelling as computationalist, but it will suffice for my purposes. All I want is a rough and ready description of what it means to be an anti-computationalist. And I will be using the label to refer to anyone who rejects the rough picture of mental representation I have just sketched.

Anti-computationalists are sometimes, but not always, anti-representationalists. Anti-computationalists sceptical about the explanatory role of mental representations in general include Dreyfus (1992; 2007), Brooks (1991), Ramsey (2007), Hutto and Myin (2013, cf. pp.2-3). Whereas anti-computationalists who are less sceptical about the explanatory importance of mental representations include Wheeler (2005), Grush (2003) and Clark (2013). I count myself in the latter group. If the s-representation story developed so far is on the right lines, and if this kind of representation turns out to be explanatorily potent, then scepticism about the explanatory role of mental representations is misguided. But, as should already be clear, it is possible to have mental representations whose functional architecture is radically different from a classical computational architecture. Take, for example, our toy case: the rudder-groove mechanism. It would be quite a stretch to interpret this system as a symbol-manipulating computational mechanism, yet it was supposed to be a case of s-representation. Similarly, as Grush is keen to emphasise (2003), it is possible to implement emulators with neural networks and connectionist systems (systems often touted as alternatives to classical computational architectures). This is not to say that one could not have computational systems that satisfy the conditions I set for s-representation. I imagine there are many computer models that satisfy the conditions for s-representation I set down in the previous chapter. The point is only that you don't have to be a computational system to be an s-representing system. So one can be sceptical (as I am) about computationalism, while still being a representationalist.

[69] Jerry Fodor (1975) is the *locus classicus* for a view of this sort. See Carruthers (2006) for a more modern manifestation of the view.

And one can do so by thinking that a certain class of s-representations are ubiquitous in cognitive systems.

In taking this position, I find myself at odds with the spirit, if not the letter of Ramsey's analysis. Ramsey often implies that the only really *bona fide* representationalist theories are computationalist theories. He claims that alternatives, like connectionist explanations of mental phenomena, are representationalist in name alone (2007, chapter 6). He does acknowledge that there are some exceptions to this analysis, however. He writes that 'there are also a few connectionist-style theories that invoke model-based representation in their explanations. See Grush (1997, 2004) and Ryder (2004) for nice illustrations' (2007, p.80 n.5).[70] But he plays down the significance of these exceptions. He relegates them to the footnotes, and doesn't let them interfere with his broader narrative, which suggests that a victory for connectionist and similar non-standard theories of cognition would be a defeat for representationalism (2007, chapter 6). I think that the exceptions to Ramsey's general narrative are more significant than Ramsey makes out, as §1 should have made clear. But this is ultimately an empirical question, and one on which I suppose only time will tell, but in what follows I will try to undermine some arguments that might seem to speak in favour of something like Ramsey's position.

But before doing so I want to consider a different attempt to tie the fate of anti-computationalism to that of anti-representationalism – an attempt I take to be demonstrably unsuccessful. The attempt I have in mind is that made by Hutto and Myin (2013). I am not concerned here with the status of their argument for anti-representationalism. As I have said, there is little I can say in the face of this argument, other than "here's my theory of representation, tell me what's wrong with it", or something to that effect. What I am concerned with is the question of whether or not the anti-representationalism they argue for somehow gets them anti-computationalism. They seem to think it does. They write in the introduction to their book that '[d]efenders of REC argue that the usual suspects – representation

[70] I do not discuss Ryder's theory in detail here. As I mentioned in the previous chapter, his theory relies on the teleological notion of proper function , so my analysis differs from him in that respect. Moreover, I do not think Ryder's (2004) SINBAD system satisfies my conditions for s-representation. Indeed I do not think this system is a genuinely representational one. But I will not argue for this claim here.

and computation – are not definitive of, and do not form the basis of, all mentality' (2013, p.3). Since 'REC' (which stands for 'Radical Enactive Cognition') is the position they set out to define and defend in their book, it is natural to assume that they take the arguments they go on to present as arguments against computationalism as well as arguments against representationalism. This interpretation is further buttressed by the following passage:

> The conclusion that informational content is not naturalistically respectable may come as something of a shock. If accepted, it can lead to scepticism about the tenability of classical representational and computational explanatory strategies in cognitive science. (2013, pp. 37-8).

So I think we can be fairly confident that they take their anti-representationalism to present a challenge to computational theories of mind. And it's easy to see why they might think this. After all, computationalism is a representational theory of mind, so it's only natural to suppose that arguments for the conclusion that representationalism is false might also imply that computationalism is false. But as we will see, this result isn't quite as obvious as it seems.

To show why this is, I first need to say something about Hutto and Myin's argument for anti-representationalism. They argue for this position basically by arguing that what they take to be the most promising naturalistic theories of mental content on the market are unsuccessful. And because there is no naturalistically respectable theory of content on the market, they argue, a naturalistically respectable theory of mind cannot invoke the notion of mental content as an explanatory postulate. To do so is to incur the explanatory equivalent of a 'toxic debt' (2013, p.160): an unacceptably onerous explanatory burden. This, crudely, is their argument for anti-representationalism. The main point for my purposes is that it doesn't rest on claims about the way the mind works. Rather, it takes an explanatory posit that theorists of mind invoke in their explanations (representational content) and shows that a naturalistic account of this posit's properties cannot be given.[71]

But, Hutto and Myin claim, this is no cause for despair, because we can have all that is good about teleosemantics (their favourite failed theory-of-content-

[71] They do so by somehow jumping to this conclusion from the weaker claim that such an account has not, to date, been given, despite the efforts of many clever people.

providing enterprise), just without the 'semantics' part. They use the term 'teleosemiotics' to describe the position generated by thus defrocking teleosemantics, and say the following on its behalf:

> Nothing important is lost if we put teleosemiotics in the place of teleosemantics. Teleosemiotics is (basically) teleosemantics without the semantic ambitions […] Teleosemiotics is simply the logical fallout of rejecting the ambition to understand basic responding in representationalist terms while keeping the remainder of the teleosemantic apparatus. (2013, pp. 78-9)

But if this is the upshot of their anti-representationalist argument, there is room for computationalists to accept its conclusion while remaining cheerfully committed to their basic theory about how the mind works. Let's see why.

Suppose we have a group of computationalists who are (initially) fans of whatever form of teleosemantics Hutto and Myin are aiming to de-semanticise. These computationalists claim that the mind works by processing mental representations much like those symbolic representations processed by conventional computers. And, our computationalists claim, these mental representations get their representational status by virtue of satisfying the teleosemanticist's conditions on representation. Suppose our computationalists then become convinced of Hutto and Myin's arguments, and thereby come to accept that satisfying the teleosemanticist's conditions is insufficient for counting as a genuine representation. Suppose they then give up on the idea that what they previously called 'representations' are, in fact, bearers of representational content. They can do so without making substantive changes to their theoretical apparatus. They can claim that the mind has exactly the functional architecture they previously attributed to it, while simply rolling back on their claim that this architecture is a representational one. And, I submit, doing this does not require them to relinquish their claim to being computationalists. They still think the mind works in much the way a computer does. They've just dropped talk of representations. And if this is right, there is plenty of room for a computationalist to accept Hutto and Myin's anti-representationalism while rejecting their anti-computationalism. So the former view does not imply the latter.

And suppose now that our computationalists are initially committed to some theory of content other than the teleosemantic one Hutto and Myin favour. The

argument of the last paragraph can be rerun by just substituting whatever this theory might be for 'teleosemantics' above. This is true because nothing I said above hinges on any details about what 'teleosemantics' amounts to. The situation still runs as follows: computationalists ground the representational status of their theoretical posits with theory of content *C*, Hutto and Myin convince our computationalists that *C* is a failure, computationalists drop *C*, de-semanticise their theory of mind, and carry on as they were before. One might object to this generalisation that Hutto and Myin's de-semanticising move only works if you started off with teleosemantics, and won't work for any old theory of content. This might be the case, but I see no reason to think that it is. And the burden is on Hutto and Myin to convince us otherwise.

§6: The death of the representational bottleneck argument

It is worth pointing out that there is a certain argument for anti-representationalism, not discussed so far here, which is potentially undercut by Grush's emulator theory. This is the argument from representational bottlenecks. The basic idea of this argument is that computational systems tend to be very bad at tasks that involve real-time interaction with a changing environment (Brooks 1991, Wheeler 2005, van Gelder 2005).[72] Wheeler describes the robot 'Shakey', a robot built on computationalist principles, as a stark illustration of the limitations of classical computationalism in this regard (2005, p.69). Shakey operates on a 'sense-represent-plan-move' framework (2005 p.68). So Shakey first receives visual images from a television monitor, and then creates a model of the world on the basis of these images. This model is 'built as a set of first-order predicate calculus expressions' (2005, p.69). The robot then plans by manipulating the model 'according to problem-solving rules' (2005, p.69). Once planning has been carried out, the robot can move on the basis of this planning. Because of the nature of this cycle, Shakey is unable to perform fast real-time interaction with an environment. This is because Shakey must constantly update its model every time it moves (since it will receive a different set of images from each new location).

[72] I should make clear that Wheeler does not move from this claim to the claim that all representations result in such bottlenecks. He allows action-orientated representations might avoid such problems (2005, pp. 159-200).

Brooks famously claimed that his robots performed much better in such real time tasks by using 'the world as its own model' (1991, p.139). By doing without internal models, Brooks' robots avoid the problem of constantly having to update models on the basis of new information. They are thus able to move around more smoothly and perform simple tasks with greater ease. In a similar vein, van Gelder makes much of the ability of simple control systems like the Watt Governor to perform complex tasks that require extremely fast real-time adjustment without using representations of any sort. Such advocates of the representational bottleneck argument claim that representations typically get in the way of fast real-time interaction with the world, creating delays of the sort encountered by Shakey. They argue that systems that make no use of representations perform much better at these kinds of tasks.

Grush agrees that there are serious problems with classical computationalism, and agrees that the control systems van Gelder and Brooks describe do not require representations of any kind to work their magic (2003, p.87). But he maintains that systems like Brooks' robots and the Watt governor described by van Gelder could carry on working their magic, and indeed could benefit from significantly improved performance, if they were augmented with emulators. The reason his argument has the potential to undercut bottleneck arguments based on the temporal constraints of certain types of tasks is that at least some of his emulators are designed with exactly these constraints in mind (as the bio-reactor example discussed above illustrated). For example, Grush claims that, when engaged in fine-grained control of movement and posture, our nervous systems are sometimes able to react to feedback that has not yet arrived. He writes that 'it appears as though motor centres make corrections to the motor plan as quickly as 70ms or so after movement onset, corrections apparently made on the basis of peripheral information (2003, p.76). But peripheral information takes longer than 70ms to reach the motor centres, suggesting that 'the motor centres are getting and acting on feedback *before peripheral feedback should be available!'* (2003, p. 76).

Grush explains this phenomenon by suggesting that the motor centres involved send efferent copies (copies of their motor commands) to emulators which in turn emulate the effect of these motor commands and thereby provide 'mock'

feedback ahead of time. While this all sounds very speculative, Grush does provide substantial evidence to back up his hypothesis. I will not discuss this evidence here. For now the main point is that Grush's theory, if correct, suggests that representation may be used in parallel with non-representational processes (of the sort described by Brooks and van Gelder) in order to *speed up* their ability to react to feedback. If true, this would certainly undercut the 'representational bottleneck' argument for anti-representationalism. Of Brooks' robots, Grush writes that 'Brooks' robots are under no pressure to be terribly fast, and get their sensory information carried via electrical signals in very fast wires, not slow biological axons' (2003, p.84). While non-representational systems like Brooks' robots may be faster than Shakey, systems with such design may still not be fast enough to be biologically plausible. What's more, s-representations may be able to speed them up.

To see how an augmentation of Brooks' robots might work, we can imagine a version of one of Brooks' robot built with extremely slow wires. In Brooks' robots, sensors are wired directly to motors (directly in this context means without any representational intermediary). So, suppose we have a Brooksian robot that is built to avoid things. When the sensor senses something, it activates some motor which will move the robot away from that thing (this is a simplification of Brooks' robotic systems, but it will do for present purposes). Normally, very little time elapses between the sensor sensing something and the movement starting. But with our slow wires in place, let' s say the signal takes 3 seconds to travel. The robot will therefore only start moving backwards after 3 seconds. Now, after it has moved backwards for a little while, the sensor will stop sensing anything, meaning that the robot should come to a halt. But the robot will not halt immediately, as it is still receiving the 'move' command sent 3 seconds ago. It will carry on moving for another 3 seconds before stopping. Suppose we want to reduce this delay, but cannot improve the speed of the wires. We could augment the robot with a Grushian emulator to do this. This emulator might emulate the effect that the movement command sent to the motor would have on sensor's proximity to the object detected. It can then predict that, once the motor has been active for a certain amount of time (x seconds, let's say), the object will be so far away that the sensor no longer detects it. It could then, on the basis of this prediction, turn off the motor

at time x. In this way, the emulator can cause the robot to stop the second the sensor stops sensing an object. Although this is a toy example, Grush provides evidence to suggest that something similar to this might be going on in our own nervous systems. I intend this case as an illustration of the idea that it is possible to reject computationalism in favour non-computational models of cognition, without giving up on representationalism. The positing of Grushian emulators is perfectly compatible with non-computational theories of cognition, and might serve to increase the explanatory power of such theories.

§7: Conclusion

I hope here to have shown that s-representations are theoretically significant, that my account of what it is to be an s-representation is distinct from that developed by Clark and Grush, and that it is possible to be enthusiastic about s-representations while being sceptical about computationalism. I also argued that some of the standard reasons for being sceptical about computationalism do not obviously transfer across to create reasons for being sceptical about s-representational theories of mental representation.

Part II

Chapter 4: Action and Perception

Abstract:

The arguments in this chapter are primarily negative. The purpose is to isolate those parts of Noë's theory that I want to reject. This will free me up to focus in subsequent chapters on the aspects of Noë's theory I want to accept, refine and build upon by deploying some of the resources developed earlier in the thesis. Noë is best known as an advocate of the thesis that there is a close relation between action and perception. This thesis takes centre-stage in his *Action in Perception* (2004), and in much of his other writings. Despite the centrality of this thesis to much of Noë's work, it is often unclear what it actually amounts to. Noë is clear and emphatic on the point that a close relation between action and perception obtains, but it is difficult to determine exactly what he takes this close relation to be. In the discussion that follows, I will not spend much time trying to determine exactly what Noë takes the relation between action and perception to be. Rather, I will examine in turn a number of different theses about the relation between action and perception, each of which seems to be endorsed by Noë at one point or other. My aim will be to determine which (if any) of these theses are plausible.

§1: Metaphysical Versus Nomological Necessity Claims

First some words on the kinds of claims I will not be discussing in this chapter. As we saw in chapter 1, some of Noë's claims about the relation between action and perception can be taken as the expression of a version of the extended mind thesis (what I called EM).[73] I will not deal with such claims in this chapter, but I will have something to say about them in the next. Second, Noë sometimes claims that the phenomenal nature of certain perceptual experiences is determined in large part by our capacities for action. For example, Noë sometimes claims that the phenomenological differences between tactile experience and visual experience can be explained in terms of the differences in the kind of perceptual skills on which these two different kinds of experience depend.[74] Again, I will not focus on claims of this sort here. I will ask whether action (or a certain kind of action-involving capacity) is in some way necessary for perception (or for certain perceptual capacities). There are lots of different possible formulations of the claim that action is in some way necessary for perception, many of which are discussed in what

[73] Or at least some of Noë claims, if true, can be taken as considerations that speak in favour of EM.
[74] See, for example, Noë (2004, p.27). For a critical discussion of claims of this sort, see Clark and Eilan (2006).

follows. In this section, I deal with the strongest (and in my view least plausible) formulations.

In his introduction to *Action in Perception*, Noë writes that '[T]he main idea of this book is that perceiving is a way of acting. Perception is not something that happens to us, or in us. It is something that we do' (2004, p.1). On the strongest (and probably least charitable) reading, this is a conceptual claim about the nature of perception. To see what this reading would amount to, let's first consider an analogy. If I say perception is essentially experiential, I might mean that as it happens, our perceptions are always accompanied by phenomenology – perceptions are always accompanied by certain feelings. Let's call this a 'contingent generalisation' – a generalisation about perceptual experiences undergone by humans (and perhaps other animals), which is taken to be true as a contingent matter of fact.

Alternatively, I might be making the slightly stronger claim that given the way the laws of nature are, perception is necessarily experiential. This would be to say that experience is nomologically necessary for perception (where 'nomologically necessary' here means 'necessary, given the laws of nature'). But I might be making an even stronger claim: a conceptual claim according to which a mental state/process only counts as a perceptual state/process if it is experiential. This conceptual claim would have the modal consequence that a philosophical zombie (a being who had no conscious experiences) would not count as a perceiver, even if he/she was fully capable of making accurate reports about his/her environment on the basis of visual information, walking around without bumping into things, and so on. The weaker claims introduced above do not have this consequence. Let's start with the contingent generalisation. One might think that, as it happens, human perception is always experiential, but nonetheless maintain that philosophical zombies would still count as perceivers. Next, the nomological necessity claim has the consequence that zombies are not possible given the laws of nature, but it is consistent with the idea that there is a possible world in which the laws of nature are different and in which there are zombies who count as perceivers.

Likewise, we can distinguish a weak and a strong version of the claim that perceiving is a way of acting. We might make the contingent generalisation that, as

it happens, human perception is a form of action. We could make this claim while still allowing for the possibility that there could be a perceiver for whom perception was an entirely passive process (where passive here just means wholly inactive). And we could even admit that this possibility is consistent with the laws of nature. Next, we could make the slightly stronger claim that perception is, by force of nomological necessity, active; given the laws of nature, perception must always be active. Finally, we could claim that perception is active by force of what I will call *metaphysical necessity*. To make this claim is to say that there is no possible world in which perception is an entirely passive process.

To tease out the modal consequences of this metaphysical necessity claim, let's consider a case that might be taken as counterexample to it. The counterexample I have in mind is the 'sentient trees' example used by Joel Smith (forthcoming) to show that certain claims about the relation between perception and action are false.[75] Imagine a world in which the laws of nature are different from ours and in which a species of sentient trees exists. These trees are incapable of action. Indeed, they have no concept of action and no concept of how action might relate to perception. Yet they are capable of perceiving. They have perceptual experiences with phenomenology and content not unlike ours. And their perceptual experiences inform their beliefs in much the same way that ours do. If the existence of such creatures really is metaphysically possible, then their possibility constitutes a counterexample to the claim that action is somehow metaphysically necessary for perception. But as I point out in a response to Smith (forthcoming), such cases do not constitute a counterexample to the claim that action is somehow nomologically necessary for perception. Sentient trees may be conceivable, and conceivability may be a good guide to *metaphysical* possibility, but it is not a good guide to *nomological* possibility (possibility given the laws of nature). We can conceive of something moving faster than the speed of light, but our ability to do this should not tempt us to conclude that it is possible for anything to do so. Finally, as should be obvious, the metaphysical possibility of sentient trees would not threaten any contingent generalisation about any relation between action and perception that happens to obtain in all humans (or even all terrestrial animals).

[75] See Smith (forthcoming) for the philosophical provenance of counterexamples of this type.

So how should we treat Noë's claims about the relation between action and perception? His claim that perception is a way of acting is neutral between the three options I have sketched. It could be a claim about how perception happens to be in all humans, or a view about how perception must be, given the laws of nature. Or, finally, it could be a claim about the way perception must be, in any possible world. As far as I know, Noë doesn't take an explicit stand on this question, but I think we can infer from his argumentative methodology that he is trying to make a nomological necessity claim. While Noë does not consider cases like sentient trees as potential counterexamples to his view, he does consider paralysed and quadriplegic perceivers as potentially threatening to his thesis (2006, pp.9-13). Indeed, his general methodology, as we will see in later sections, involves leaning heavily on certain empirical results regarding perception in real-world cases. And he rarely if ever presents his claims as falling out of the very concept of 'perception'. Given this, it would not be charitable to read Noë as making metaphysical necessity claims or conceptual claims.

But he seems to be going for something stronger than a contingent generalisation. As well as considering actual, real-world cases, he considers non-actual but nomologically possible cases too. For example, he considers the possibility that a scientist might be able to generate perceptual experiences purely by manipulating a person's brain, and ends up arguing that such a scenario is not in fact possible (2009b, pp.176-7). The fact that he takes on such a task suggests that he is concerned not only with making claims that happen to hold true of all perceptual experiences undergone by sentient terrestrial creatures; he seems to be making claims about the conditions under which perceptual experience is possible, given the laws of nature. Given this, I will concern myself in what follows primarily with different ways of cashing out the idea that action (or some capacity for action or some conceptual grasp of action) is nomologically necessary for perception.

§1.2: Is Action Nomologically Necessary for Perception?[76]

I will call '*Act Necessity'* the view that action is nomologically necessary for perception. More precisely, *Act Necessity* is the view that each and every 'act' of

[76] I will henceforth assume that all the necessity claims I discuss are *nomological* necessity claims.

perception is dependent on some accompanying action.[77] While this is probably not Noë's considered view, it is worth discussing briefly as some of the things he writes might suggest that it is. For example, Noë might seem to endorse *Act Necessity* when he writes that 'we *enact* our perceptual experience; we act it out' (2004, p.1) and that 'perceiving is a kind of skilful bodily activity' (2004, p.2). While these statements do suggest something like *Act Necessity*, they are probably most charitably read as eye-catching slogans that do not do justice to the nuances of Noë's view. We will soon be looking at more plausible formulations of Noë's view, but it is worth quickly pointing out what is wrong with this formulation.

A seemingly compelling, but in fact flawed objection to *Act Necessity* is given by Jesse Prinz (2009). He points out that paralysis of ocular muscles does not prevent people from having conscious visual experience, writing that 'Land, Furneaux and Gilchrist (2002) describe an individual who has relatively normal vision despite the fact that she had life-long congenital ocular fibrosis, which prevents her eyes from moving' (2009, p.428). It might seem that this is a counterexample to *Act Necessity*. After all, one would think that someone incapable of the actions most commonly associated with vision – eye movement –should, according to *Act Necessity*, be incapable of seeing. But careful reading of the paper cited by Prinz shows that the patients with ocular fibrosis use unusual head-movements to compensate for their inability to perform eye-movements. Specifically, they perform head movements that resemble saccades and fixations. In the abstract of the paper, for example, Land et al. write that the subject studied 'made saccades with the head' (2002, p. 80). Given this point, Prinz's proposed counterexample poses no problems to *Act Necessity*. While it might seem plausible to think that people with ocular paralysis are incapable of performing the kinds of actions usually associated with visual exploration, we have reason to think this is not the case. The *Act Necessity* claim only requires that visual perceivers are capable of performing actions relevant to visual exploration. It might initially seem plausible to think that only eye-movements can satisfy this requirement. But the compensatory head movements performed by people with ocular fibrosis are clearly actions relevant to visual exploration.

[77] This is distinct from the weaker view that a subject who had never acted in their life would, by virtue of this fact, be incapable of perception.

In a more successful attempt to undermine *Act Necessity,* Kevin O'Regan writes that although perception is often accompanied by some skilful action, like eye-movement, 'there are cases, as when you recognise something in a flashed display, where eye movements are not used. There need not be any covert eye movements either […]' (2011, p.67). O'Regan doesn't cite any particular study to back up this claim, but it is likely that he had in mind something like the literature on subitization.[78] The subitization literature demonstrates that subjects can effectively enumerate displays of up to 4 dots, even when they are presented for a moment so brief that eye-movement, or any other movement that could be construed as relevant to perceptual exploration, is impossible. Given this fact, it seems that some perception is indeed possible even in the absence of eye-movement, or other exploratory movements, meaning that *Act Necessity* is false.[79]

Nivedita Gangopadhyay (2010, p.403) offers an ultimately unsuccessful rebuttal to this line of argument:

> When the reader fixates on the dot in the figure her eyes are *not* paralysed, the gaze is merely at a point. It is an unfortunate, and somewhat widely held, misconception that during fixation the eyes do not move at all. The studies by Martinez-Conde, Macknick, and Hubel (2004) and Martinez-Conde, Macknick, Troncoso, and Dyar (2006) clearly dispel the myth of complete absence of eye movements in fixation. Martinez-Conde et al. (2004) and Martinez-Conde et al. (2006) discuss in detail three types of movements that persist in fixation, namely microsaccades, drifts and tremors. Findlay and Gilchrist (2003) point out that the movements during fixation may not be involuntary as revealed by the studies of Ditchburn (1973), Steinman, Cunitz, Timberlake, and Herman (1967) and Kowler and Steinman 1979).

The problem with this argument is that not all movements count as action; the beating of my heart, for example, or the contraction of my arteries, does not count as action. Bodily movements must be voluntary, or at least to some degree under the control of volition if they are to count as actions. Gangopadhyay does address this point, in a manner, by citing Findlay and Gilchrist (2003) in support of the claim that eye-movements during fixation 'may not be involuntary'. This citation of

[78] For an example of a subitization study see Simon and Vaishnavi (1996)
[79] Aizawa (2007) develops a line of argument similar to this.

a whole monograph without any page-numbers is unhelpful, but the passage Gangopadhyay seems to have in mind goes as follows:

> The original use of 'involuntary' for these fixation eye movements was shown to be inappropriate by the demonstration of some higher-level input into the fixation mechanisms. The incidence of microsaccades could be changed with instructions (Steinman et al. 1967) and directed drift movement was found to occur in anticipation of a subsequent target following a saccade. (2003, p.14)

Let's first take the claim that the incidence of microsaccades 'could be changed with instructions'. I assume this means that subjects can, with instruction, either increase or decrease the rate at which their eyes microsaccade. Suppose this is true. Does it follow that such microsaccades are voluntary? It might be possible for me to change my heart rate with instruction, but it does not follow that the beating of my heart is a voluntary action. Next, let's take the claim that 'directed drift' happens in advance of directed saccades. Let's assume that directed saccades are voluntary. Does it follow that the 'directed drift' that precedes these saccades is also voluntary? Before I move my arm in a certain manner, certain motor neurons fire in my brain. Does it follow that these neural firings are voluntary? Not obviously. My arm-movement is voluntary, but it does not follow that everything that is associated with that arm-movement is also voluntary.

Notice also that even if we accept the idea that microsaccades and drifts might, in some special circumstances, be brought under voluntary control (and hence count as voluntary actions), it does not follow that *all* microsaccades and drifts are therefore voluntary. This is important, because the subitization case described above is a case where it is *extremely* implausible to think that any microsaccades and tremors that occur are voluntary. Suppose you are put in front of a blank screen, and up flashes a number of dots. You see that there were three dots. Did you also make some decision about how many microsaccades you would perform while the dots were on display? Do you know how many microsaccades you performed? Could you have refrained from performing those microsaccades if you had wanted to? It seems that 'clearly not' is the answer to all these questions. Apart from anything else, there would have been no time for any such decision to be made, or for any kind of voluntary control to be exercised over such matters.

Even if we suppose that, under certain circumstances, microsaccades and drifts can count as voluntary actions (a supposition that is anyway deeply implausible), it is extremely unlikely that those eye-movements that occur during subitization are going to count as under voluntary control in any sense. They are, therefore, not actions in the traditional sense.

§2: Is Sensorimotor Knowledge Necessary for Perception?

Early on in *Action in Perception*, Noë claims that sensorimotor knowledge is necessary for perception, writing that: 'For mere sensory stimulation to constitute perceptual experience – that is, for it to have genuine world-presenting content – the perceiver must possess and make use of *sensorimotor knowledge*' (2004, p,10). The claim is that, without sensorimotor knowledge, we would have sensory stimulation, but this sensory stimulation wouldn't tell us anything about the world. It would be mere sensation, or something like that. At this point in the book, Noë seems to be conceiving of sensorimotor knowledge as an understanding of the way sensory stimulation we receive varies as a function of movement.[80] Noë supports the claim that sensorimotor knowledge is necessary for perception by citing experiments in which sensorimotor knowledge has in some way been disrupted, or is for other reasons absent. He claims that in these cases, subjects experience sensory stimulations, but these stimulations do not amount to perceptions.

Talking of the subject S.B., Noë writes that 'S.B. lacks understanding of the sensorimotor significance of his impressions; he lacks knowledge of the way the stimulation varies as he moves or would move. As a result, or so I propose, his impressions are without content and he is, to a substantial degree, blind' (2004, p.10). The idea behind this argument is that S.B. is subject to perfectly normal sensory stimulation but, because his sensorimotor knowledge has been disrupted, he is unable to see. To see why this might be the case, we need to examine the case of S.B in more detail. S.B. was a patient who had recently undergone a cataract operation. Noë cites evidence suggesting that, when patients have their cataracts removed, they do not regain their vision right away. First, the patients experience a 'blur' or a mix of light, colour and movement, but it takes a little while before they

[80] As we will see, it is not quite clear exactly what sensorimotor knowledge amounts to on Noë's theory. But this brief statement should do for current purposes.

start to actually see objects in the normal sense. Noë wants to explain this temporary visual impairment by arguing that the patients take a while to understand the 'sensorimotor significance' of the sensory stimulations they are receiving. The idea is that the sensory stimulations they are receiving are perfectly normal, so they should be able to see normally. But, since the patient has not seen for a long time, they may have lost some of their understanding of the way sensory stimulation varies as a function of movement. The lag between the removal of the cataracts and the return of normal vision might be explained by the fact that they need to relearn the appropriate sensorimotor knowledge before they can see again.

As Noë acknowledges, the evidence from cataracts patients does not supply very strong support for his theory. After all, there are all sorts of explanations for the patients' temporary inability to see. It might be that 'inactivity of retina and visual cortex could lead to some degree of stunting of the development of neural connections needed for mature adult vision' (2004, pp.6-7). If Noë wants to find empirical support for his claim that sensorimotor knowledge is necessary for perception, he'll have to find cases of 'experiential blindness' that do not so easily admit of alternative explanation. [81] Noë acknowledges this point, but thinks he has identified such cases. That is, he thinks there are cases where the inability of subjects to perceive can only be explained by the fact that their sensorimotor knowledge is in some way deficient. For the rest of this section, I will examine these putative cases.

[81] Gangopadhyay disagrees with this analysis, arguing that cataract cases provide strong evidence for enactivism. She writes that '[o]bservation of abnormal eye movements in post-operative patients supports the hypothesis of breakdown of sensorimotor anticipation and sensorimotor/cognitive processing rather than a purely sensory deficit as the underlying cause of their visual impairment [… the post-operative patients] report a lack of perceptual content because they are unable to deploy exploratory sensorimotor behaviour and exhibit a breakdown of sensorimotor/cognitive processing' (2010, p.402). The argument starts with the observation that post-operative patients, who are still unable to see, normally exhibit abnormal eye movements. Gangopadhyay then infers from this observation that it is the patients' inability to perform normal eye-movements that explains their inability to perceive properly. This inference is shaky, because the order of explanation could easily go in the other direction. It is possible that the patients' inability to perceive properly renders them unable to perform normal exploratory eye-movements. Indeed, it is difficult to see how cataract patients, who are initially unable to perceive objects, could *possibly* perform normal eye-movements; performing normal eye-movements requires one to fixate on and track salient objects. If you cannot see objects, this exploratory feat will obviously be impossible. So the opponent of Gangopadhyay, who does not think that the cataract patient's inability to perceive should be explained in terms of a breakdown of sensorimotor capabilities, need not be in the least bit embarrassed by the finding that these patients exhibit abnormal eye-movements. Any sensible person would predict that someone who is unable to perceive will be unable to perform ordinary eye-movements.

Noë first calls upon experiments involving inverting goggles in support of his idea that sensorimotor knowledge is necessary for perception. Inversion goggles are fitted with prisms that invert the light passing through them. This means that, when I am wearing the goggles, light that would usually hit the right side of my retina will now hit the left side. So an object which is off to my right, and which was in the right hand side of my visual field before I put on the goggles, will look like it is off to my left once I have put on the goggles. That is, the goggles will essentially have the effect of inverting my visual field so that things that previously looked as if they were off to the left will look like they're off to the right, and *vice versa*. But, Noë claims, although we might expect the goggles to have this effect, they actually do something far more dramatic.

He claims that '[t]he initial effect of the inverting glasses of this sort is not an inversion of the content of experience (an inversion in what is seen) but rather a partial disruption of seeing itself' (2004, p.8). In support of this claim, Noë cites the report of subject K, who 'wrote of his initial experiences in Kohler's experiment with displacing spherical prism spectacles' (2004, p.8). Subject K describes how 'the most familiar forms seem to dissolve and reintegrate in ways never before seen' and how 'I felt as if I was living in a topsy-turvy world of houses crashing down on you, of heaving roads, and of jellylike people' (2004, p.8). Noë concludes that '[K] receives exactly the stimulation he would receive were he looking at an object in a different spatial location without the inverting lenses. The inability to see normally stems not from the character of the stimulation, but rather from the perceiver's understanding (or rather failure of understanding) of the stimulation' (2004, p.8). The idea is that K receives perfectly regular sensory stimulation, but he fails to understand how this stimulation changes as a function of movement. For example, suppose there is an object off to K's right. K will see this object as off to the left. If he tries to turn his head towards it, he will turn his head to the left. But this will achieve the opposite effect, since the object is actually off to the right. This is an example of the way in which K's sensorimotor knowledge has been disrupted. The idea is that the inversion goggles only disrupt sensorimotor knowledge, but leave unaffected all other perceptual mechanisms. Therefore, the loss of visual capacity can only be explained by the partial disruption of sensorimotor knowledge.

So far so good, but the problem is that Noë completely misdescribes the effects of inversion goggles. The effects described by K simply do not occur when one wears the goggles. I have tried them, and experienced no such effects, as have many others – see Joel Smith (forthcoming) for further testimony to this effect. Noë is simply mistaken. In fact, the source of his mistake becomes obvious if we look closely at the passages I quoted in the last paragraph. Noë introduces K as a subject who wore '*displacing spherical prism* spectacles', but then later says that K 'receives exactly the same stimulation he would receive were he looking at an object in a different spatial location without the *inverting* lenses' (emphases added). The problem is that inverting lenses and displacing spherical prism lenses are not the same thing. Looking through the spherical prism lenses is more like looking at your reflection in a spoon. As the name suggests, the lens contains a sphere which distorts the light passing through it, creating a bulge in the visual field. For this reason, it is not true, as Noë writes, that 'K receives normal stimulation' (2004, p.8). Given that K's sensory stimulation is massively distorted, it is no longer true that the only disruption to K's visual system is the disruption of sensorimotor knowledge. For this reason, Noë's argument from 'inversion' goggles fails.

Another line of evidence upon which Noë relies relates to the TVSS device invented by Paul Bach-y-Rita, which allows blind subjects to 'see', using a camera transduced to a series of vibrating nodes (which are attached to some part of the blind subject's body). Using this device, blind subjects are able to discern features of their environment. As Noë reports, they are able to 'make judgements about the number, relative size, and position of objects in the environment' (2004, p.26). They are also able to perform facial recognition and accurately judge the speed of moving balls (Bach-y-Rita, 2004, p.86). But, as Noë and O'Regan report,

> In the earliest trials with the TVSS device, blind subjects generally unsuccessfully attempted to identify objects in front of the camera, which was fixed. It was only when the observer was allowed to actively manipulate the camera that identification became possible and observers came to "see" objects as externally localised (White et al. 1970). This important point constitutes empirical verification of the mainstay of the present theory of visual experience, namely, that seeing constitutes the ability to actively modify sensory impressions in certain law-obeying ways (O'Regan and Noë: 2001, p.958).

The idea here is that the subjects can only 'see' using the TVSS device once they can actively move the camera around. Or, to put the claim less contentiously, the subjects can only identify distal objects on the basis of the stimulation they receive once they are allowed to control the camera attached to the TVSS system.[82] This is important, because it is only by moving the camera around that the subject can create the conditions in which it is possible to exercise sensorimotor knowledge. So, before moving the camera around, the subject experiences mere sensory stimulation. Once he/she moves the camera around, the sensory stimulations change as a function of movement. That is, they obey the laws of sensorimotor contingency. Once the sensory stimulations are obeying the laws of sensorimotor contingency, it is possible for the subject to exercise sensorimotor knowledge. It is this possibility that allows the subject to see, or at least to identify distal objects on the basis of the sensory stimulation they receive.

So, on this interpretation, the TVSS experiment provides evidence in favour of the claim that sensorimotor knowledge is necessary for perception. But notice that in the above quotation, Noë and O'Regan state that subjects were *generally* unable to identify objects in the fixed camera condition. This is an important caveat, since subjects did have some perceptual abilities even when they were not allowed to move the camera. As White et al. report, subjects who were asked to distinguish between circles, triangles and squares without moving the camera reached accuracy levels of 60% after 54 trials, an accuracy level way above chance (1970, p.24). It is true that, when the subjects were able to manipulate the camera, they reached 100% accuracy very quickly, and were able to perform much more complex tasks of object and facial recognition. So it looks like at most the TVSS experiments support a more modest necessity claim: the claim that sensorimotor knowledge is necessary for reliable and sophisticated perceptual capacities.

But how are we best to understand this necessity claim? Is it the case that, without sensorimotor knowledge, we merely experience meaningless sensations whereas, when sensorimotor knowledge is added to the mix, these sensations magically coalesce into perceptual content? This interpretation seems to be ruled out by the fact that the subjects of the TVSS experiment have some perceptual capacities

[82] This claim is repeated by Robert Briscoe (forthcoming).

even before they are able to exercise sensorimotor knowledge. It looks like there is a degree of continuity between the initial sensory stimulations and the more sophisticated perceptual content. Closer examination of White et al.'s (1970) results suggests a more plausible interpretation of Noë's necessity claim. White et al. report that the subjects allowed to manipulate the camera distinguished shapes by panning 'the camera so that the figure came into and passed out of the field, suggesting that the discrimination is largely based on contour changes in the leading edge of the figure' (1970, p.24). The idea seems to be that the contour changes specified by, say, circular shapes are far easier to pick up than the static contour specified by circular shapes. So once the perceiver is able to manipulate the camera, he/she can move it in a scanning motion that creates the kinds of contour changes that make shape-discrimination easy. In order to do this, the perceiver needs two abilities. The first is the ability to create the right kind of contour changes by creating the right kind of scanning motion. The second is to associate, say, circularity with a particular type of contour-change. This two-fold ability contains a motor dimension – the ability (perhaps the skill) of creating the right kind of scanning motion – and a sensory dimension – the ability to categorise different patterns of contour-change as specifying a certain shape.

But is it necessary that the subjects actually perform the scanning motion themselves? What if you set up a system that performed the scanning motion automatically, and simply told the test subjects how this scanning motion worked? Fortunately, examination of a different sensory substitution device can answer this question. The device in question is called vOICe. The vOICe software can be downloaded on android smartphones, meaning that anyone with the time and inclination can test it out themselves (as I have done). If you hold the camera of your phone up to a scene when the vOICe software is running, it will scan the scene presented on your camera screen every second. The scanner moves from left to right, and translates the visual information it picks up into auditory information. The scanner makes noises when it scans light (as opposed to dark) objects. The vOICe system then makes a noise when its scanner picks up light, and the character of this noise depends on the position of the light on the screen. Light picked up at the bottom of the screen causes the system to emit a low sound. The higher up on the screen the light is when it's picked up, the higher the pitch of the noise

produced. So suppose the device is pointed towards a dark screen on which there is a white diagonal line which goes up from bottom-left to top-right. As the scanner scans the screen, the system will produce a note that starts low and gets higher. If the camera is not moved, it will produce exactly the same noise every second, unless the display on the screen is changed.

Now it turns out that if you are told how this device works, you can learn to identify progressively more complicated objects, even if you have no control of the camera. Suppose the camera is fixed on a screen, on which progressively more complicated objects are presented. If you guess what is being presented, and are given feedback on whether you are guessing right, you can learn to identify progressively more complex objects. I have tried this myself and spoken to people who develop the technology and train others to use it – the upshot is that you don't need to control the camera in order to use the device. Of course, if the device is to be of any practical benefit to you, you will need to walk around with the camera strapped to you. And, if you do this, you will have to learn to distinguish between sound-changes caused by your own motion and sound-changes caused by events in the world. But the basic point is that you do not need to control the camera in order for the strange sounds it makes to tell you about the world.

Does this show that sensorimotor understanding is unnecessary for perception? Perhaps not quite. After all, you need to know that there is a scanner going across a scene from left to right if you are to learn how to use the device. The 'motor' component is still present because there is still a scanning action going on. It's just that the scanning doesn't have to be performed by the perceiver.[83] But, if we combine this evidence with the fact that people can identify simple objects with the TVSS device without any scanning at all, it starts to look like sensorimotor knowledge isn't really necessary for at least very minimal perceptual abilities. At most, in these cases, sensorimotor knowledge plays the minor role of allowing us to understand that scanning is going on so that we can identify complex objects. This falls far short of what Noë wants from sensorimotor knowledge in this context – his

[83] Note that, if we take the experience of trained vOICe users to be genuinely perceptual, this might provide more reason (if more reason were needed) for rejecting the *Act Necessity* claim discussed in the previous section. It looks like people can learn to perceive spatial properties through the device without any need for bodily action.

claim was that, without sensorimotor knowledge, all we can experience is meaningless sensation.

Note also that the vOICe device clearly falsifies another key claim made by Noë about sensorimotor knowledge. Noë writes that 'only through self-movement can one test and so learn the relevant patterns of sensorimotor dependence' (2004, p.13). If we take the relation between the movement of the vOICe scanner across the scene and the noises the device produces to be a 'pattern of sensorimotor dependence', it looks like it is just false to say that we need to perform 'self-movement' in order to learn this pattern of sensorimotor dependence. Noë might try to deny that this is an instance of genuine sensorimotor dependence, but if he does this, then it looks like sensorimotor understanding is in *no* sense necessary for learning to use the vOICe device (since there is nothing *else* even remotely resembling patterns of sensorimotor dependence going on in the case I described above). This would be an even more embarrassing result for Noë.

§3: Sensorimotor knowledge and Perspectival Content

Sensorimotor knowledge also figures in Noë's theory of perception as an explanation of our ability to see perspective-independent properties. An example of a perspective-independent property is the circularity of a plate. Whatever perspective I view the plate from, it will be circular. Its circularity is independent of perspective. By contrast, if I look at a plate from an oblique angle, it will have an elliptical perspectival shape. That is, there is a sense in which the plate will appear elliptical. Noë calls such an elliptical apparent shape a p-property (in this case, a p-shape). The p-properties visible to me at any given moment will depend on the particular spatial relation between me and the objects in my environment. Put another way, the p-properties visible to me vary with my perspective. Noë sets himself the task of explaining the fact that, even though the visible p-properties of objects change as I move about my environment, I am still able to see the invariant, perspective-independent properties of objects. So, even though the visible p-shape of a plate varies as I move about it, I am still able to see the plate as circular. Barring illusory cases, I can see the plate as circular from any perspective. And, according to Noë, I can do this while also seeing a non-circular plate-p-shape.

Noë claims that this perceptual feat is achieved because our perception of p-properties is accompanied by the right kind of sensorimotor knowledge. He writes that:

> We see its circularity *in* the fact that it looks elliptical from here. We can do this because we understand, implicitly, that circularity is given in the way *how things look with respect to shape* varies as a result of movement […] to see the actual size of a thing is to see how its perspectival size varies as we move (2004, p.84).

The idea can be put as follows. As we move about in our environment, the visible perspectival properties of objects change in certain ways. But the exact way in which an object's visible p-properties change as a function of movement is determined by the actual (perspective-independent) properties of the object. Consider a contrast between two plates. One is actually circular, whereas the other is actually slightly elliptical. You can imagine an experimenter setting up the plates in such a way that, viewed from a particular perspective, the visible p-shape is the same for both plates. But if the perceiver was then to move around the two plates, viewing them from different perspectives, the two plates would not present the same sequence of p-shapes. Noë's idea is that we implicitly understand how the visible p-shape of circles should change as a function of movement, and it is this understanding that allows us to distinguish circles from, say, ellipses. It is in this sense that to see the actual shape of a thing is to see how its perspectival shape varies as one moves. Formulated as a necessity claim, this idea might be put as follows: 'sensorimotor knowledge is necessary for the perception of perspective-independent properties'.

It is worth noting at this point that by making this necessity claim as well as the necessity claim discussed in §2, Noë is guilty of an equivocation. This point has been made by Taylor Carman, who writes that:

> Noë equivocates with respect to exactly what the 'sensory' component at work in sensorimotor contingencies amounts to. He sometimes calls the 'sensory' element 'stimulation'. At other points he seems to abandon that notion in favour of the idea that sensorimotor dependencies obtain between bodily movements and sensory 'appearances'. Yet according to his own view, stimuli and appearances are different kinds of things: stimuli are proximal causal ingredients in perception; appearances are supposedly phenomenally manifest relations in which we stand to objects (2009, p.636).

Carman's charge of equivocation can be reconstructed as follows. Noë claims that sensorimotor knowledge is necessary for perception. Sensorimotor knowledge is the practical understanding of sensorimotor contingencies. But Noë vacillates between describing these sensorimotor contingencies as 1) the way in which appearances (the way things look) change as a function of our movement and 2) the way in which the sensory stimulations we receive change as a function of our movement. Evidence for this equivocation (or conflation of terms) is easy to find in *Action in Perception.* For example, Noë writes that 'the sensorimotor profile of an object is the way its appearance alters when you move with respect to it (strictly speaking, it is the way sensory stimulation varies as you move)' (2004, p.78).

We are now in a position to see why Noë makes this conflation. The reason is that he is trying to make two necessity claims:

1) Sensorimotor knowledge is necessary for perception *simpliciter.*
2) Sensorimotor knowledge is necessary for the perception of perspective independent properties.

For 1) to be true, it has to be the case that, without sensorimotor knowledge, we would only experience non-perceptual sensory stimulation. The role that sensorimotor knowledge plays in 1) is getting us from non-perceptual sensory simulation to perceptual content that tells us about the world. The idea, roughly, would be that only when you come to understand how sensory stimulation varies as a function of your movement are you able to move from experience of mere sensory stimulation to full-blown perception. So, for 1) to be true, sensorimotor knowledge must be the understanding of the way in which (non-perceptual) sensory stimulation varies as a function of movement. But then when Noë argues for 2), he is trying to explain why it is that we see invariant properties despite the fact that, as we move, the visible perspectival properties of objects are constantly changing. He is trying to explain how we get past this constant flux of appearances to see the unchanging properties of objects. When he invokes sensorimotor knowledge to explain this fact, the idea is that sensorimotor knowledge is the understanding of the way the visible perspectival properties of objects vary as a function of movement. But for this explanation to work, the ability of an agent to perceive must be presupposed. Only when I enjoy perceptual content can I see the perspectival properties of objects, and

only when I see perspectival properties of objects can I understand how these properties vary as a function of movement.

There are two possible routes to take in trying to resolve this tension. One is to distinguish two different types of sensorimotor knowledge. One is the knowledge of how sensory stimulation varies as a function of movement, the other is an understanding of how appearances vary as a function of movement. Once these two types of sensorimotor knowledge are distinguished, Noë can appeal to the former in his argument for claim 1) and to the latter in his defence of claim 2). The second option is to drop either claim 1) or claim 2), and then refrain from postulating the type of sensorimotor knowledge that would have to be postulated in order to make that claim stick. For example, we could drop necessity claim 1) – the claim that sensorimotor knowledge is necessary for perception simpliciter. We have already seen in §2 that the evidence Noë marshals in defence of this claim is suspect. And the evidence from sensory substitution devices (particularly vOICe) might be taken to further undermine claim 1). Given this, and given the fact that claim 1) is generating tensions within Noë's theory, it might be best just to drop it.

But this is not the only option. Grush (2007) has argued that his emulator theory (which was discussed in the previous chapter) is capable of doing the theoretical work that Noë hoped sensorimotor knowledge would do. In later chapters, I will ask whether something like Grush's emulator theory might be used to bolster something like 2), but for now I want briefly to consider the possibility that it might in principle be used to buttress both 1) and 2). Recall that sensorimotor knowledge is knowledge of sensorimotor contingencies. And sensorimotor contingencies have something like the following form: if I perform such-and-such a movement, sensory input will change in such-and-such a way. And, as I will illustrate in the next chapter, emulators are perfectly capable of modelling counterfactuals of exactly this sort. The key point for now is that for emulation purposes, it doesn't really matter what form the 'sensory input' in the above counterfactual takes. Emulators are just systems that model certain counterfactuals. They just emulate the causal dynamics that would unfold in a given system, given certain initial conditions (or 'input'). There is no principled constraint on the kind of systems an emulator can emulate, or on the kind of input they can receive, as Grush

(2007) is at pains to point out. So it could be the case that we need a certain kind of sensorimotor knowledge, made possible by a certain kind of emulator, to get to perception *simpliciter*, and another kind of sensorimotor knowledge, made possible by another kind of emulator, to get to perception of perspective-independent properties. If this were the case, then the emulator theory would provide us a way of accommodating both 1) and 2). I am not saying that this situation is actual. I suggest only that it is an open possibility.

In later chapters, I will focus more closely on something like 2), attempting to improve upon Noë's story about how we get from perspectival to perspective-independent content. I will ultimately drop Noë's notion of p-properties and replace it with a different, though related understanding of perspectival content. I will nonetheless retain the idea that perspectival content, coupled with sensorimotor knowledge, gets us perception of perspective-independent properties. But this part of the story will have to wait. My reason for dropping Noë's p-properties story is that it is inconsistent with certain other aspects of his theory of perception, aspects I want to hold onto. But in order to argue for this inconsistency, I first need to say something about what these other aspects are. And that is the job of the next chapter.

Chapter 5: The Virtual Content Thesis

Abstract

In this chapter, I develop an alternative to what is sometimes called the 'filling in' hypothesis. The filling in hypothesis is basically the idea that the brain makes a series of educated guesses in order to convert relatively sparse information, registered by the retina, into an experienced visual field that seems rich in colour and detail. I develop an alternative to this hypothesis by refining Noë's promising but flawed 'virtual content' theory of perceptual presence. In its current form, Noë's virtual content theory is incoherent and phenomenologically implausible. And by the time I have rectified these shortcomings, the resultant theory will be quite a long way from Noë's original proposal.

§1: Introduction

We left off the last chapter wondering whether sensorimotor knowledge might be necessary for certain perceptual abilities. In this chapter, I examine another perceptual capacity that might require sensorimotor knowledge: the capacity to experience a full visual field. I start by introducing a puzzle concerning visual-field-perception, and then sketch out Noë's solution to this puzzle. As we will see, Noë's solution will require significant work if it is to be successful. Having shown why this is, and having made the necessary alterations to Noë's theory, I will finish by asking what role sensorimotor knowledge must play in the account I endorse. I then argue that the 's-representation' notion can help us understand how sensorimotor knowledge might perform the explanatory function required of it.

§2: The Puzzle of the Visual Field and the 'Filling in' Hypothesis

A body of empirical findings appear to show that the sense we have of our visual field is in fact illusory.[84] It seems that we experience a visual field that is richly coloured and fairly detailed all the way out to the periphery. It does not seem to us that the parts of the visual field to which we are not attending are severely sapped of colour and sparse in detail. But it turns out that only the central region of the retina is capable of registering detailed information about colour and form. This might seem to imply that we should see in rich colour and detail only those things at

[84] I use the term 'visual field' to mean something like 'ordinary visual phenomenology' – whatever you take yourself to experience when you open your eyes. I will proceed by first making some claims about what I take to be the character of ordinary visual phenomenology, and then trying to construct a theory that does justice to that phenomenology.

which our eyes are pointing directly. The rest of the visual field should be seriously lacking in colour and detail. But this is not how the visual field seems to us – it seems pretty colour-rich and detailed all the way out. It might seem, therefore, that the sense we have of our visual field is a trick played on us by our brains; the brain is essentially enriching the information received at the retina, and thereby giving us the false impression that we are, at any moment, picking up far more information from our visual environment than we actually are.

And there is additional experimental evidence that might seem to support this conclusion. A number of recent experiments have established that, in certain circumstances, experimental subjects fail to notice dramatic changes and events in their environment. As O'Regan has shown in a number of 'change-blindness' experiments, it is quite easy to render perceivers incapable of noticing dramatic changes in images to which we are attending directly (2011, pp.55-9). For example, if an image presented on a computer screen flickers slightly at the moment when a dramatic change is made, we will not notice the change. In a demonstration of a related phenomenon, called inattentional blindness, Simons and Chabris (1999) asked subjects to watch a video in which two teams, one team dressed in white, the other in black, passed a basketball around. The subjects were asked to count the number of passes made by the white team. But while the video was playing, a man dressed in a gorilla outfit walked into the centre of the court, did a little dance, and then walked off again. A surprising number of people failed to notice the gorilla. These cases might again be taken to show that our sense of the visual field is in some sense illusory. Our visual experience seems to us far more informative than it is. Our retinas are capable of registering far less detail than we take ourselves to perceive, and we are surprisingly incapable of noticing major events and changes occurring right before our eyes.[85] Let's call this problem the 'problem of retinal paucity'.

A natural way to explain this discrepancy would be to hypothesise that the brain somehow fills in a huge amount of the visual scene in order to compensate for the sparse input actually picked up by the perceiver, thereby giving perceivers the illusory impression that they are registering more detail than they actually are (see

[85] For a dramatic illustration of a similar effect, see Simons and Levin (1998).

Palmer, 1999, p.617). A natural way of cashing out this idea is to suggest that much of the content we take ourselves to experience visually has been augmented with, or filled in by, representations cooked up by the brain. Noë's 'virtual content' thesis aims to provide an alternative to this representation-invoking explanatory strategy (2004, p.47).

§3: Virtual Content and Perception as Access

Noë's alternative to the filling in hypothesis is a development of an idea found in the work of Daniel Dennett (1969, 1991) and Marvin Minsky (1985). The basic idea is that we have quick and easy access to any object in our visual field, and this fact explains the sense we have that the objects in our visual field are perceptually present. According to Dennett, the fact that we have quick and easy access to visual information about objects in our visual field creates in us the misleading impression that this information is present all of the time. He claims we fall into an 'introspective trap' (1991, p.360) – a version of what Minsky calls the 'immanence illusion', which Minsky formulates as follows: 'Whenever you can answer a question without noticeable delay, it seems as though that answer were already active in your mind' (1985, p.155). So, in the visual case, whenever we look for some piece of information in our visual field, we find it. And the fact that we do so creates in us the misleading impression that it was present to us all along.

Noë's view is similar but not identical to this. Noë agrees that the impression we have that visual objects are present to us is often explained by the fact that we have quick and easy access to them. But he does not think this impression is misleading. For Noë, 'it is no part of our phenomenological commitments that we take ourselves to have all [the] detail at hand *in a single fixation*' (2004, p.57). Rather, according to Noë, we take ourselves to have visual contact with a world whose detail we can readily access as and when it is needed. So, for Noë, we are not misled in this respect about the character of our visual experience. We take ourselves to have ready access to the detail visible in our visual field, and we are right to do so.[86] On this account, there is no introspective trap. We do not think, wrongly, that all the detail in our visual field is present to our minds all the time.

[86] For Noë's account of the distinction between his own and Dennett's position, see Noë (2004, pp.55-8).

Rather, we rightly take it to be the case that all the detail in the visual field is present to us – it is present in the sense that it is easily accessible by us. Noë uses the term 'virtual content' to refer to content that is present to the perceiver by virtue of being accessible to the perceiver.

§4: If Virtual Content is Accessible Content, What is Access?

I want to follow Noë in denying that virtual presence is illusory presence, but clarification is in order at this point. Noë's view is that many of the objects in our visual field are only *virtually* present to us – that is 'present to perception as accessible' (2004, p.63). But to say something is accessible is to say that it can be accessed. So to know what it means for something to be accessible, we need to know what it is for something to in fact be accessed. Noë is vague on this score. He talks of what is 'at hand', what is 'occurrent' to consciousness and what can be 'embraced' in consciousness (2004, p.57, pp.134-5). According to the view I will be advancing, we should think of accessing some object or property as making it the focus of one's conscious visual awareness. I take it that focal awareness is a phenomenologically recognisable notion – when I focus my attention on some visible object or property, that object or property is the focus of my awareness.

When we think of access in this way we can, accordingly, think of visually accessible objects or properties as those which can easily be made the focus of our visual awareness (more will be said in §8 about how we should cash out 'easily' in this context). Note that going down this route precludes us from accounting for virtual content in a way that reduces it to something non-conscious. But a reductive analysis of this sort is not my aim. My aim is to account for peripheral content (content which should, given the nature of our retinas, be sparse in detail) without appeal to 'filling in'. And we should in principle be able to do this by explanatory appeal to what might be called 'focal content' (visual content that is the focus of our visual awareness). When accounting for focal content, we don't have to confront the problem of retinal paucity; the retina is fully capable of registering the detail presented by objects that are the focus of our visual awareness. So if we can somehow explain peripheral content in terms of focal content, then we will have explained the problematic in terms of the unproblematic, thereby finessing the problem of retinal paucity.

It is important to notice at this point that the proposal I have just put forward is flatly incompatible with Noë's notorious claim that content is virtual *all the way in*. Let's take a look at this idea. Noë asks us imagine focusing our visual attention on a tomato, inviting us to find plausible the following claims:

> [Y]ou can no more embrace the *whole* of the [tomato's] facing side all at once in consciousness than embrace the *whole* tomato in consciousness all at once. This is clear on reflection I think […] This shows that we cannot factor experience into an *occurrent* and a *merely virtual* or potential part. Experience is fractal in this sense. (2004, pp.134-5)

And he later writes:

> When you peel away the layers of potentiality and merely virtual presence, you are not left with pure phenomenal content, that which, as it were, is present to your mind now. You are always presented with qualities that in turn have qualities and that are presented against a structured background. (2004, p.216)

The rough idea is that all content is only virtually, or potentially present, because no object or property can be entirely present, all at once, to consciousness; no content is 'present' or 'occurrent' in its entirety to consciousness, so all content is virtual. Content is virtual *all the way in*. Two objections. First, you cannot explain something's presence in terms of its accessibility, define access in terms of occurrence, and then claim that nothing is occurrent. To do so is to commit oneself to the incoherent claim that content is accessible, but cannot in fact be accessed. If something cannot in fact be accessed, there is no meaningful sense in which it is accessible. The point can be put as follows. Because virtual content just is accessible content (2004, p.63; 2009a, p.476), the claim that content is only ever virtual is just the claim that it is only ever accessible (and never in fact accessed). This claim looks incoherent, for the reasons just stated. Second, Noë's motivation for claiming that nothing can in fact be accessed is the claim that experience is fractal – no object or property can be grasped in consciousness in its entirety. But he gives no good reason for thinking that the following entailment holds: if you can't grasp something in its entirety then you can't grasp it at all.

To see how implausible Noë's two claims are, suppose I told you that the coins in my pocket are accessible to me. You nod in agreement and suggest that they are accessible in the sense that I can, whenever the mood takes me, reach in and grab any coin I want. Not quite, I respond. It's not *actually* possible for me to

reach in and grab them. At this point you struggle to see how the coins can be accessible in any meaningful sense of the term. You say so and ask, out of interest, why the coins cannot be grasped. I respond that the coins have a special, fractal surface structure, and this structure makes it impossible for me to wrap my whole hand around any given coin all at once. On hearing this response, the natural thing to think is that I am confused on both points. The sensible thing to think would be that the coins *are* accessible – accessible in the sense that they can be grasped at any time. They *can* be grasped even though it isn't possible to grasp every part of any of them all at once.[87]

And this, I suggest, is what we should say about the tomato and the tomato-face. The tomato-face is accessible to consciousness in the sense that it can be accessed by consciousness. And it can be accessed because it can be made the focus of conscious awareness. And this is true even if I can't focus on every single aspect of the tomato-face all at once. We can accommodate *this* idea by claiming that one can access a visible object or property without accessing it in its entirety, all at once.

§5: Sufficiency and the Fine-Grained Dependence Desideratum

Jesse Prinz develops a counterexample designed to undermine the idea that accessibility is sufficient for visual presence. He notes that when I am watching TV, I can easily access other channels; all I have to do is press a button on the remote controller (2009, p.423). But these other channels are not visually present to me. It would seem to follow, then, that accessibility is not sufficient for something to figure in one's visual experience in the way that the periphery features in our experience. Noë elaborates the accessibility thesis in more detail than Dennett, and his elaboration allows us to see how Prinz' objection might be met. Noë stipulates that a set of conditions must be met if an object is to be perceptually accessible.

These conditions are not just cooked up to deal with tricky counterexamples, but are, Noë claims, designed to capture an important set of insights. First, Noë credits the causal theory of perception with capturing the key insight that 'experience depends, in a fine-grained, counterfactual-supporting way on things

[87] Michael Martin (2008) disagrees with Noë on similar grounds, appealing to Thompson Clarke's (1965) claim that one can nibble at a block of cheese by only nibbling at a part of it. But Martin does not attempt to rescue the virtual content idea from Noë's excesses.

being thus and so' (2009, p.477).[88] But he aims to go further, by developing conditions that capture the fact that 'how things look must depend (in a suitably find-grained counterfactual-supporting way) not only on how things are, but on one's relation to how things are' (ibid). The basic idea is that how things look depends not only on how things are, but on how I am situated with respect to the things seen. I aim to develop a virtual content theory which meets these requirements and which can deal with counterexamples of the sort that Prinz develops. Indeed, I aim to develop a theory that does these jobs better than Noë's theory does. But first let's see how Noë's proposal is supposed to work.

§6: Noë's Requirements for Virtual Content

Noë's conditions are as follows:

> [A]n object or quality is perceptually present (i.e., it is an object of perceptual consciousness) when the perceiver understands, in a practical, bodily way, that there obtains a physical, motor-sensory relation between the perceiver and the object or quality, satisfying two conditions:
>
> (i) Movement-dependence: movements of the body manifestly control the character of the relation to the object or quality
> (ii) Object-dependence: movements or other changes in the object manifestly control the character of the relation to the object or quality.[89]
>
> In short, an object or quality is present in perceptual experience when it is perceptually available (2009a, p.476).

And

> An object is perceptually present when our access to it depends on *sensorimotor* skills (2009a, p.480).[90]

Conditions (i) and (ii) describe a relation that must obtain between a perceiver and an object or property if that object/property is to be accessible. Before discussing the two conditions in detail, let's first look at the suggestion that if this accessibility relation is to obtain, the perceiver must have some sensorimotor understanding, or

[88] The causal theory Noë has in mind here is that developed by Peter Strawson (1979).
[89] These conditions were first introduced in *Action in Perception* (2004, p.64), but I will work with the more recent and fully worked out formulations (2009a). These more recent formulations are repeated verbatim in Noë's *Varieties of Presence* (2012, p.22).
[90] While Noë seems to be talking here about the conditions that must be satisfied for perceptual presence in general, I am only concerned with how these conditions apply to visual presence.

skilful sensorimotor grasp of the fact that the two conditions hold. Noë is a little vague about whether this sensorimotor grasp amounts to skill or understanding (or somehow both). Let's use the term 'sensorimotor grasp' as neutral between both. In §9 I say something about how sensorimotor grasp should be conceptualised, making use of resources developed earlier in the thesis, but this part of the story can wait for now.

Noë's basic idea is that when (i) and (ii) obtain, and when sensorimotor grasp obtains, the object/property whose relation to the perceiver is described by the two conditions is perceptually 'available' (or 'accessible', in the terminology I prefer). Noë's thought, expressed at the end of the above quotation, is that our access to virtual content depends on sensorimotor grasp – so sensorimotor grasp is a precondition for accessibility. But, as we have seen, Noë doesn't actually think that 'access' ever obtains. Given this, it isn't clear what we should make of the claim that access depends on sensorimotor skills. 'Accessibility' only makes theoretical sense if we have some idea of what 'access' amounts to. If this is right, then we need to recast Noë's theory in order to make room for the idea that virtual (accessible) content can actually be accessed.

To do this, I suggest that we start with the focal awareness account of access I sketched in §4. On this view, to access some object or property is to make it the focus of one's visual awareness. If we start with this view of access, and retain Noë's idea that sensorimotor grasp is that on which access depends, we get the view that sensorimotor grasp is just that set of abilities which is necessary for making an object/property the focus of one's visual awareness in a given situation – the set of abilities necessary for accessing visible detail.[91] By doing this, we can render coherent the idea that virtual content is accessible content.

But achieving coherence isn't the only challenge. We also have to deal with sufficiency worries of the sort raised by Prinz. I can visually access all sorts of things (like the other channels while I am watching TV) but it does not follow from this that these sorts of things are perceptually present to me before I have accessed them. So we must find some way of placing constraints on the kinds of accessibility

[91] For now, we can remain neutral on the question of whether the abilities definitive of sensorimotor grasp are conceptual, skilful, or whatever. I deal with issues of this sort in §9 below.

relations that must obtain if an object is to be virtually present. We must do so in a way that generates phenomenologically plausible results. And we must motivate these constraints by means of independent theoretical considerations (like those listed in §5) if we are to avoid being accused of making *ad hoc* moves just to save the virtual content theory. Noë's conditions (i) and (ii) are designed to do just this job. But, as I will argue, Noë's conditions must be altered considerably if they are to do all the work required of them. Once I have performed all the necessary alterations, Noë's conditions (i) and (ii) will be transformed into a different set of conditions – conditions (i)*, (ii)* and (iii)*. I will argue that this altered condition-set is able to satisfy the criteria outlined in §5.

§7.1: A First-Pass Reconstruction of Noë's Conditions

Since I will be reconstructing (i) and (ii) so that they fit with my own theory of access, I am moving away from straightforward exegesis of Noë. Whereas the first condition will need only minimal reconstruction, the second will need considerable work. My reconstruction of the first condition goes as follows:

(i)* If some object or property is to be virtually present, it must be the case that I can, at will, bring that object/property into focal awareness.

To see how this repurposed condition works, suppose there is a cup in my visual periphery. By moving my eyes in the appropriate way, I can bring the cup into my focal awareness, and thereby access it. I can also focus on the cup's visible properties – its colour, for example, or its shape. So to say that the cup, or its visible properties, satisfy condition (i)* is just to say that I can, by moving in certain ways, visually access the cup and its properties. I can exercise this ability right now, by moving my eyes in a certain way and attending to the cup. The requirement that this ability obtains is the first requirement that must be satisfied if an object/property is to be virtually present.

Reconstructing condition (ii) will take more work. In this section, I will give a first-pass way of understanding it, then show how the account of virtual content generated by fitting (i)* together with the first-pass version of (ii) might go some way towards meeting the desiderata set out in §5. Once we have the rough idea in

place, I will show, in §7.2, that (ii) stands in further need of alteration. On the first-pass reconstruction, the condition might be set out as follows:

> (ii) For an object/property to be virtually present to me, it must be the case that suitably dramatic movements or other changes to the object/property would cause the property-bearing object to become the subject of my focal awareness.

Let's consider my cup again. If the cup were to move or change in some way, say by falling off the table, or turning pink, this fact would draw my attention, and the cup would become the focus of my awareness. The fact that dramatic changes in my periphery will draw my attention is not a trivial fact. I must somehow have attentional mechanisms that are sensitive to such changes, and the ability to move my eyes and my attention to the area in which change is happening.[92] It is worth noting that condition (ii) is satisfied by object-properties as well as objects. Take the size of the cup in my periphery. If the cup's size were suddenly to change, I would notice this, and my attention would be drawn to the cup. Since (ii) can be satisfied by object-properties, and since (ii) is a precondition for virtual perceptual presence, it can be used to make sense of the idea that object-properties are virtually present.

Note also that the condition applies to perspectival properties as well as perspective-independent properties. Suppose I have on my table a coin standing on its edge. From the perspective I am occupying, the visible perspectival shape of the coin is elliptical – it looks elliptical from here. But suppose it rotated, quickly and silently, on the spot. Suppose that after the rotation, the coin is facing me head-on, meaning that the perspectival coin-shape visible from here is more-or-less circular. This change in the visible perspectival shape of the coin would, if it was suitably dramatic, draw my visual attention; it would cause me to make the coin the focus of my visual awareness. So this change in perspectival shape would bring the object bearing that perspectival property into the focus of my awareness. This is just to say that the perspectival coin-shape visible from here satisfies (ii). It also satisfies (i)*, since I can, at will, bring it into focal awareness. One of Noë's key aims is to build into his account the resources necessary to accommodate the perception of

[92] This ability is sometimes referred to as the ability to keep tabs on objects in our environment – see Dennett (1991, pp.360-1) and Kevin O'Regan (2011, pp.31-2). In section §9, I will say something about what is necessary for this kind of ability.

perspectival properties. I suggest we can achieve this aim just by thinking of perspectival properties (like the coin's apparent ellipticality) as a special class of properties that can be virtually present. They are special in that they are properties that can only be viewed from particular perspectives. But they can still be virtually present, because they can satisfy the conditions necessary for virtual presence.

Now notice how, on my construal, the above theory of access satisfies the desiderata set out in §5. First, consider counterfactual dependency. The conditions require that if I am to experience the cup virtually, my experience of how the cup looks depends in fine-grained, counterfactual-supporting ways, on how the cup is and on the nature of my relation to the cup (be that relation spatial/perspectival or attentional). The relevant counterfactuals are as follows. If the cup changes sufficiently, I will notice. If my perspectival relation to is altered (by movements made either by me or the object seen), I will also notice. And, by appropriate shifts of attention, I can, at will, bring the cup into focal awareness.

Also note how the conditions can deal with Prinz' TV channel objection to the sufficiency of accessibility for visual presence. Prinz' idea was that the events depicted on the other TV channels are accessible to me, but they are not visually present. But condition (ii) is obviously not satisfied by Prinz' TV case. Let's suppose for the sake of argument that we can describe another channel (or some event depicted thereon) as an object or a property. No amount of change in, or movement of, this object/property will result in the relevant object/property becoming the focus of my awareness. Condition (i)* is also unlikely to be satisfied, but there seems little point going into this, given how far short of condition (ii) Prinz' case falls. Since the case does not satisfy the relevant conditions, it is not a genuine counterexample to their sufficiency.

§7.2: Restricting (ii) to Avoid Insufficiency Problems

I now want to focus on Noë's condition (ii) in more detail. I want to show that the condition stands in need of further alteration. In its current form, the condition goes as follows:

> (ii) For an object/property to be virtually present to me, it must be the case that suitably dramatic movements or other changes to the object/property

> would cause the property-bearing object to become the subject of my focal awareness.

I want to focus on the disjunction 'movement or other changes', which is found in Noë's original formulation of the condition. I agree with Noë that we ought to distinguish between movement-involving and non-movement-involving changes, but more needs to be said about the distinction. When thinking about the 'movement' side of the disjunction, we should include growth, shrinkage and shape-change as well as ordinary movement. This way, 'movement' picks out any change that alters the spatial relations that obtain between the perceiver and the object (or object-part) seen: if my mug suddenly moves, grows, shrinks or changes shape, it will take up a different area of space in my environment and in my visual field. Whereas it can change colour or surface-texture, say, while taking up exactly the same space in my visual field.

But once we have distinguished movement-involving changes from changes that do not involve movement, it becomes apparent that there is an ambiguity as to what it takes for condition (ii) to be satisfied. To see this ambiguity, consider Noë's most notorious examples of the virtually present – the tomato-backside or the occluded cat-parts (2004, p.63). Noë claims that the tomato-backside is virtually present to some degree. And he repeats this claim in the paper from which I take his two conditions (2009a, p.479). So Noë must be construing (ii) in such a way that it is satisfied by the tomato-backside. Let's see how the condition might be construed in order for accommodate this point. It looks right to say that movement-involving changes to the tomato-backside could potentially bring the tomato into focal awareness. Suppose the whole tomato suddenly rotated through 180 degrees, or suppose the tomato-backside were to bulge out suddenly, or detach from the tomato and fly into my field of vision. I would notice such movement-involving changes. But now imagine that the tomato-backside suddenly changed colour, turned to stone, or became translucent. These non-movement-involving changes would not draw my attention. I would notice one sort of change – the movement-involving sort. But I would not notice changes of the other sort – the non-movement-involving sort.

Now let's return to condition (ii). On one reading, this condition is satisfied so long as the perceiver would notice changes of at least one sort (the movement-involving sort, say). On this reading, the tomato-backside satisfies (ii). As far as Noë is concerned this is the right result (since he wants to say that the tomato-backside is virtually present). But on another, more demanding reading, (ii) is only satisfied if the perceiver would notice *any* suitably dramatic change to the tomato-backside – whether this change is movement-involving or not. On this reading, the tomato-backside would not satisfy condition (ii), since the perceiver would not notice changes to the tomato that fall into the non-movement-involving category. Like Martin (2008, p.678), I don't find it plausible to think that there is any sense in which I see the tomato's backside. It simply does not feature in my visual phenomenology. Granted, I can see the whole tomato, but this is because I can see a whole thing without seeing all its parts at once (this was the point that the earlier fractal-coin-grabbing example was supposed to illustrate).

So I want to disambiguate (ii) in such a way as to exclude tomato backsides from the realm of the virtually present. The easiest way to do this is to separate it into two conditions:

> (ii)* For an object/property to be virtually present to me, it must be the case that suitably dramatic *movement-involving* changes to the object/property would cause the property-bearing object to become the subject of my focal awareness
>
> (iii)* For an object/property to be virtually present to me, it must be the case that suitably dramatic *non-movement-involving* changes to the object/property would cause the property-bearing object to become the subject of my focal awareness

We can then say that for an object to be accessible to a perceiver, the following must be true: the relation between a perceiver and an object/property must satisfy *all three* conditions: (i)*, (ii)* and (iii)*. We can now compare the two rival theories (the permissive one and the demanding one) against some test-cases. Remember that according to both theories, to be accessible is to be perceptually present in a meaningful sense.

First, take the tomato-backside. As we have already seen, the tomato-backside is accessible (and therefore perceptually present) on condition (ii), read permissively. But it is not accessible once we split condition (ii) as I recommend we should. It doesn't satisfy (iii)* because the back of the tomato could change suddenly change colour, turn to stone, become translucent, and so on without my noticing. For this reason, it is not perceptually present (not even virtually) on the demanding account. This seems to me the right way of describing the phenomenology – the tomato-backside is not visually present. But let's try to make a more compelling case for those who aren't convinced.

Imagine I construct a perfect occluder for some tomato – an occluder which, when placed between the tomato and a perceiver occupying a very specific perspective, will occlude the tomato and nothing else. Imagine I set up an experiment such that the occluder is occluding your view of the tomato perfectly. If you were to move a fraction, you could bring the tomato's edge into view. And if the tomato was to move a fraction, it would peep out from behind the occluder. Suppose also that I tell you there is a tomato behind the occluder, so you can bring whatever sensorimotor grasp you like to bear on the situation. Conditions (i)* and condition (ii), read permissively, are both satisfied in this situation, since even the slightest movement on the part of either you or the tomato could bring a tomato-edge into view. So the tomato should be accessible, and hence perceptually present.

But this seems wrong – before you or the tomato move, the tomato is making absolutely no impact on your visual phenomenology. Now contrast this with my separated conditions. Again, the tomato doesn't satisfy (iii)*, since it could suddenly turn green, become translucent or turn to stone, without drawing the perceiver's attention (provided there were no nearby colour-reflective surfaces).[93] This seems like the right result. Noë might bite the bullet at this point, maintaining that the tomato is visually present in this scenario, even though it is having no appreciable effect on your visual phenomenology. But in doing so, he would be riding roughshod over the phenomenological facts in order to save his theory. And he

[93] One might object at this point that the backside *does* satisfy (iii)* since I would notice if the tomato started emitting intense beams of light. This looks like a non-movement-involving change. Although this change would draw my attention, it would not cause the tomato-backside itself to become the focus of my visual attention. The tomato-backside would still be absent from my visual field and for this reason, it would still not satisfy condition (iii)* if this scenario were to obtain.

can't fall back on his usual line, which is to claim that we need to posit the visual presence of occluded object-parts in order to make sense of the idea that we see whole objects. He can't do this because there is no tomato-whole whose visual presence stands in need of explanation. This being the case, it looks like there is a presumption in favour of my amended conditions.

Another consideration in favour of my amended conditions is that it scores better against one of Noë's own desiderata: it better captures the idea that perception is dependent in a fine-grained, counterfactual-supporting way on the way things are. The idea is that if a peripheral object is going to be accessible (and thereby perceptually present in a virtual way), it had better be the case that I would notice if it suddenly changed colour, turned translucent, and so on. On the less demanding theory of virtual presence, this is not strictly necessary – as the occluded tomato case illustrates. And notice that the amended view still captures Noë's additional thought, which was that perception is in fine-grained ways dependent not only on how things are, but on our relation to how things are. This idea can still be captured by condition (i)*, which I have not changed, and by (ii)*, which I have made into a condition exclusively about movement-involving changes – changes that alter in some way the perceiver's spatial relation to how things are.

§8: Degrees of Perceptual Presence

But now let's consider what *would* count as accessible (and hence virtually present) on this revised theory. Take the mug just off to my right. The virtual content theory was supposed make room for the idea that peripheral objects of this sort can still be perceptually present, while keeping consistent with what the scientists tell us. So my new conditions should be satisfied by the mug. We've seen already how (i)* is satisfied by my mug. And (ii)* looks to be satisfied, since if the mug were to fall off the desk, or to move in a similarly dramatic fashion, it would draw my attention. And (iii)* is also satisfied, since if the mug were suddenly to change from blue to pink, I would most likely notice.

But at this point one might legitimately ask how extreme a change must be in order to draw my attention. Conditions (i)*, (ii)*, and (iii)* can all obtain to varying degrees. Take (iii)*. It might be the case that I would notice, say, a sudden, dramatic

colour-change in some object behind my computer screen. But perhaps I would not notice a less dramatic change, or a dramatic change that occurred gradually. Indeed, this is precisely the lesson to learn from empirical studies of the sort mentioned in §2. Something similar is true of condition (ii)*. If an object is to satisfy this condition, it must be the case that suitably dramatic changes in my spatial relation to that object would draw my attention. But what counts as a suitably dramatic change in this context? And now take (i)*: enough of a movement and/or attentional shift on my part could bring just about anything into the focus of my awareness. Where, one might ask, should we set the bar for meeting condition (i)*?

Following Noë's lead, I want to argue that it makes sense to speak of degrees of perceptual presence. *Ceteris paribus*, those things that are closer to the centre of one's focal awareness are perceptually present to a higher degree than those things in your visual periphery. I say *ceteris paribus* because being easily accessible to focal awareness isn't simply matter of being close to the centre of the visual field. I take the inattentional blindness studies like the gorilla experiment to show us that when engaged in an attention-absorbing task, for example, even quite dramatic changes in our environment, near the centre of the visual field, can fail to capture our attention. The exact degree to which a given environmental change occurring in our visual field will grab our attention is an empirical question – one that is unlikely to admit of an easy solution. But the rough idea is that the degree to which some object/property is virtually present to a perceiver is determined by the extent to which the perceiver's relation to the object/property satisfies conditions (i)*, (ii)* and (iii)*. This is very similar to Noë's (2009a, p.279) proposal. But I have changed the conditions, and changed what it means to say that these conditions obtain. What's more, on my proposal, all three conditions must obtain to at least some degree in order for something to be virtually present at all. Since my conditions are different from Noë's, this will lead to different results. Let's see how the two proposals perform when dealing with some examples.

Let's take some of Noë's own examples. He writes that 'the front of the tomato is maximally present; the back a little less so; the hallway even less so. To these gradations of degree there correspond gradations in the degree to which the motor-sensory relation we bear to the object, quality, or situation, is movement- and

object- dependent' (2009a, p.479). We have already seen why Noë's theory commits him to the claim that the tomato-backside is present. But here he goes further, making similar claims about the hallway, and so on (he even goes as far as to say that the Eifel Tower is ever so slightly present to him visually). On my view this is not the case. Take an object in the hallway – perhaps a person standing just round the corner of my doorway. If I were to move in the right way, or if he/she were to do so, the person would come into vision. So, for Noë, if enough movement would make them present, they are present already, to some degree. Not so on my story. If the person were to suddenly turn to stone, turn translucent, or change colour, I would be none the wiser. So, on my view, they are not virtually present at all, since (iii)* is not satisfied to any degree. Again, this seems to me the right result.

By contrast, the redness of the wall behind the computer-screen I am looking at is virtually present, but it is not maximally present. As the change-blindness literature shows, when I am focusing on something in the foreground, the wall-colour can change gradually without my attention being drawn away from the foregrounded object and towards the wall.[94] So the wall's colour does not satisfy (iii)* maximally. It satisfies it to some extent, since a very dramatic and sudden colour-change *would* capture my attention. Again, this seems like a phenomenologically plausible result. It does seem that the wall in the background is present to some degree, but it also seems right to say that it's not as present as that which is currently the focus of my visual awareness.

§8.1 Counterexamples to the Necessity of (i)*, (ii)* and (iii)*

But now one might wonder if my conditions for virtual presence are too demanding. We can take something that seems uncontroversially to be virtually present to me, and then engineer certain scenarios in which it would satisfy none of the three conditions I laid down. Let's take the cup in my periphery. Right now, it is featuring in my visual phenomenology, but it is firmly in my periphery (let's assume). So I want my account to deliver the result that it is virtually present. But imagine that right now there is an evil scientist watching me, and monitoring my brain activity, such that he is ready to zap the cup into a million pieces the moment I come even

[94] See Simons and Levin (1998).

close to turning my attention towards it.[95] In this case, it looks like I don't satisfy (i)*, since I am not able to bring the cup into focal awareness at will. And similarly, we can imagine the scientist is willing to kill me the moment the cup moves, or changes colour, or changes in any other ordinarily visible manner. Let's stipulate that the scientist has resolved to kill me so swiftly that I do not have time to notice the cup changing before I vanish out of existence. In this scenario, conditions (ii)* and (iii)* would also fail to hold true of my relation to the cup. But, *ex hypothesi*, the cup is featuring in the periphery of my visual field, so its presence is still the kind of phenomenon that my account should be accommodating.

In response to this worry, I suggest we treat the three conditions as governed by typicality clauses. To do this is to say that for an object to be virtually present, it must be the case that my relation to it typically satisfies my three conditions. The first thing I need to do is distinguish this proposal from that canvassed in the previous section. It seems that, in the evil scientist case, the cup is perceptually present to exactly the same degree as it would be if the evil scientist was on a lunch-break. My proposal is that this is because in both situations, *barring unusual circumstances*, the three conditions *would* hold to the same degree. The same sets of relevant circumstances would result in my attending to the cup in both circumstances (so long as the scientist stayed out of things). The obvious worry is that this response cheats by making my account trivially true; my conditions are necessary for perceptual presence, except in cases where they don't intuitively seem to be.

To answer this worry, let's review the desiderata my account was supposed to serve. The first desideratum was capturing the insight that perception is sensitive (in fine-grained, counterfactual-supporting ways) not only to how things are, but to the perceiver's relation to things seen. One way to think of this 'insight' is to think of it as a folk platitude about the functional role constitutive of perceptual states; it is the kind of statement that can be recognised as obviously true by anyone who reflects on their own experience. But it would only count as a folk platitude if it was governed by a *typicality* clause similar to the one I'm trying to vindicate. To count as a platitude, the 'counterfactual dependency' principle must be true. But, as devious

[95] This case, is extremely similar to the 'censor' case discussed in Lewis's seminal (1980b, p.248) paper, as are the permutations on the case that I discuss in what follows.

scientist cases show, the principle is only *typically* true; it is not true in all cases. When the scientist is meddling, I am not sensitive to changes in the way things are (or to changes in my relation to how things are) in a counterfactual supporting way. David Lewis seems to think of the 'counterfactual dependence' principle central to the causal theory should be governed by a typicality condition of the sort I recommend. The following quotation (taken from his seminal paper on the causal theory) should make this point clear:

> The difference between [non-seeing] and genuine seeing is not sharp, on my analysis. It if fuzzy; when the requirement of suitable counterfactual dependence is met to some degree that falls far short of the standard set by normal seeing, we may expect borderline cases. (1980b, p.247)

As discussed in chapter 1, and as Lewis intimates in the above quotation, there are reasons for thinking that in analysing mental kinds in terms of functional roles, we must respect the fact that our analyses will admit of borderline cases. I will not rehearse these reasons here; they were discussed at some length in the first chapter. The point for now is that we allow for such borderline cases by making our analyses fuzzy, and we do this by qualifying them with typicality clauses. And this is just what I suggest we do for our current analysis.

But accommodating the key insight of the causal theory was not the only desideratum that the virtual content theory was supposed to meet. It was ultimately supposed to provide an alternative to the 'filling in' hypothesis. To do this, it must go beyond capturing folk platitudes about functional role. It must say something about the manner in which these functional roles are realised: it must say something about the mechanisms that explain our perceptual capacities. The next section explores the potential empirical consequences of the virtual content theory canvassed here.

§9: Sensorimotor Grasp and S-Representation

Recall that in addition to the conditions so far discussed, Noë claims that there is another condition, a sensorimotor grasp condition, which must also be satisfied if virtual presence is to obtain. In §4 I noted that this sensorimotor understanding condition, as cashed out by Noë, is a little vague. But I delayed detailed discussion

of it until this section. Although Noë claims that sensorimotor knowledge is a kind of skilful mastery, what he says about it often seems to suggest that it has a propositional, or at least representational component. This is not a new point. It has been made by Rowlands, (2006, 2007, 2010a), Block (2005) and Hutto (2005). To see the basic point, consider the following quotation from Noë:

> Our perceptual sense of the tomato's wholeness – of its volume and backside, and so forth – consists in our implicit understanding (our expectation) *that* the movement of our body to the left or right, say, will bring further bits of the tomato into view. (2004, p.63; emphasis added).[96]

The implicit understanding Noë talks of here is supposed to be an instance of sensorimotor understanding (or sensorimotor knowledge). Sensorimotor knowledge is supposed to be a form of practical skill or mastery. It is *not*, for Noë, supposed to be propositional knowledge (2004, pp.65-67; pp.117-120). But having the kind of knowledge invoked in the above quotation seems to be a case of representing certain facts about the world (or about one's perceptual relation to the world). It seems that sensorimotor knowledge here is a matter of knowing (or at least representing implicitly) certain counterfactuals of the form 'if I do such and such, so and so that will become visible to me.'

To see how this point applies to the material discussed so far, let's consider the three conditions I set for virtual presence and ask what sensorimotor expectations a perceiver plausibly needs to have in order to satisfy these conditions:

> (i)* If some object or property is to be virtually present, it must be the case that I can, at will, bring that object/property into focal awareness.
>
> (ii)* For an object/property to be virtually present to me, it must be the case that suitably dramatic *movement-involving* changes to the object/property would cause the property-bearing object to become the subject of my focal awareness
>
> (iii)* For an object/property to be virtually present to me, it must be the case that suitably dramatic *non-movement-involving* changes to the object/property

[96] This passage was originally quoted in Rowlands (2010a, p.76). The addition of emphasis is also originally due to Rowlands.

would cause the property-bearing object to become the subject of my focal awareness

Looked at in one way, each of these conditions picks out a perceptual ability that a perceiver must have, relative to some object/property, in order for that object/property to be virtually present. But in order to have these abilities, we must have certain expectations, or so I will argue.

To satisfy condition (i)*, I must be able to bring an object/property into focal awareness at will. But it is plausible to think that in order to have this ability, I must have accurate (albeit tacit, sub-personal) understanding of certain sensorimotor counterfactuals. To bring the cup in my periphery into focal awareness, I need to perform an appropriate eye-movement. The fact that I can perform such an eye-movement suggests that my visual system has somehow mastered a relevant counterfactual of the form: 'eye-movement *X* will bring object *Y* into my focal awareness'.[97] So a full story about how it is that I can satisfy condition (i)* must say something about how it is that I have mastery of the relevant counterfactual.

And satisfying conditions (ii)* and (iii)* requires sensitivity to changes occurring in my visual field. Changes made to objects/properties in my visual field will plausibly lead to changes in the pattern of sensory input registered at my retina. But in order to identify such changes *as* changes, I must have some expectations as to what input I would receive if *no* changes occurred, such that these expectations would be violated in cases where changes *do* occur in my environment. So I must have expectations of the sort 'if there are no significant changes in my visual field, the patterns of input I receive will be such and such'. Only when I have expectations of this sort, and a means of noticing when these expectations are violated, will I be able to notice the relevant changes in my visual field. So it looks like I need (implicit) knowledge/representation of certain counterfactuals of this sort if I am to satisfy conditions (ii)* and (iii)*.

I think we can meet these theoretical demands by explanatory appeal to s-representations. S-representation can get us the requisite counterfactual expectations while accommodating Noë's idea that sensorimotor knowledge is not encoded in

[97] One might think that the eye could just dart around randomly for a bit until it happened upon the desired object. It could do, but this quite clearly isn't how the eye actually works.

propositional form. And this proposal also generates a plausible way of explaining how sensorimotor knowledge is instantiated in perceivers (or so I will claim). To see how this proposal works, suppose that when I am about to move relative to the tomato, the motor centre in charge of this movement sends a copy of the motor command to an emulator. This emulator then simulates the effect this motor command will have on the location of the tomato relative to my visual system, thus providing an expectation about what I will see once I have completed this movement. This hypothetical emulator thus explains the ability to build expectations about the manner in which the perceiver's movements will determine what they subsequently see. One can hypothesise the existence of such emulators while also claiming that sensorimotor knowledge is not propositional knowledge, since emulator representations do not represent in the way that propositions represent (as should have been made clear in chapters 2 and 3). But is this emulator hypothesis plausible? I hope to show that it is.

This hypothesised emulator might sound like a hopelessly speculative idea, but it isn't. Grush sees Noë's claims about sensorimotor contingencies as an 'underspecified and confused' precursor to his own emulator theory of visual anticipations (2007, p.390), so the modelling of sensorimotor knowledge as emulator representations (or 's-representations', in my terminology) is not a new idea. As Grush has already argued (2007), there is evidence that our visual system actually makes use of such emulators. I will briefly describe one strand of evidence Grush cites. This evidence is taken from a study by Duhamel et al. (1992). The study involves a monkey looking at a small square on a blank background, where there is a small disc diagonally down and to the right of the square. Suppose that the monkey is currently looking at the square, and that the disc is in the periphery of the monkey's visual field. This being the case, the peripheral disc will be causing the activation of some region of the monkey's visual cortex. But now suppose that the monkey moves its eyes from the square to the disc. Once this movement is complete, the disc will start to cause the activation of some other new region of the monkey's visual cortex. One of the cells in this new region is the PPC neural cell, whose activity is being monitored. One might predict that this new cell will only start lighting up once the monkey's eye-movement is completed. After all, it is only then that the light reflected off the disc will be stimulating the part of the retina that

is retinotopic with the PPC cell. But in fact, the PPC cell starts lighting up *before* the monkey's eye-movement has been completed. Grush claims that the 'PPC neuron appears to be anticipating its future activity as a function of the current retinal projection and the just issued motor command. That is, it is a visual modal emulator' (2007, p.402).

The idea, basically, is that the monkey's visual system is predicting the effect of the monkey's motor system's activity. The monkey's motor system 'tells' the eye to move in a certain direction. A copy of this command is then fed into the visual modal emulator. Given the magnitude and direction of the movement the eye has been 'told' to make, the emulator can predict that something which was previously on the periphery of the monkey's visual field will soon come to its centre. The emulator thus allows the monkey's visual system to anticipate the effects of eye-movement on the (visual) sensory stimulation the monkey will receive. And it looks like this is exactly the kind of emulation that would be required in order to grasp the kinds of counterfactuals required for the ability picked out by (i)*. The emulator emulates the sensory consequences of a certain kind of eye-movement; it 'says' something like 'if such and such a movement is performed, such and such a sensory input will be received.' This is just one of the studies that Grush cites as evidence for the existence of such a 'visual modal emulator'. But hopefully such evidence is enough to render less speculative the suggestion that we model sensorimotor knowledge as an instance of s-representation.

And now, briefly, to the kind of expectations necessary for the abilities required by (ii)* and (iii)*. As argued above, satisfying these criteria requires expectations about what patterns of input one would receive in the no-change condition, such that deviations from these expectations would be detected. This seems to me exactly the sort of expectations that the predictive coders, discussed at the start of chapter 3, were supposed to generate. Recall that these coders generated expectations about future patterns of sensory input, and that they only registered input that deviated from the expected patterns (i.e. they only registered so-called 'prediction error'). Systems of this sort seem to be exactly what is needed in order for (ii)* and (iii)* to be satisfied; they can form expectations about future input, and

detect situations in which these expectations are violated.[98] Obviously, much more work would need to be done to establish that predictive coders are indeed capable of generating expectations of exactly the sort necessary for the satisfaction of (ii)* and (iii)*, but I hope to have established at least the idea that investigating this question is a promising avenue for future research.

But one might wonder at this point whether giving this explanatory role to s-representations undermines the project with which I started. After all, the virtual content hypothesis was supposed to be an alternative to the 'filling in' hypothesis. The claim was supposed to be that because peripheral content is virtually present, we can explain it without claiming that our visual field is 'filled in' by representations cooked up by the brain. But if the s-representation story just sketched is right, it looks like we're reverting back to a representational story. This problem can be finessed if we distinguish between experience *of* representations and experience *that requires* representational capacities of a certain kind. The filling in hypothesis was the idea that when we experience peripheral content, what we are experiencing is just a whole set of representations cooked up by the brain. On the current proposal, by contrast, this is not what is happening. We are experiencing the objects, not representations of those objects. It's just that in order to experience those objects as accessible, we need to be deploying certain representational capacities.

§10: Conclusions

The virtual content theory I have developed is better than Noë's theory at accounting for what I take to be obvious phenomenological facts. And it can do so without relying on the incoherent claim that perceptual content is accessible but impossible to access. It also avoids sufficiency counterexamples to which Noë's view is vulnerable. And it captures the idea that perceptual experience is sensitive, in a fine -grained counterfactual-supporting way to (1) the way things are and (2) the perceiver's relation to the way things are. What's more, I have offered a promising way of understanding the notion of 'sensorimotor knowledge/grasp' on which Noë's account of virtual content depends. And, as we have just seen, the virtual

[98] See Clark (2013, p.20) for the suggestion that attentional mechanisms might be explainable in terms of prediction error minimisation.

content hypothesis can remain distinct from the 'filling in' hypothesis, while still giving a fairly central role to a certain kind of representational capacity.

I'll finish this chapter by remarking on the relation between the virtual content theory developed here and the extended mind thesis (EM) discussed in chapter 1. At the end of the first chapter, I argued that one could endorse the virtual content thesis while remaining neutral on the EM debate. I hope here to have shown in more detail how this can be done. I have described and argued for a version of the virtual content thesis. And at no point did my argument rest on assumptions about the plausibility or otherwise of EM.

Chapter 6: The Problem of Invisible Contents[99]

Abstract

This chapter identifies, and attempts to resolve, a serious inconsistency in Alva Noë's theory of perception. I argue that a key feature of Noë's theory of perception, what I will call his 'indigestibility thesis', is incompatible with his 'p-properties' account of perspectival content. The indigestibility thesis implies that p-properties, as characterised by Noë, cannot be reliably visible. P-properties play an important role in Noë's theory of perception, and they could not play this role if weren't reliably visible. This problem (the 'problem of invisible contents') must be solved by amending either the indigestibility thesis, or Noë's account of perspectival content. I argue that the indigestibility thesis should not be rejected, and then try to solve the problem of invisible contents by amending Noë's theory of perspectival content.

§1: Introduction

This chapter aims to tie up a loose end identified in chapter 4, but postponed until this point. In chapter 4, I examined the claim that sensorimotor knowledge might be necessary for getting from perception of perspectival properties to perception of perspective-independent properties. I suggested a way in which we might understand the role of sensorimotor knowledge in this context, but postponed (until now) the identification and ironing out of a wrinkle in Noë's story. This wrinkle, the 'problem of invisible contents', is generated by two incompatible Noëan theses. I will start by getting clear on what these theses amount to.

§2: The Indigestibility Thesis

The indigestibility thesis featured briefly in the last chapter, although I did not give it a name at that point. It was a key claim that Noë deployed in attempting to justify the claim that perceptual content is virtual 'all the way in'. I offer the following definition of the indigestibility thesis: visible objects and properties cannot be taken in perceptually (or presented to consciousness) in a single instant. Indigestible visible objects or properties are those that cannot be taken in, or 'digested', in a single moment. Noë seems to think that everything visible is also indigestible, as the

[99] Much of the material presented in this chapter appeared in a collection published by *Springer* (Wadham; 2014). The copyright agreement I signed licences me to use this material here, so long as the *Springer* publication is cited as the original source of publication. See Wadham (2014) for the necessary citation.

quotations below should go some way towards showing (I have underlined the key parts of the passages):

> [Y]ou can no more embrace the *whole* of the facing side all at once in consciousness than embrace the *whole* tomato in consciousness all at once. This is clear on reflection I think […] This shows that we cannot factor experience into an *occurrent* and a *merely virtual* or potential part. Experience is fractal in this sense. (2004, pp.134-5)

> When you peel away the layers of potentiality and merely virtual presence, you are not left with pure phenomenal content, that which, as it were, is present to your mind now. You are always presented with qualities that in turn have qualities and that are presented against a structured background. (2004, p.216)

> Experience is not something that happens in us. It is something we do; it is a temporally extended process of skilful probing. […] Experience has content only thanks to the established dynamics of interaction between perceiver and world. (2004, p.216)

In these passages, Noë repeatedly emphasises the idea that visible objects and properties cannot be present to the mind in their entirety, all at once. And he seems to think that this point holds true of all visible properties. He asks us, for example to imagine ourselves looking at the face of a tomato. As he claims, 'you will admit that you don't actually experience every part even of its visible surface all at once' (2004, p.217). And he goes on to claim that this point generalises to every visible quality. He challenges us thus: 'Pick any candidate for the occurrent factor. Now consider it. It too is structured; it too has hidden facets or aspects. It is present only in potential' (2004, p.217). Here, he is pushing us towards the claim that no visible property is so simple that it can be taken in at a single instant. He concludes that 'qualities are available in experience as possibilities, as potentialities, but not as completed givens' (2004, p.2007).

I'll start by saying two things about the indigestibility thesis. First, it is deeply implausible to think that *all* visible objects and properties are indigestible. To see why, consider the case of subitization. Coined by Kaufman et al. (1949), 'subitization' is the ability to enumerate grouped objects that are presented only for a fraction of a second. For example, if I showed you a picture of the side of a die with five dots on it for only a fraction of a second, you would be able to tell me the number of dots on the die. You can do this without counting the dots one by one; you take in the information in an instant. Subitization is just one example of our

ability to take in perceptual content in an instant. But the fact that we have such an ability seems to run counter to Noë's claim that every visible property is indigestible. I can, in an instant, perceptually apprehend a property instantiated by a dot-cluster: the property of being five in number, for example. Noë could counter by saying that even in subitization cases, the time-frame is a fraction of a second, which is still an extended temporal interval. But such a move threatens to render the indigestibility thesis hopelessly trivial. Of course we can pick a time frame so short that no object presented for that amount of time is visible.

But Noë is probably right to think that most objects and properties are complex enough that they cannot be apprehended in their entirety in a single moment. To take his tomato-face example, it seems that I can switch attention from one aspect of the tomato-face to another, and it seems that I cannot focus on all of its aspects all at once. So it seems likely that the indigestibility thesis must be true at least of most complex visible objects and properties (where even the facing side of a tomato counts as complex). This might also seem like quite a trivial claim, but as we will see, it has serious ramifications for other aspects of Noë's theory. The second point to make is that the indigestibility thesis doesn't imply the claim that content is virtual 'all the way in', as Noë seems to think it does. This was the point made in §4 of the last chapter, and illustrated there with the fractal coin example.

§3: P-Properties

Recall that Noë gives the name 'p-properties' to the visible perspectival properties objects have from certain perspectives. He uses the cases of p-shape and p-size to illustrate the basic idea:

> That a plate has a given p-shape is a fact about the plate's shape, one determined by the plate's relation to the location of the perceiver, and to the ambient light. The p-shape is the shape of the patch needed to occlude the object on a plane perpendicular to the line of sight. The p-size of the tree is, in turn a fact about how the trees look, with respect to size, from the location of the perceiver: it is identical to the size of a patch we can imagine drawn on the occlusion plane (2004, p.85).[100]

[100] Noë doesn't explain what he means by the claim that the perceiver's relation to the 'ambient light' also determines an object's p-shape. This point need not concern us here though.

So the idea is as follows. Suppose you look at a plate from a given position. The particular spatial relation between yourself and the plate will determine the exact p-property that you will see from that location. Imagine that you stayed in a fixed position relative to the plate and held out an opaque material perpendicular to your line of sight. If this occluder were exactly the right shape, it could totally occlude the plate, without occluding anything else in your field of vision. If you managed to construct such an occluder, you would have captured the p-shape of the plate from your particular position.[101]

Noë is emphatic in claiming that, relative to any given position at any given moment, an object has a *single* and *determinate* p-shape/p-size etc. He writes that 'there *is* a single apparent size of an object – namely, the unique way that an object looks with respect to size from a particular position. This is secured by phenomenology' (2004, p.84).[102] What we also see in this quotation is the claim that the visibility of p-properties is, for Noë, something that is 'secured by phenomenology'. Noë is emphatic in asserting that we *see* p-properties and he clearly regards this as a claim that accurately describes our phenomenology: 'P-properties are themselves *objects of sight*, that is, things that we see. They are visible' (2004, p.83).

Now let's look at the role that p-properties play in Noë's theory of perception. As we saw briefly in chapter 4, Noë explains our ability to perceive perspective-independent properties in terms of our ability to perceive p-properties. Let's remind ourselves of the basic idea with an example:

> To see a circular plate from an angle, for example, is to see something with an elliptical P-shape, and it is to understand how that perspectival shape would vary as a function one's (possible or actual) movements with respect to the perceived object. We see its circularity *in* the fact that it looks elliptical from here. (2004, p.84)

The idea is that as we move around our environment, the p-properties of the objects in our visual field will change. For example, if I have my head immediately above

[101] Notice that the size of this occluder could vary. If you made a small occluder and held it up close to your eye, this would have a similar effect to a larger occluder held closer to the plate (assuming there's no difference caused by binocular disparity). This point shouldn't worry Noë; he can still claim that the plate has a single determinate p-shape, since shape is different from size (a square is a square no matter how big or small it is).

[102] For a similar 'occlusion' theory of perspectival features in depiction theory, see John Hyman (2006).

my plate, the plate will have a circular p-shape. But if I now sit back gradually in my chair, the plate's p-shape will become gradually more elliptical (because my perspective on the plate is becoming more oblique as I move). Noë thinks that it is because I see such variation of p-properties and because I understand the way in which this variation is a function of my changed perspective that I am able to see the perspective-independent properties of objects. To continue with the plate example, we might say that one particular p-property of a plate would not uniquely specify the plate's perspective-independent shape. A circular plate viewed from a certain angle might have a certain p-shape but equally, a slightly elliptical plate, viewed from a slightly different angle might have exactly the same p-shape. But when I move, I see a variation in p-shape. And when I move in a particular way, the particular variation in p-shape I see would only have occurred if the plate in question was circular. This is an example of the general claim that movement-relative variations in p-properties uniquely specify perspective-independent properties, and our ability to see the former explains our ability to see the latter. So, for Noë, our perception of p-properties explains our perception of perspective-independent properties.

§4: The Problem of Invisible Contents

We are now in a position to see why Noë's account of p-properties is incompatible with his indigestibility thesis. Recall that, for Noë, perceptual content is not given to us all at once. It is only because we engage in the extended activity of perceptual exploration that perceptual content becomes available to us. Presumably, this applies to p-properties in the same way that it would apply to non-perspectival content. And if this is right, then we can't, say, take in the p-shape of a person's face in profile all at once.

So what? One might ask. If I hold still, and look at the person's face over an extended period of time, I can see the face's p-shape. There will be a single p-shape visible to me for as long as I hold still, and I can take it in at my leisure. But suppose I don't stay still. If I am moving around while looking at the face, the p-shape visible to me from my particular perspective will change at every instant. When I move around the face, the 'here' (the location from which I am viewing it) will be different at every 'now' (the time at which I see the face). And so, at every 'now', a different

face-p-shape will be visible to me. But if content is not given to me all at once, I can't see anything that is only visible for the duration of any 'now'. If content is not given to me in an instant, I cannot see some feature of the world that is only visible to me for an instant. And when I am moving, a given p-property will, by necessity, only be visible to me for an instant. From this it follows that, while in motion, we cannot see individual p-properties.

Notice that my argument above assumes that it is only possible to see a given p-property of a given object when occupying the perspective at which the object presents that p-property. The idea was that, while moving, we can only occupy a given determinate perspective on an object for a brief instant. And, given that p-properties are only visible from such determinate perspectives, this means that any given p-property will only be visible for a brief instant (the instant at which we occupy some determinate perspective). But if we can't take in any content in an instant, then we can't see a property of the object that is presented to us only for an instant. This argument assumes that we can only see a given p-property of an object while occupying the precise location at which the object presents that p-property. This assumption seems plausible and, as the following quotations should establish, it is one that is shared by Noë:

> Elliptical is just how circular plates viewed from an angle look. Indeed, we experience the plate *as circular* precisely because we encounter its elliptical look from here, and we understand the transformations the elliptical apparent shape (aspect) would undergo as we move. (2004, p.78)

The key point I want to draw out of this quotation is that, for Noë, we experience some particular elliptical p-shape that the plate presents only because we occupy a position at which the plate presents that particular p-shape. So Noë shares the assumption upon which my argument against him rests.

We can now see why my argument to the effect that p-properties must be invisible while we are moving poses a serious problem for Noë's theory of perception. If we can't see p-properties while we move, this undermines Noë's explanation of the fact that we can see the perspective-independent properties of objects. Noë's explanation of this fact is that, as we move, we experience a variety of p-properties presented by the object we are looking at. It is through implicitly understanding the way in which the p-properties we see vary as a function of our

movement that we come to see the actual, perspective-independent properties of objects. But if we can't see p-properties while moving, this explanation doesn't get off the ground. Noë wants to explain our perception of perspective-independent properties by appealing to our understanding of the way in which perspectival properties vary as a function of our movement. But if he is going to do so, he's going to have to give an account of our perception of perspectival properties such that these properties are actually visible while we are in motion. For ease of reference, I will henceforth call this problem **the problem of invisible contents**.

The problem of invisible contents arises from the conjunction of Noë's indigestibility thesis and his claims about p-properties. So we have a choice. We can either eliminate the problem by revising Noë's p-properties story, or we can alter his indigestibility thesis. But, as we have seen, a restricted version of the indigestibility thesis looks eminently plausible. And it looks eminently plausible to think that many perspectival properties are going to be complex enough that they cannot be taken in at a single instant. A person's facial profile is a particularly complex example, but even something more simple, like the perspectival shape of a coffee cup, is most likely far too complex to be taken in at an instant. So it looks like we will have to revise Noë's p-properties story.

§5: Appearance Patterns: Reliably Visible Perspectival Features

I now want to sketch an alternative to Noë's p-property story: the appearance-pattern theory. This theory, when combined with the indigestibility thesis, will not give rise to the problem of invisible contents. To get a sense of what an appearance-pattern is, imagine I move relative to the cup on my table. As I do so, the perspectival properties of the cup visible from my location will change. Let's take p-shape. The visible p-shape of the cup's rim, for example, will change constantly as I move. And it seems right to say that I see this pattern of p-shape-change; the cup's perspectival shape seems to change before my eyes. I want to say that what I see in this case is an appearance-pattern.

One could think of an appearance-pattern as just a change in the way things look from a perceiver's perspective, but this doesn't quite capture what I want to the notion to capture. Suppose I stand perfectly still with respect to my cup and my cup

doesn't move. In this scenario, the perspectival features of the cup visible from here do not change. The cup looks the same way over a stretch of time. But I want to cash out the notion of 'appearance-pattern' in such a way that in this case, it still makes sense to say I see appearance-pattern. To do so, I suggest we think of an appearance-pattern as a temporally extended unfolding of an object's visible perspectival properties. The idea is that I can see an object's visible perspectival properties unfold over time where 'unfolding' can mean either 'changing in some way' or 'staying the same'. The idea can be illustrated with an example. I can see a freshly baked cake's shape unfold over time either by watching as it holds its shape or by watching as it slowly sinks in on itself. In both cases, an unfolding is an 'event' in the sense that it happens over a temporal interval.

So an appearance-pattern is an event in the sense that it is the kind of thing that happens over a period of time. The changing, or the staying the same, of the perspectival property of an object visible from some perspective is something that occurs over a temporal interval. And an appearance-pattern is a visible kind of event: I can see my cup's perspectival features changing or staying the same. And, crucially, appearance-patterns are *reliably* visible. As the visible perspectival shape of my coffee cup changes, it goes from having one visible p-shape to another, to another, and so on. And, if what I have said so far is right, when the p-shape-change is quick enough, and the relevant p-shapes are complex enough, I will not be able to see the individual p-shapes in this pattern of p-shape-change. But it does not follow from this that I cannot see the overall pattern of p-shape-change that unfolded.

To see why this is, suppose I make a film in the following way. I start with a camera held in a fixed position, pointed at my cup. I turn the camera on for a tiny fraction of a second, and then turn it off, creating a short '*still*' of the cup. The *still* is a very short piece of footage, depicting the cup as it would look from a particular determinate location. So the still will present the cup as having a particular visible p-shape, p-size and so on. Having created my first *still*, I then move the camera an imperceptible fraction of an inch, to a new fixed point, ensuring that the camera remains fixed on the coffee mug. I then create a new *still* from this location. Suppose I repeat this process again and again until I have a short film starring my coffee cup. I end up with smooth footage of the coffee cup's perspectival properties slowly

changing. But now suppose each *still* of which this footage is composed is so short in duration as to be imperceptible to the human eye; if I was just shown one *still*, it would flash up and disappear so quickly that I would fail to see anything. Despite being unable to see any particular still, I am capable of seeing a pattern composed of multiple stills, presented one after the other. Notice that while each individual still presents the cup as having a particular determinate p-shape, p-size and so on, the pattern made up of stills played in succession presents me only with a pattern of p-property-change. And it seems I am able to see this pattern of p-property-change without seeing any individual still of which this pattern is ultimately composed.

So appearance-patterns can be reliably visible even when p-properties are not. Notice that what I have said so far does not commit me to claiming that p-properties are never visible. If I want to see the coffee-cup-p-shape visible from a particular perspective, I can simply occupy that perspective for a certain amount of time, and take in the relevant p-shape at my leisure. By doing so, I am able to see a stable appearance-pattern: an appearance-pattern in which the object's visible p-properties are held constant for a certain period of time. And because the relevant p-properties are held constant for a time, they can be taken in at my leisure. In this case, I want to say that perception of an appearance-pattern has explanatory priority over perception of the relevant p-properties. To see why, imagine that I create another short film of my cup, only this time each *still* I take is taken from exactly the same position. The resulting film is just footage of a cup whose visible perspectival properties are unchanging. Suppose again that each *still* of which this footage is composed is invisible to the human eye. I cannot see any individual *still*, or the properties depicted thereon. But because I can see a pattern composed of multiple stills, and because this pattern presents me with exactly those visible p-properties that are depicted on each individual still, I am able to take in those p-properties at my leisure. So I am able to see individual p-properties *by* seeing suitably stabilised appearance-patterns. And this is just to say that the perception of appearance-patterns here has explanatory priority over the perception of p-properties. And while I can only sometimes see p-properties (by seeing a particular kind of appearance-pattern), appearance-patterns are reliably visible.

§6: Appearance-Patterns and Perspective-Independent Properties

Now I have introduced appearance-patterns, I need to show how they can solve the problem of invisible contents. Recall that the problem is as follows. Noë explained the perception of perspective-independent properties in terms of our ability to perceive p-properties, and our understanding of the way in which p-properties varied as a function of movement. But the problem of invisible contents arose because p-properties are not reliably visible. If we need to see p-properties in order to see perspective-independent properties, then we would expect our ability to see perspective-independent properties to be as unreliable as is our ability to see p-properties. But this is not what we find. Unlike p-properties, appearance-patterns are reliably visible. So we should, in principle, be able to explain the perception of perspective-independent properties in terms of our perception of appearance-patterns without generating any 'invisible contents' problems. But I want to do more than showing that such an explanation is in principle possible; I want to say something about how it might work.

Appearance-patterns can replace p-properties in Noë's theory, without any great theoretical overhaul. Noë's idea was that we see perspective-independent properties by seeing p-properties and by understanding how p-properties change as a function of our movement (or, more precisely, as a function of changes in our spatial relation to objects seen). The idea is that when see such patterns of p-shape-change, and we see them with sensorimotor understanding, we thereby come to see perspective-independent properties. We understand (implicitly) that we only perceive some particular pattern of p-shape-change because the object seen has particular non-perspectival, invariant properties. If the object had had different non-perspectival properties, we would have experienced a different pattern of p-shape-change as we moved around the object:

> [W]hen we see, we experience the way the environment structures sensorimotor contingency. So, for example, it is in the pattern of changing P-shape of the table as one moves around it (a pattern in the structure of sensorimotor contingency) that informs the perceiver of the rectangularity of the table' (2004, p.103).

As this quotation brings out, it is patterns of p-shape-change that are doing the theoretical work in Noë's theory. According to Noë, we see patterns of p-shape-change and understand that the nature of the pattern-change is partly determined by changes in our spatial relation to the object seen. Because we understand the way in which changes in our spatial relation to the object seen impacts the pattern of change we experience, we can determine which aspects of the perceived pattern-change are the result of our own movement relative to the object, and which aspects of the pattern-change are determined by the actual (perspective-independent) properties of the object seen. This, in essence, is what Noë's theory amounts to. So it is patterns of p-properties, rather than individual p-properties, that do the theoretical heavy lifting in Noë's theory. And since appearance-patterns just are patterns of p-property-unfolding, they can do exactly the same theoretical work. The claim that we see the individual p-properties that make up these patterns doesn't need to feature in Noë's explanation of our ability to perceive perspective-independent properties. Once we notice this point, we can see that it is the perception of appearance-patterns, not p-properties, that is crucial to Noë's account of perspective-independent property perception.

Notice also that there is no reason to think that the appearance-pattern account developed here should pose any additional problems for the idea that sensorimotor knowledge might be understood as a kind of s-representation. The basic idea behind this proposal, canvassed in §3 of chapter 4, was that sensorimotor knowledge might be understood as the emulation-based prediction of the sensory consequences of certain movements. To remain consistent with the current proposal, an emulator would have to take as input a set of motor commands, and create as output an anticipated patterned unfolding in the way things look (a certain appearance-pattern, in other words). Evidently, working out exactly how such emulation might work is a complex task. But all I want to establish here is that endorsing the appearance-pattern view over the p-properties view does not create extra difficulties for this kind of story.

§7: Some More Advantages of the Appearance-Patterns View

We've already seen some of the potential advantages of the appearance-patterns view, but I want to highlight another potential consideration in its favour. One advantage of the appearance-pattern view is that it fits better than Noë's view with the Gibsonian notion of optic flow. Noë claims that there is a close affinity between his own and Gibson's view (e.g. 2004, p.85). But the appearance-patterns view is much closer than Noë's to Gibson's theory. In fact, Gibson explicitly criticises a view very like Noë's conception of p-properties.[103] Proximity to Gibson on this point is a desideratum because considerable research has been conducted into optic flow, providing explanatory models for certain perceptual capacities. For example, Withagen and van der Kamp (2010) describe the work on the optical variable 'tau', which 'specifies the time-to-contact between an object and a moving animal. More precisely, the inverse of the relative rate of change of optical angle subtended at the point of contact relates one-to-one to the time remaining before contact' (2010, p.150). The variable referred to is a measure of the looming phenomenon (the fact that objects loom larger in our perceptual field as we approach them). The idea is that, if we are close to an object, the same amount of movement towards that object will create a much bigger looming effect than it would if we were further away. So, as a gannet plummets towards the water, the looming effect will become faster and faster as it gets close to the water into which it is about to plunge. Tau is a measurement of this increase in looming-speed. When tau reaches a certain value, the gannet must tuck in its wings to prepare for contact with water. This variable has been seen as important because it elegantly explains what seem to be cognitively demanding real-time calculations (like that involved in the gannet case) with relative ease. The gannet needs only to be sensitive to one variable: tau. It does not need to make implicit calculations of speed, distance etc. which would be difficult to make accurately in such real-time scenarios. The variable has also been used to model the

[103] Consider the following quotations from Gibson's *Ecological Approach To Perception*: 'The optic array *flows in time* instead of going from one structure to another [...] When the moving point of observation is understood to be the general case, the stationary point of observation is more intelligible. It is no longer conceived as a single geometrical point in space but as a pause in locomotion, as a temporarily fixed position relative to the environment.' (1986, p.75) Appearance-patterns fit better with Gibson's description of the dynamic nature of perception. According to Gibson, perception from a fixed perspective is a special case of perception, rather than the basic phenomenon in terms of which dynamic perception must be explained (1986, p.2).

ability of somersaulters to land on their feet, and the ability of sportspeople to catch balls.[104]

This is not the right place to go into a detailed defence of these models, but there's a *prima facie* case for thinking that Noë would be better off with a theory that fits well with this line of research. To see how the appearance-pattern would fit with this research, notice that looming is an appearance-pattern; it is a pattern of change to the apparent size of an object. To fit with the research on 'tau', we need to hypothesise only that perceivers can be sensitive to features of these appearance-patterns, features like their rates of change. Noë's theory adds to this picture a superfluous extra first step: we experience appearances, we somehow synthesise these into appearance-patterns, and only then can be we sensitive to features of this pattern. Adding this extra step undermines the primary appeal of the tau-variable model: its elegant simplicity.

§8: Conclusion

As we have seen, the 'appearance-pattern' notion is more useful explanatorily than is the 'p-property' notion. By explaining the perception of perspective-independent properties in terms of appearance-pattern-perception, we avoid the problem of invisible contents. And we accrue some additional explanatory advantages in the process.

[104] See Chemero (2009) for an overview of this literature.

Afterword

In the first half of the thesis, I developed a theory of content for s-representations. I did so in part by drawing on some resources developed by writers in the common-sense functionalist tradition, writers whose position I had, in chapter 1, defended from Sprevak's *reductio*. I then gave some general reasons for thinking that the 's-representation' notion, for which I had provided a theory of content, might be biologically ubiquitous and cognitively significant. In the second half of the thesis, I put the notion of s-representation to work by showing how it might be used to shed light on the notion of 'sensorimotor knowledge' central to Noë's theory of perception.

As well as using Noë's theory as a testing ground for s-representations, I also gave a detailed evaluation of some of Noë's key theoretical commitments. I started with what I took to be some of his least promising (though perhaps most notorious) ideas about the role of action in perception. I gave reasons for finding many of these claims unappealing. Next, I identified what I took to be two of Noë's most interesting ideas: his virtual content theory and his theory of how it is that we get from perception of perspectival properties to the perception of perspective-independent properties. I argued that both these theories need serious modification if they are to hold water. And I made what I took to be the necessary modifications. I also showed that both these theories rely crucially on the notion of 'sensorimotor knowledge' which, as I have argued, might usefully be treated as a form of s-representation.

It might be noted that the first chapter, in which I dealt with the extended mind thesis, is rather detached from the rest of the thesis. This structural feature reflects my view on the relation between the extended mind thesis and the other positions I have been developing. As I hope I made clear, it is my view that these positions swing free of the extended mind thesis. And given that this is my view (a view that seems not to be shared by many of the writers whose work I have examined) I have tried to develop the positions I wish to defend without taking a stand on the 'extended mind question'.

References

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

Adams, F. and Aizawa, K. 2007. *The Bounds of Cognition*. Oxford: Blackwell.

Adams, F. 2010a. Why we still need a mark of the cognitive. *Cognitive Systems Research*. 11, pp. 324-331

Adams, F. and K. Aizawa. 2010b. Defending the Bounds of Cognition. In *The Extended Mind,* ed. R. Menary. Cambridge, Mass.: MIT Press.

Aizawa, A. 2007 Understanding the Embodiment of Perception. *Journal of Philosophy*, *CIV* (1), pp. 5-25

Anscombe, G.E.M. 1957. *Intention*. Oxford: Basil Blackwell.

Bechtel, W. 1994. Natural Deduction in Connectionist Systems. *Synthese* 101 (3), pp. 433-463

Bickhard, M. 1999. Interaction and Representation. *Theory Psychology*. 9, pp. 435-58

Block, N. 1990. Inverted Earth. *Philosophical Perspectives.* 4, pp. 53-79

Block, N. 2005. Review of Alva Noë, *Action in Perception. Journal of Philosophy* 102, pp. 259-272

Bortolotti, L. 2009. *Delusions and Other Irrational Beliefs*, Oxford: Oxford University Press.

Bortolotti, L. 2010. Double bookkeeping in delusions: Explaining the gap between saying and doing. In *New waves in the philosophy of action*, eds. K. Frankish, A. Buckareff, and J. Aguilar. Palgrave

Brooks, 1991. Intelligence without representation. *Artificial Intelligence* 47, pp. 139-159

Carman, T. 2009. Merleau-Ponty and the Mystery of Perception. *Philosophy Compass*. 4/4: 630-638

Carruthers, P. 2006: *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought.* Oxford: Oxford University Press

Chalmers, D.J, Jackson, F. 2001. Conceptual analysis and reductive explanation, *Philosophical Review,* 110:3, pp. 315-61

Chemero, A. 2009. *Radical Embodied Cognitive Science.* Cambridge, Mass.: MIT Press

Choi, S. and Fara, M. Dispositions, *The Stanford Encyclopedia of Philosophy*. E. Zalta (ed.) URL = <http://plato.stanford.edu/archives/spr2014/entries/dispositions/>.

Clark, A. 1997 *Being There*. Cambridge, Mass.: MIT Press

Clark, A. 2003. *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence.* New York: Oxford University Press.

Clark, A. 2005. Intrinsic content, active memory, and the extended mind. *Analysis*, 65(285), pp. 1-11.

Clark, A. 2008a. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension.* Oxford: Oxford University Press.

Clark, A. 2008b. Pressing the flesh: A tension in the study of embodied, embedded mind? *Philosophy and Phenomenological Research*. 76 (1), pp. 37-59

Clark, A. 2009. Spreading the joy? Why the machinery of consciousness is (probably) still in the head. *Mind*. 118:472, pp. 963-993

Clark 2010a. Finding the Mind. *Philosophical Studies,* 152.3, pp. 447-61

Clark 2010b. *Memento's* Revenge: The Extended Mind, Extended. In R. Menary (ed.) *The Extended Mind*. Cambridge, MA.: MIT Press, pp.43-66

Clark, A. 2010c Coupling, Constitution and the Cognitive Kind. In R. Menary (ed.) *The Extended Mind*. Cambridge, MA.: MIT Press, pp.81-100

Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioural and Brain Sciences,* 36:3, pp. 1-73.

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

Clark, A. and Eilan, N. 2006. Sensorimotor Skills and Perception. *Proceedings of the Aristotelian Society, Supplementary Volumes*. 80, pp. 43-65

Clark, A., and Grush, R. 1999. Towards a cognitive robotics. *Adaptive Behaviour,* 7, pp. 5-16

Clarke, T. 1965. Seeing surfaces and Physical Objects. In M. Black (ed.), *Philosophy in America*. London: George Allen.

Cummins, R. 1989. *Meaning and Mental Representation.* Cambridge, MA: MIT Press.

Cummins, R. 1991. The role of representations in connectionist explanations of cognitive capacities. W. Ramsey, S. Stich, D. Rumelhart *Philosophy and Connectionist Theory*. Hillsdale, NJ: Lawrence Erlbaum, pp. 91-114.

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

Davidson, D. 1987. Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, 61, pp. 441–58

Dawkins, R. 1982. *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press.

Dennett, D. C. 1969. *Content and Consciousness*. London: Routledge & Kegan Paul.

Dennett, D. C. 1991. *Consciousness Explained*. Boston: Little, Brown.

Dennett, D.C. & Kinsbourne, M. 1992. Time and the observer: The where and when of consciousness in the brain. *Behavioural Brain Sciences*. 19, pp.183-247

Di Paolo, E. 2009. Extended Life. *Topoi.* 28, pp.9-21.

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

Dreyfus, H. L. 1992: *What Computers Still Can't Do: A Critique of Artificial Reason*. New York: MIT Press.

Dreyfus, H. L. (2007) Why Heideggerian AI Failed and How Fixing It Would Require Making it More Heideggerian. *Philosophical Psychology* 20:2, pp. 247-268

Duhamel, J., Colby, C., and Goldberg, M.E. 1992. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255* (5040), pp. 90-02

Fara, M., 2005. Dispositions and Habituals, *Nous*. 39, pp. 43-82

Fernandez-Duque, D., & Thornton, I.M. 2000. Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system. *Visual Cognition*. 7, pp. 324-344

Findlay, J. M., & Gilchrist, I. D. 2003 *Active Vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Fodor, J. 1975. *The Language of Thought*. New York: Thomas Cromwell.

Fodor, J. 1981. *RePresentations*. Cambridge, MA: MIT Press.

Friston, K. J. 2010. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* 11:2, pp. 127-38

Gangopadhyay, N. 2010. Experiential blindness revisited: In defence of a case of embodied cognition. *Cognitive Systems Research.* 11, pp. 396-407

Gettier, E. 1963. Is Justified True Belief Knowledge? *Analysis*, 23, pp. 121-3.

Gibson, J. J. 1986. *An Ecological Approach to Visual Perception*. New York: Taylor and Francis Group.

Godfrey-Smith, P. 2006 Mental representation, naturalism and teleosemantics. In *Teleosemantics,* ed. G. Macdonald and D. Papineau. Oxford University Press.

Grush, R. 1995. *Emulation and cognition*. PhD Dissertation, UC San Diego Cognitive Science and Philosophy, UMI.

Grush, R. 1997. The architecture of representation. *Philosophical Psychology* 10:1, pp. 5-15

Grush, R. 2003. In Defence of some 'Cartesian' Assumptions Concerning the Brain and its Operations. *Biology and Philosophy* 18:1, pp. 53-93

Grush, R. 2004. The emulation theory of representation: motor control, imagery and perception. *Behavioural and Brain Sciences* 27:3, pp. 377-396

Grush, R. 2007. Skill theory v2.0: dispositions, emulation and spatial perception. *Synthese* 159, pp. 389-416

Grush, R. 2008. Review of William M. Ramsey, *Representation Reconsidered*. *Notre Dame Philosophical Review*.

Grush, R. (manuscript) *The Machinery of Mindedness*

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea.* Cambridge, Mass.: MIT Press.

Heil, J. 2004. *Philosophy of Mind: A Contemporary Introduction*. New York: Routledge.

Hohwy, J. and Rajan, V. 2012. Delusions are forensically disturbing perceptual inferences. *Neuroethics*, 5:1, pp. 5-11

Hurley, S. 1998. *Consciousness in Action.* Cambridge, MA.: Harvard.

Hurley, S. 2010. Varieties of Externalism. In R. Menary (ed.) *The Extended Mind*. Cambridge, MA.: MIT Press, pp.101-154

Hurley, S. and Noë, A. 2003. Neural Plasticity and Consciousness. *Biology and Philosophy*, 18, pp. 131-68

Hutto, D., 2005. Knowing what? Radical versus conservative enactivism. *Phenomenology and the Cognitive Sciences* 4, 389-405

Hutto, D., and Myin, E., 2013 *Radicalizing Enactivism: Basic Minds without Contents*. Cambridge, Mass.: MIT Press.

Hyman, J. (2006) *The Objective Eye: Colour, Form and Reality in the Theory of Art*. Chicago University Press, Chicago

Jackson, F. 1995. Mental Properties, Essentialism and Causation, *Proceedings of the Aristotelian Society*, 95, pp. 253-68

Jackson, F., Braddon-Mitchell, D. 2007. *Philosophy of Mind and Cognition: An Introduction*. 2nd ed. Oxford: Blackwell Publishing

Kaufman, E.L., Lord, M.W., Reese, T.W., & Volkmann, J. (1949) The discrimination of visual number. American *Journal of Psychology.* 62 (4), pp. 498–525.

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

Land, M. F., Furneaux, S. M., & Gilchrist, I. D. 2002. The organisation of visually mediated actions in a subject without eye movements. *Neurocase*, 8, pp. 80-87

Lewis, D. 1969. *Convention*. Cambridge: Harvard University Press.

Lewis, D. 1980a. Mad Pain and Martian Pain. In N. Block (ed.) *Readings in Philosophy of Psychology*, *Vol. 1.* Cambridge, MA: Harvard University Press, pp. 216-222

Lewis, D. 1980b. Veridical Hallucination and Prosthetic Vision. *Australasian Journal of Philosophy*, 58, pp. 239-249

Lewis, D. 1994. Reduction of Mind. In S. Guttenplan (ed.) *A Companion to the Philosophy of Mind.* Oxford: Blackwell, pp. 412-31.

Lycan, W. 1982. Towards a Homuncular theory of believing. *Cognition and Brain Theory* 4: 139-59.

Macdonald, G. and Papineau, D. (eds.) 2006. *Teleosemantics: New Philosophical Essays.* Oxford: Clarendon Press.

Martin, M. G. F. (2004). The limits of self-awareness. *Philosophical Studies,* 120**,** pp.37-89

Martin, M.G.F. (2008) Commentary on *Action in Perception*. *Philosophy and Phenomenological Research*. Vol. LXXVI, No. 3, pp. 691-706.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. 2004. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5, ppp. 229-239

Martinez-Conde, S. Macknik, S. L. Troncoso, X. G., & Dyar, T. A. 2006. Microsaccades counteract visual fading during fixation. *Neuron, 49, pp.* 297-305

Mitroff, S. R., Simons, D. J., Franconeri, S. L. 2002. The siren song of implicit change detection. *Journal of Experimental Psychology – Human Perception and Performance*. 28:4, pp. 798-815

Millikan, R. 1989 Biosemantics *The Journal of Philosophy*, 86:6, pp. 281-297

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

Millikan, R. 1996. Pushmi-pullyu representations. *Philosophical Perspectives,* IX, pp. 185-200

Millikan, R. 2005 *Language: A Biological Model*. Oxford: Clarendon Press.

Minsky, M. 1985. *The Society of Mind*. New York: Simon & Schuster.

Noë, A. 2004. *Action in Perception*. Cambridge MA: MIT Press.

Noë, A. 2006. Précis of *Action in Perception*. *Psyche Electronic Journal* 12(1).

Noë, A. 2008. Reply to Campbell, Martin and Kelly. *Philosophy and Phenomenological Research*. Vol. LXXVI, No. 3, pp. 691-706.

Noë, A. 2009a. Conscious References. *Philosophical Quarterly* 59:236, 470-82.

Noë, A. 2009b. *Out of Our Heads.* New York: Hill and Wang.

Noë, A. 2012. *Varieties of Presence*. London: Harvard University Press.

Noë, A. and Thompson, E. 2004a. Are There Neural Correlates of Consciousness? *Journal of Consciousness Studies,* 11, pp. 3-28.

Noë, A. and Thompson, E. 2004b. Sorting Out the Neural Basis of Consciousness: Authors' Reply to Commentators. *Journal of Consciousness Studies*, 11, pp. 87-98

O'Regan, J. K. 2011. *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*. Oxford: Oxford University Press.

O'Regan, J.K., Rensink, R.A., & Clark, J.J. 1999. Change-blindness as a result of "mudsplashes". *Nature*, 398, 34.

Palmer, S. E. 1999. *Vision Science: Photons to Phenomenology*. Cambridge, MA.: MIT Press.

Peirce ,C. S., 1931-58 *The Collected Paperes of C.S Peirce, vols 1-8.* A. Burks, C. Hartshorne, and P. Weiss (eds). Cambridge, Mass. Harvard.

Pierce, C.S. 1998 *The Essential Pierce: Volume 2*. Indiana: Indiana University Press.

Prinz, J. 2009. Is Consciousness Embodied? In P. Robbins, M. Aydede (eds.) *Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press, pp. 419-436.

Putnam, H. 1975/1985. The meaning of 'meaning'. *In Philosophical Papers, Vol. 2: Mind, Language and Reality*. Cambridge University Press.

Ramsey, W. 2007. *Representations Reconsidered*. Cambridge: Cambridge University Press.

Rowlands, M. 2006. *Body Language: Representations in Action.* Cambridge, Mass.: MIT Press.

Rowlands, M. 2007 Understanding the 'active' in 'enactive'. *Phenomenology and Cognitive Science* 6, 427-443

Rowlands, M. 2009. Extended Cognitions and the Need for a Mark of the Cognitive. *Philosophical Psychology,* 22:1, pp. 1-20

Rowlands, M. 2010a. *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. London: MIT Press

Rowlands 2010b. Consciousness, Broadly Construed. In R. Menary (ed.) *The Extended Mind*. Cambridge, MA.: MIT Press, pp. 271-294

Rumelhart, D., McClelland, J., and the PDP Research Group. 1986. *Parallel Distributed Processing* , 3 vols. Cambridge, Mass.: MIT Press.

Rupert, R. 2004. Challenges to the hypothesis of extended cognition. *Journal of Philosophy.* 101, pp. 389-428.

Rupert, R. 2009. *Cognitive Systems and the Extended Mind*. Oxford: Oxford University Press.

Rupert, R. 2010. Extended Cognition and the Priority of Cognitive Systems. *Cognitive Systems Research*, 11, pp.343-356

Rutkowska, J. C. 1994. Emergent functionality in human infants. In Cliff, Husbands, Meyer, and Wilson (eds) *From Animals to Animats 3: Proceedings of the Third Conference on Simulation of Adaptive Behavior. Cambridge, Mass.:* MIT Press/Bradford Books, pp. 179-188

Ryder, D. 2004. SINBAD Neurosemantics: A Theory of Mental Representation *Mind and Language*. 19:2, pp. 211-40

Shapiro, L. 2004. *The Mind Incarnate*. Cambridge: Cambridge University Press.

Smith, J. Forthcoming. Egocentric Space. *International Journal of Philosophical Studies.*

Shapiro, L. 2008. Functionalism and Mental Boundaries. *Cognitive Systems Research*, 106, pp. 503-527

Simon, T. and Vaishnavi, S. 1996. Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages. *Perception & Psychophysics,* 58(6), pp. 915-926

Simons, D. J., Levin, D. T. 1998. Failure to detect change during real-world interaction. *Psychonomic Bulletin and Reviews* 5: pp. 644-649.

Simons, S. J., Chabris, C. F. 1999. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception* 28: pp. 1059-1074.

Sprevak, M. 2009. Extended Cognition and Functionalism. *Journal of Philosophy*. 106, 9, pp. 503-527

Sprevak, M. 2010. Inference to the Hypothesis of Extended Cognition. *Studies in the History and Philosophy of Science*. 41:4, pp. 353-362

Sprevak, M. 2011. Representations Reconsidered (book review). *British Journal for the Philosophy of Science*. 62: 3, pp. 669-675

Sterelny, K. 2004. Externalism, epistemic artefacts, and the extended mind. In *The externalist challenge* ed. Richard Schantz. New Studies on Cognition and Intentionality. New York: de Guryter.

Strawson, P. F. 1979. Perception and its Objects. In A. Noë and E. Thompson (eds.) (2002) *Vision and Mind: Selected Readings in the Philosophy of Perception*. Cambridge, Mass.: MIT Press.

Swoyer, C. 1991. Structural Representations and surrogative reasoning. *Synthese.* 87, pp. 449-508

van Gelder, T. 1995. What Might Cognition Be If Not Computation? *Journal of Philosophy.* 92:7, pp. 345-381

Wadham, J. 2014. The Problem of Invisible Contents. In Bishop, J.M. and Martin, A.O. (eds) *Contemporary Sensorimotor Theory: Studies in Applied Philosophy, Epistemology and Rational Ethics (SAPERE)*, Vol. 15, London: Springer, pp. 117-126

Wadham, J. Forthcoming. Nomological Necessity, Noë and Merleau-Ponty. *International Journal of Philosophical Studies.*

Wheeler, M. 2005. *Reconstructing the Cognitive World*. London: MIT Press.

Wheeler, M. 2010a. Minds, things, and materiality. In *The Cognitive Life of Things,* ed. C. Renfrew and L. Malafouris. Cambridge University Press, pp. 29-37

Wheeler, M. 2010b. In defense of extended functionalism. In *The Extended Mind*, ed. R. Menary. Cambridge, Mass.: MIT Press, pp. 245-270

Wilson, R. 2010. Meaning Making and the Mind of the Externalist. In *The Extended Mind*, ed. R. Menary. Cambridge, Mass.: MIT Press, pp. 167-188

Withagen, R., van der Kamp, J. 2010. Towards a new ecological conception of perceptual information: Lessons from a developmental systems perspective. *Human Movement Science*. 29, pp. 149-163