Durham E-Theses

Consciousness and the Heterogeneity of Attention

TAYLOR, JOHN,HENRY

How to cite:

TAYLOR, JOHN,HENRY (2015) *Consciousness and the Heterogeneity of Attention*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/10973/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, Durham University, University Office, Old Elvet, Durham DH1 3HP
e-mail: e-theses.admin@dur.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

CONSCIOUSNESS AND THE HETEROGENEITY OF ATTENTION

Thesis Submitted for the Degree of

Doctor of Philosophy

by

John Henry Taylor

Department of Philosophy

University of Durham

2014

This thesis is dedicated to the memory of Edward Jonathan Lowe.

1950-2014.

Declaration

I confirm that no part of the material contained in this thesis has previously submitted for any degree at this or any other university. All of the material is the author's own work, except discussion of others' work, which has been appropriately indicated in the text. Some of this work has appeared in print before, and has been appropriately cited.

The copyright of this thesis rests with the author. No quotation from it should be published without prior written consent from the author and information derived from it must be acknowledged.

Acknowledgements

First and foremost, I would like to thank my supervisors. Thanks to Jonathan Lowe for giving over many hours of his time, even though he was the busiest person I have ever met. This thesis is dedicated to him, as it has been the greatest honour of my life to have been his friend. Thanks also to Peter Vickers, for volunteering to take over when Jonathan tragically became ill and died in January of 2014. Pete has dedicated colossal amounts of his time to discussing and debating issues with me, despite having a million and one more important things to do. Finally, thanks to my secondary supervisor Sophie Gibb, who has offered constant support and guidance ever since I began my journey into academic philosophy. Jonathan, Pete and Sophie have all shown huge amounts of patience with me, even in my most unreasonable and neurotic moments. This piece of work would not exist without them.

Just as important as the support of my supervisors has been that of Bob Kentridge from Durham University's psychology department. I am ridiculously lucky that I ended up studying attention at a university that just happened to have Bob in it. Were it not for Bob's magnanimous and generous nature, I would not have had the confidence to discuss his work in a philosophical manner.

I would like to thank Durham University for offering me a fellowship which sustained me for the first three years of my PhD. Thanks also to the Royal Institute of Philosophy, which supported me financially through the final year.

Thanks to my parents, Karen and Mark Taylor, both of whom have worked for decades in jobs they hate in order to give me opportunities that they could never have dreamed of. They have taught me that the best career move you could ever make is to do something you enjoy. Without this lesson, I would never have studied philosophy.

Thanks to Natasha Kriznik, for constant mutual emotional support. In the same spirit, thanks to all of my friends, for making my PhD years immensely more fun than they would have been otherwise.

I have received huge amounts of helpful comment and criticism over the course of this thesis. Just a small number of those I would like to thank for this are the following: Keith Allen, Bernard Baars, Ned Block, Alan Bowden, David Chalmers, Tony Cheng, Tim Crane, Felipe De Brigard, Naiomi Eilan, Marc Ereshefsky, Philip Goff, Mette Hansen, Clare MacCumhaill, Mohan Matthen, Paul Noordhof, Matt Nudds, Aude Oliva, David Papineau, Ron Rensink, James Stazicker and Wayne Wu.

Some parts of this thesis have appeared in print before. Chapter 2 is an unrecognisable descendent of:

Taylor, J. H. 2013a. "Is attention necessary and sufficient for phenomenal consciousness?" *Journal of Consciousness Studies.* 20 (11-12), pp.173-194.

The appendix of this thesis appeared (with only minor changes) as:

Taylor, J. H. 2013b. "Is the grain of vision finer than the grain of attention? Response to Block." *Thought: a journal of philosophy.* 2 (1), pp.20-28.

The ancient ancestor of chapter 4 is:

Taylor, J. H. 2014. "Against unifying accounts of attention." *Erkenntnis.* Available online. DOI: 10.1007/s10670-014-9611-3.

Abstract

This thesis will primarily consider issues related to attention and its relationship to phenomenal consciousness. The main argument of this thesis is that we cannot hope to make progress in this domain unless we greater appreciate the role that the concept 'attention' plays within these debates. I argue that by focussing on this, we can reach new and surprising conclusions about both the nature of attention, and its relationship to other philosophically interesting faculties of the mind.

I begin by offering a summary of the arguments of the thesis, and a brief history of the debates I will be addressing. Then I turn to a detailed examination of the debates over whether attention is necessary for consciousness or sufficient for consciousness, with a particular emphasis on the latter of these. I examine much of Robert Kentridge's work in this regard. I will argue that this debate has reached an impasse, and that the point of apparent disagreement is conceptual, not empirical. I then go on to investigate the concept which is the point of friction, which is 'attention'. I investigate various ways of resolving the difficulties in this debate, and argue that none of them work. I then examine some particular accounts of attention that have been offered in the philosophical literature. I argue against all of them, and also argue that the underlying assumptions on which they are based should be rejected.

I then suggest my own positive proposed solution to these problems. I use elements from the material so far covered to build an original argument in favour of pluralism about attention: the view that 'attention' is ambiguous between several importantly different concepts, no one of which is privileged or 'more correct' than the others, and several of which are worthy of acceptance in our theorising. I defend this view extensively from

criticisms, and apply it to the issues that have been raised in the former part of the thesis. I also expand on this position by putting forward an argument for *eliminativism* about 'attention': the view that we should abandon using the term 'attention' in certain contexts in psychology and philosophy. In the course of making these arguments, I also consider various issues to do with the natural kinds of psychology, and classification in general. I also include an appendix where I discuss one particular argument in favour of the claim that attention is *unnecessary* for consciousness.

Table of Contents

Chapter 1 Introduction....13
§1.0-Summary....13
§1.1-Preamble....13
§1.2-Methodology....14
§1.3-A brief summary of the project....16
§1.4-The debates....19
§1.4.1-Is attention necessary for consciousness?....19
§1.4.2-Is attention sufficient for consciousness?....27
§1.4.3-An explanation of consciousness?....30

Chapter 2 Attention and consciousness....33
§2.0-Summary....33
§2.1-Conceptual preliminaries....33
§2.2-Evidence against the sufficiency claim....35
§2.2.1-Blindsight and the sufficiency thesis....36
§2.2.2-Rival views on the data....42
§2.2.3-The meta-contrast masking experiments....49
§2.2.4-De Brigard and Prinz vs Kentridge....55
§2.2.5-Orienting?....61
§2.2.6-A reason to accept Prinz's definition of attention?....63
§2.3-Declan Smithies' argument in favour of ST....66
§2.4-More empirical data relevant to ST....69
§2.4.1-The erotic images studies....70
§2.4.2-The Soto et al. data....74
§2.4.3-A summary of the situation we are in....78
§2.5-Is Attention Necessary for Consciousness?....79
§2.6-Could any empirical data refute De Brigard and Prinz's position?....83
§2.7-Why is this important?....86
§2.8-Two possible reactions and the nature of attention....87

§2.8.1-The two reactions. 87
§2.8.2: Monism and pluralism. 88
§2.9-What next? 90

Chapter 3 The Road to Pluralism: first steps 91
§3.0-Summary. 91
§3.1-Solving the problems of chapter 2. 91
§3.2-Unpacking the claim: Folk psychology. 93
§3.3-What the folk say about attention. 96
§3.3.1-Experimental philosophy 96
§3.3.2-Comparing Mole and De Brigard. 98
§3.3.3-More experimental philosophy. 102
§3.3.4-What the folk say: more general concerns. 105
§3.3.5-An objection and a reply. 112
§3.3.6-Folk psychology: Conclusions. 115
§3.4-Empirical Psychology 115
§3.4.1-Back to the debates over ST. 116
§3.4.2-Triviality. 120
§3.4.3-Empirical psychology and pluralism. 121
§3.4.4-An objection and a reply. 122
§3.5-Theoretical Use. 124
§3.5.1-Plethora of theoretical uses 126
§3.5.2-Objection: the future of attention research. 133
§3.6-The meta-question. 134
§3.6.1-From Folk Psychology to Empirical Psychology and back again? 136
§3.7-Some reasons to accept pluralism. 138
§3.8-Conclusion. 141

Chapter 4: The Attention Essentialist Programme 142
§4.0-Summary. 142
§4.1-Attention essentialism. 142

§4.1.1-Attention essentialism: what are we trying to do? 144
§4.2-Judging attention essentialist theories. 146
§4.2.1-Attention essentialism and pluralism. 148
§4.3-Jesse Prinz. 150
§4.3.1-Replies. 153
§4.3.2-Extending the criticism to other accounts: Wayne Wu. 156
§4.4-Sebastian Watzl. 159
§4.5-The prospects of attention essentialism. 163
§4.5.1-Failure of individual accounts and paucity of argument. 164
§4.5.2-The nature of concepts. 168
§4.5.3-Is 'attention' a special case? 173
§4.6-Theoretical use. 174
§4.6.1-Specific concepts and their uses. 175
§4.7-Putting the pluralistic pieces together. 180
§4.8-An objection to the master argument. 184
§4.9-What makes them all 'attention' concepts? 185
§4.10–Eliminativism? 189
§4A-Appendix: Christopher Mole's 'cognitive unison' theory of attention. 190
§4A.1-Mole's theory. 190
§4A.2-Problems over divided attention. 191
§4A.3-Replies. 195
§4A.4-Mole's view and Lavie's 'perceptual load' theory 196

Chapter 5: A Road to Eliminativism? 199
§5.0-Summary. 199
§5.1-The eliminativist challenge. 199
§5.2-The argument from analogy. 201
§5.3-Some similar debates. 203
§5.3.1-The general issue. 206
§5.4-Two kinds of eliminativism. 207
§5.5-Resisting the slide to ontological eliminativism. 209

§5.5.1-Step one: predicates, properties and truthmakers. 210
§5.5.2-Step two: 'Attention' and realism. 213
§5.6-Pragmatic eliminativism. 217
§5.6.1-The argument for pragmatic eliminativism. 219
§5.6.2-Premise (2). 219
§5.6.3-Premise (3) and monism. 227
§5.6.4-Pragmatic eliminativism and realism. 233
§5.7-Another reason to be an eliminativist about 'attention'? 235
§5.8-A last objection. 236
§5.9-Conclusion. 237

Chapter 6: Psychological Eliminativism and Natural Kinds 238
§6.0-Summary. 238
§6.1-The natural kinds question and three eliminativist arguments. 238
§6.2-What is a natural kind? 241
§6.2.1-Two red herrings 242
§6.2.2-A more promising approach? 244
§6.3-A space of kinds. 248
§6.4-Two case studies. 252
§6.4.1-Concepts. 252
§6.4.2-Emotions. 255
§6.5-Natural kinds and attention. 259
§6.5.1-Two natural kinds-based objections to my view 259
§6.5.2-'Mechanisms' 263
§6.5.3-Endogenous and exogenous attention again. 264
§6.6-Getting to eliminativism without natural kinds. 266
§6.7-An objection and a reply. 270
§6.8-Conclusion. 272

Chapter 7: Theoretical Use, Classification and the Future of Attention Research 273
§7.0-Summary. 273

§7.1-The story so far. 273
§7.2-Theoretical use and pluralism. 277
§7.2.1-Tomatoes and fruit. 277
§7.2.2-Theoretical uses and the different 'attention' concepts. 279
§7.2.3-Objection: redundancy. 282
§7.3-Where do we go from here? 284
§7.4-Concluding thoughts. 287

Appendix: Is the Grain of Vision Finer than the Grain of Attention? Response to Block. 290
§A.0-Summary. 290
§A.1-Attention and Crowding. 290
§A.2-An analysis of Block's argument. 292
§A.3-Objections and replies. 296
§A.3.1-Objection 1. 296
§A.3.2-Objection 2. 298
§A.3.3-Objection 3. 300
§A.4-Unconscious Seeing. 301
§A.5-Conclusion. 303
Bibliography 304

Chapter 1
Introduction

§1.0-Summary.

In this chapter the methodology of the thesis will be outlined. Then, some of the core claims and arguments of the thesis will be summarised and a brief history of the debates this thesis is concerned with shall be given.

§1.1-Preamble.

Here is a question: how much are you conscious of right now? For my own part, I think it's a pretty good bet that you're conscious of the word that you're currently focussing on, but how much more are you conscious of? Are you peeking out at the world through a tiny window, large enough to only capture a few details of your environment? Does your consciousness capture only a few scraps about the external world? Or do you possess a rich and intricately detailed phenomenal field? Are you conscious of the feeling of your feet in your shoes? Or your tongue resting on the bottom of your mouth? Or your shirt on your back? Is consciousness like a flash in the pan? Or is it a rich and varied landscape?

These are striking and fascinating questions. However, we cannot begin to answer them without also examining the knotty and difficult issue of how consciousness and attention interact. In the contemporary philosophy of psychology literature, there are a host of different debates concerning all aspects of how attention and consciousness relate to each other. Among the debates that I shall be examining are the question of whether there can be consciousness outside of attention, and whether there can be attention outside of consciousness.

We will begin by looking at these questions, and attempting to assess various sides of the debates around them, but we will soon find that in order to answer these questions,

we will have to answer some much more fundamental questions. One of these prior questions will be 'what is attention?' We will be examining this more basic question in some detail, and I will be arguing in favour of a view of attention which is currently quite unpopular in the philosophical literature on attention. The view I have in mind can be called 'pluralism'. I shall be arguing that the term 'attention' is ambiguous between several different meanings, which refer to different entities in the world. I shall also be arguing that no one of these meanings is 'privileged' or 'correct' above the others. Finally, I shall argue that we have good reason to accept several of these alternative meanings of 'attention' in our theoretical discourse. For these reasons, I argue that it is misguided to attempt to develop one unified and privileged account of all of the phenomena that we call 'attention'. If we accept this view (as I argue we should), then we can draw some novel answers to the question of whether attention is necessary and/or sufficient for consciousness.

Pluralism about attention is not original to me. However, the current piece of work represents the first sustained, dedicated and developed argument in favour of it, and defence of it from criticism. But already we're getting ahead of ourselves. We should start with a few comments on the methodology I'll be using and the history of the debates I'll be looking at.

§1.2-Methodology.

This thesis is a piece of work in the philosophy of psychology and philosophy of mind. I will give a philosophical analysis of certain experiments which have been particularly prominent in the empirical and philosophical literature regarding the interaction between attention and consciousness. I aim to raise some problems for the debate as a whole. In the process of examining various solutions to these problems (and suggesting

my own solution), I will comment on various philosophical theories of attention, develop one of my own and discuss some of the work on attention within empirical psychology.

Though I believe that a great many debates in philosophy could benefit from a more close interaction with empirical work, it cannot be denied that doing philosophy whilst drawing upon such work is extremely difficult. When it is done well, what can result is a vital perspective on a particular set of problems. However, drawing on another discipline brings with it a host of methodological difficulties and attempting to say something empirically well-informed and philosophically substantive can be a nightmare. For these reasons, when I discuss scientific work I will try to ensure that the claims that I rely upon are widely accepted within empirical psychology. When I invoke a claim which is more controversial, I shall explicitly make the reader aware of this. I shall also try to restrict my main discussion to a small number of particular experiments, which I shall discuss in detail. I think this minimises the chances of making egregious slips when it comes to the scientific details.

I will have much to say from a metaphilosophical perspective. I am interested in the claims that certain thinkers make, the arguments they put forward and the evidence that is given for particular views. However, I am also very interested in the assumptions that are made by philosophers and psychologists alike; and the different motivations behind the views that are given, which often as not go unstated. In addition to examining the debates themselves, I hope to thoroughly analyse these background issues.

I am hesitant to call myself a naturalist, as that term has been used so frequently that to say that you are a naturalist is pretty much not to say anything at all. However if we understand naturalism to be the view that we should pay attention to empirical work in our

philosophical reasoning where it is relevant, and that it is a very significant criticism of any philosophical theory if it is inconsistent with best science, then I carry a strong commitment to naturalism. In fact, I would insist that no one sensible should deny naturalism, when understood this way.

I should say at the outset that I will *not* be attempting to *refute* positions other than my own, in the sense of showing that they are completely untenable. Nor will I be attempting to give a conclusive *proof* of my own view, in the sense of demonstrating its truth beyond all doubt. I agree with David Lewis when he says that such arguments are almost never successful and philosophers who demand them are usually misguided (1986, viii). Usually, absolute proof of one's own view and decisive refutation of one's opponent's view is completely out of the question in philosophy. Instead, I will be arguing that the overall balance of considerations leans towards my view rather than the others and that this gives us good reason to accept it.

§1.3-A brief summary of the project.

I now turn to summarising the main contributions of each chapter of the thesis. In the remainder of this chapter, I shall give a brief and selective history of the debates that will begin our discussion. In chapter 2, I will examine some of the most prominent debates over whether attention is sufficient for consciousness. I will argue that the disagreement in this debate stems not from disagreement over some empirical facts, but from different thinkers using the word 'attention' in significantly different ways. I argue that this stems from the different thinkers in the debates having very different background assumptions about what counts as 'attention'. If what I argue in this chapter is correct, much of the current debate on whether attention is sufficient for consciousness is misguided. Much of the chapter will be

concerned with the debate over whether attention is sufficient for consciousness, but I will also argue that similar problems beset the debate over whether attention is *necessary* for consciousness.

Having examined these problems in chapter 2, my focus in chapter 3 will be on examining various strategies for solving them. Particularly, I will examine the proposal that we should decide which of the proposed views of 'attention' encountered in chapter 2 should be preferred. I shall examine the strategy of picking the account of 'attention' which coheres best with the use of the term 'attention' in folk psychology, or empirical psychology, or that we should select the concept of 'attention' which is the most theoretically useful. I argue that none of these strategies can help us to solve the problems pointed out in chapter 2. I also use these considerations to put forward a preliminary argument in favour of pluralism about 'attention'.

In chapter 4, I will examine the attempts that have been made to give a reductive analysis of 'attention', where this project is understood as attempting to give necessary and sufficient conditions for attention in a non-circular and explanatory way. In particular, I concentrate on the views of Jesse Prinz, Wayne Wu, Christopher Mole and Sebastian Watzl. I criticise all of their views, and then argue that any reductive analysis of 'attention' of this kind is unlikely to be successful. I will then draw together several ideas developed in chapters 2, 3 and 4 to present my main argument in favour of my particular brand of pluralism about 'attention': the view that 'attention' is ambiguous between several importantly distinct concepts, no one of which is privileged, and *several* of which are worthy of acceptance in our theoretical work. I suggest that we replace the popular views in the philosophical literature pluralism of this stripe.

In chapter 5, I consider the objection that my view leads to *eliminativism.* To assess this criticism, I make distinctions between various different kinds of eliminativism, and I argue that we have good independent reason to accept one particular kind of eliminativism. This is what I call 'pragmatic eliminativism' which is the view that the term 'attention' should be eradicated from certain debates within philosophy and psychology.

Chapter 6 will connect my arguments to various different arguments for eliminativism in psychology which are connected to issues over natural kinds. The question of whether 'attention' is a natural kind will be examined, and will be looked at in relation to other arguments in philosophy of psychology which revolve around questions to do with natural kinds. Ultimately, I argue that there is usually *no answer* to the question of whether a certain term in psychology refers to a natural kind. I use this claim to critique eliminativist *and* anti-eliminativist arguments that are based around whether a certain term refers to a natural kind. The task of chapter 6 is also constructive. I suggest a better way of arguing in favour of eliminativism in psychology, which derives its force entirely from theoretical pragmatics, independently of ontological issues connected to natural kinds.

Chapter 7 will be a concluding chapter, where I further develop many of the ideas that the thesis has been concerned with. Specifically, I will draw upon the work of John Dupré to give a plausible explanation of *how* different concepts of 'attention' have arisen, and explain why I take pluralism about attention to be a logical progression within these debates. I will also give some more ambitious normative arguments and suggest how I think attention research *should* be carried out, if my views are taken seriously.

This thesis also contains an appendix, where I discuss in some detail a specific argument given by Ned Block (2013a) for the claim that one can be conscious of an object

without being able to direct attention towards that object. I analyse Block's reasons for making this claim, and suggest a different interpretation of the evidence.

Before we begin the main discussion of the thesis, it will be helpful to have in view some of the history of the debates that we will be turning to first, and which I will be discussing in a great deal of detail. It is to this task that I now turn.

§1.4-The debates.

Currently, there is a great deal of argument over whether attention may be necessary, sufficient or both necessary and sufficient for phenomenal consciousness in some sense. These claims will require considerable unpacking. For example, it is unclear what the intended meanings of 'consciousness' and 'attention' are. However, examining these issues will generate a great many complications that will occupy us in many of the coming chapters, and such work cannot be performed here. In the remainder of this chapter, I will purposefully leave the terms in question undefined and rely upon a pretheoretical understanding of them. This will be adequate for the purposes of this section, which is simply to give a brief and selective history of the debates that centre on these questions. The history will be anything but comprehensive, but it will give the reader a flavour of the state of the field.

§1.4.1-Is attention necessary for consciousness?

We could trace the view that attention is necessary for consciousness back as far as Aristotle, who says: 'persons do not perceive what is brought before their eyes, if they are at the same time in fright, or deep in thought, or listening to loud music' (quoted in Mack, 2002, pp.105-106). Of course, it would be anachronistic to interpret Aristotle as making a specific claim about the dependence of consciousness on attention. Nonetheless, Aristotle here

voices one of the key intuitions at the centre of this debate, which is that we often fail to notice things that are right before our eyes if we are not paying attention. As we shall see, this intuition has been used to fuel the related claim that we are not phenomenally conscious of things unless we pay attention to them. This has naturally led to the claim that attention is necessary for consciousness.

As Aristotle noted, this failure to notice things due to inattention is a common feature of everyday life. Frequently when scanning the shelves of a supermarket or a bookshelf we will fail to notice what we are looking for, even when we are actively searching for it and it is right under our noses. This is a common occurrence in other sense modalities as well. We often do not notice noises that are going on around us when we are lost in thought, or doing something else. One common example is when the fridge stops making a noise to cool itself and we realise that it has been humming for several minutes before we noticed it.[1] Equally, it is safe to say that we almost never notice the feeling of our feet in our shoes, or the shirt on our backs because we almost never pay these tactile sensations any attention. It has also been suggested that we can sometimes have headaches which go on all day, but which we do not notice because our attention is distracted.

It is quite surprising how much we can fail to notice when we are distracted. This phenomenon has a sinister side as well. One disturbing result comes from Haines (1991) who found that when trainee pilots are placed in flight simulators and told to land on a runway, under conditions of distraction they often fail to notice planes that were parked directly where they intended to land. As a result they crashed into the parked planes.

[1] I think this example originates from Block (1995a).

This phenomenon is well known from day to day life and has been since at least Aristotle's time. However, its links to consciousness and to the claim that attention is necessary for consciousness have to my knowledge only become the focus of major debate in philosophy and psychology over the last few decades. One prominent proponent of the use of these anecdotal phenomena in this context is Daniel Dennett. The famous example that Dennett uses is the 'hide the thimble' game:

> 'An ordinary thimble is shown to all participants, and all but one leave the room, while the thimble is "hidden". The rules for the hider are clear: The thimble has to be hidden *in plain sight*…. Once it is hidden, the rest of the children come back in the room and proceed to hunt for the thimble… The last few children to find the thimble can usually be counted on to *look right at* the thimble without actually *seeing* it. In these delicious moments, everyone else can see that the thimble is right in front of Betsy's nose… well lit and subtending a healthy angle in her visual field.' (1991, p.334).

From this example, Dennett draws the conclusion that the children were not conscious of the thimble until they noticed it. Dennett attributes this to the lack of attention directed towards the thimble. Dennett then uses these observations to support the claim that consciousness is actually far sparser and more 'gappy' than we may think (Dennett, 1991, ch.11).

Other prevalent examples of philosophers who link these kinds of phenomena to phenomenal consciousness include those thinkers that attempt to explain phenomenal consciousness in terms of 'Higher-order-representations'. For example, David Rosenthal argues that a mental event is conscious if and only if it is the target of a suitable kind of

higher order thought. Following the literature, we can call Rosenthal's view 'higher-order-thought' theory or HOT theory.[2]

In order to motivate the claim that mental events are not conscious unless the subject thinks about them, HOT theorists often point to a claim that we have already encountered: that mental events that are not subject to attention often go unnoticed. It is then argued that unattended events are unconscious, which lends support to the claim that some minimal kind of attention (in the form of a thought of some kind) which is directed upon the mental event in question is required in order to make the mental event conscious. This claim then leads to HOT theory. The claim that attention is necessary for consciousness is, of course, of central importance in this kind of strategy. The strategy is employed by Rosenthal, who says this:

> 'Some mental states seem to be conscious only some of the time, largely through shifts in attention; examples considered earlier are pains or auditory sensations from which we are temporarily distracted' (1997, p.745).

We can also see the claim that attention is necessary for consciousness being put to work in other 'higher-order' theories of consciousness. William Lycan argues that a mental event is conscious iff. that very event is in some sense *perceived* by the subject.[3] Lycan explains this claim by way of postulating an attentional mechanism which operates on

[2] Strictly speaking, we should refer to it as 'actualist HOT' theory, see Rosenthal (1986, 1997 and 2005). The 'actualist' criterion is added here to separate it from 'dispositionalist' HOT theories such as those of Carruthers (2000). These details will not matter for the present discussion.

[3] For work on this 'higher-order-perception' ('HOP') theory, see Lycan (1987, 1996, 2004 and 2008). See Sauret and Lycan (2014) for a more sceptical discussion of the relationship between HOP theory and attention).

internal mental events. Lycan then claims that just such an attentional perception of a mental event is necessary and sufficient for the event to be conscious. He writes:

> 'As I would put it, consciousness is the functioning of internal *attention mechanisms* directed upon lower-order psychological states and events. I would also add an explicit element of teleology: Attention mechanisms are devices which have the *job* of relaying and/or coordinating information about ongoing psychological events and processes.' (2004, pp.99-100).

Such discussions relied largely upon anecdotal evidence such as that outlined above, rather than experimental evidence taken from empirical psychology. This has recently changed, and a great deal of empirical data has come to bear on these questions. Some of the early experiments from contemporary psychology that tried to study these phenomena systematically were focussed on auditory attention (e.g. Cherry, 1953). In these 'dichotic listening' studies, subjects were given headphones and told to pay attention to the auditory stimuli that came through one earpiece, whilst 'distractor' words were played into the other earpiece. It was found that subjects failed to notice dramatic changes in the unattended auditory stream, such as changes in the language of the spoken words.

Though these results are interesting and surprising, it is only recently that empirical data have come to the forefront of *philosophical* debates concerning consciousness and attention. One main reason for this was the emergence of some particularly dramatic and counterintuitive results. Chief among these was the 'gorillas in our midst' experiment of Simons and Chabris (1999) where it was observed that over half of subjects given an attention-demanding task (counting the number of passes between basketball players) failed to notice the presence of a woman in a large gorilla outfit walking onto the scene, beating

her chest for a few seconds, and then walking off the scene. This famous experiment has been one of the focus points of philosophical debate over the interaction of attention and consciousness ever since.

Running alongside these surprising results are ones from a different paradigm, focussed upon 'change blindness'. In these experiments, images were shown to subjects, followed by a brief break, and then a new image was displayed which was identical to the first image in most ways, except for one large change. It was found that subjects are unable to notice the change (even though they were very dramatic) until they focus their attention upon the object or objects that changed. These empirical results, and many related ones, are described in Mack and Rock's 1998 book *Inattentional Blindness.* This book has been one of the loci of debate since its publication, and 'inattentional blindness' has also become a popular term for all such phenomena.[4]

In one particularly entertaining experiment, which also makes use of the now infamous gorilla, expert radiologists were given scans of patients' lungs and asked to look for any anomalies. Though the radiologists could locate tiny indications of lung cancer after only a brief glance at the image, three quarters of those radiologists given the task failed to notice a large picture of a dancing gorilla superimposed directly onto the scan of the lungs.[5] So famous are these effects that the BBC used them to develop a campaign demonstrating

[4] For just some of the extensive philosophical discussion of these data, see Noë (2002), Siewert (2002), Mack (2002), Levin (2002), Block (2008), Dretske (2007) and Tye (2009b and 2010). See Noë (2007) for a philosopher's survey of the terrain and Simons (2008) for a psychologist's view. The term 'inattentional blindness' is sometimes used to refer to such effects *outside* the visual sense modality.
[5] This was reported at: http://www.bbc.co.uk/news/health-21466529. Accessed 10/June/2013.

how easy it is to miss dramatic changes in one's environment as a way of warning drivers to be more alert to the possible presence of cyclists on the roads.[6]

These empirical results have been used by many in a similar way that Rosenthal and Dennett's used anecdotal phenomena such as the 'hide the thimble' game. That is, they are deployed in order to argue that the unattended items were not phenomenally conscious. This leads many thinkers to the claim that without attention, there is no phenomenal consciousness, and thus to the claim that attention is necessary for phenomenal consciousness. This is the view of Mack and Rock themselves.

Many thinkers have claimed that this position is deeply counterintuitive, as it apparently implies that only a very small portion of the world around us is represented in phenomenal consciousness, because only a small portion of it is attended at any one time.[7] This 'sparse' view is usually contrasted with the claim that phenomenal consciousness is 'rich' or 'abundant', which is the claim that consciousness contains many items, some of which are outside of attention.[8] This 'rich' or 'abundant' position is normally attributed to the 'folk' or to 'naïve' people.

In order to explain how 'naïve' people could be so deluded about this issue, sparse theorists have proposed that most people suffer from the famous 'refrigerator light illusion'. This involves imagining someone who believes that the light in his refrigerator is always on because whenever he looks in it, it is on. Obviously, it is the very action of opening the door that causes the light to turn on. In just the same way it is claimed that whenever we try to

[6] The video can be seen at: http://www.youtube.com/watch?v=ubNF9QNEQLA. Accessed 22/May/2014.
[7] This characterisation is a bit crude, but Dennett (1991, 2006 and Cohen and Dennett, 2011) and Kouider et al. (2010) hold something like it.
[8] See positions 3 and 5 in the list given in §1.4.

check whether a certain item is represented in consciousness, we attend to the item (or to a representation of it in our visual system). This causes us to be conscious of the item, and this fosters the mistaken belief that we were conscious of the item all along. It is this illusion that many sparse theorists use to explain why their view is apparently 'counterintuitive'.

Indeed, so prominent are these results in the literature that a whole debate has arisen focussed purely on the question of whether 'normal' or 'naïve' folk *really do* believe that we have a rich phenomenal world. Some thinkers (Noë, 2002 and 2007 and O'Regan and Noë, 2001a and 2001b) claim that normal naïve subjects *do not* really have the belief that they have a rich phenomenally conscious field, whilst Dennett (2001a and 2002) and Mole (2008) claim that normal folk *do* have this assumption. A third position in this debate is that of Schwitzgebel (2007), who argues that people's intuitions on these matters are malleable and can vary from person to person. 'Normal' people's intuitions about attention will be examined in chapter 3 of this thesis.

I think it is fair to say that most of the impetus for the claim that attention is necessary for consciousness has stemmed from the anecdotal phenomena and the experimental results mentioned above. However, there is another route to a particular version of the necessity claim which is associated with Michael Tye (2009a, 2010 and 2014). Tye argues that one condition on being able to consciously see an object is that one must be able to have *de re* thoughts about the object purely on the basis of one's experience. When Tye uses the term '*de re*' thoughts he means something akin to 'demonstrative' thoughts. That is to say: Tye claims that a necessary condition on seeing something is that one must be

able to ask of it 'what is *that*?'[9] Tye then argues that if a subject is to be able to have such demonstrative thoughts about the object, then she must be able to direct attention towards it. Therefore, if we are to consciously see an object we must at least *be able* to direct attention towards it. Interestingly, Ned Block (2014a) has not only replied to Tye's argument specifically, but engaged in a metaphilosophical critique of the entire methodology in which it is framed. As I have already said, there are a great many fascinating metaphilosophical questions raised by these debates, and this thesis will only be able to examine some of them.

This concludes our selective history of the debates over the claim that attention is necessary for consciousness. We can now turn to our other central claim, which is that attention is *sufficient* for consciousness.

§1.4.2-Is attention sufficient for consciousness?

Historically, the claim that attention is sufficient for consciousness has been assumed by many thinkers without much argument. Indeed, for some, the claim that attention is sufficient for consciousness has been *trivially* true. Consider the following famous comment from William James:

> 'Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence' (James, 1890/1981, pp.403-404).

For James, it is of the 'essence' of attention that it operates on consciousness. This seems straightforwardly to imply that attention is impossible in the absence of consciousness, and thus that attention is sufficient for consciousness. So, for James, the idea

[9] Cf. Dretske (1969 and 2007) and Siegel (2006 and 2010).

of attention in the absence of consciousness is analytically impossible. Even if we do not go so far as to claim that attention's being sufficient for consciousness is an *analytic* truth, it is frequently said that it accords with our normal folk psychological intuitions. Mole (2008) is especially clear on this point:

> 'we expect the facts about what a person is attending to to make an immediate difference to what it's like to be that person and we expect a person to be able to know what she is attending to in the immediate first-person… way that characterises facts about consciousness… The fact that we expect attention and consciousness to behave in these ways is made intelligible if we understand commonsense psychology to treat paying attention to something as a way of being conscious of that thing… According to commonsense psychology, then, attention requires consciousness' (2008, p.89).

The claim that 'attention requires consciousness' directly implies that when we have attention, we must have consciousness; and thus that attention is sufficient for consciousness. As Mole says:

> 'For all persons and all things, if the person is attending to the thing then the person is conscious of that thing' (2008, p.100).

Note that Mole thinks that the claim that attention is sufficient for consciousness is part of normal folk psychological intuition. This is interesting because whilst the claim that attention is *necessary* for consciousness has usually been accused of running counter to commonsense, the converse claim that attention is sufficient for consciousness has

sometimes been seen as having the advantage that it *accords* with our commonsense intuitions.

The claim that the sufficiency thesis runs in line with commonsense is a popular way of arguing in favour of it, but there are other ways as well. One route is by means of similar empirical data as are used to argue in favour of the necessity claim. The studies usually cited in this regard generally show that attention can make subjects aware of something that they would otherwise be unaware of (e.g. Mack and Rock, 1998, Simons and Chabris, 1999 and Treisman and Gelade, 1980. See De Brigard and Prinz, 2010; Prinz 2010, 2011 and 2012 for presentation of these and other relevant data). From these results it is argued that attention is likely to be the faculty that brings information about things to phenomenal consciousness, and thus that attention is sufficient for consciousness.

A final popular way to argue in favour of ST is to claim that there is a conceptual implication between attention and consciousness. For these thinkers, to say that we could attend to something without being conscious of it is in some sense contradictory (See Smithies (2011a and 2011b). So in this debate, we find the claim that attention is sufficient for consciousness defended on intuitive, empirical *and* conceptual grounds.

Perhaps because the claim that attention requires consciousness was often seen as intuitively plausible, it was not to my knowledge challenged in a systematic way until some empirical results that emerged at the turn of the 21st century. The most prominent challenger of the claim is the psychologist Robert Kentridge, who has argued that there are certain empirical cases where attention to a stimulus occurs in the absence of consciousness of that stimulus. Thus (argues Kentridge) attention is insufficient for consciousness. The interpretation of these results has usually been taken to dictate the fate of the claim that

attention is sufficient for consciousness. These results will be the focus of a great deal of discussion throughout the thesis, especially in chapter 2.

With this brief history in place, it will be helpful to provide a (non-exhaustive) list of the main positions in this debate. In the interests of brevity, I will skim over the subtleties of each position. Such complexities will be addressed later in the thesis:

1) Attention is necessary and sufficient for consciousness (De Brigard and Prinz, 2010 and Prinz, 2010, 2011 and 2012).
2) Attention is necessary but not sufficient for consciousness (Cohen et al. 2012, Watzl, 2014).
3) Attention is sufficient but unnecessary for consciousness (Smithies, 2011a and Mole, 2008).[10]
4) Attention is *not* sufficient for consciousness (Kentridge, 2011 Kentridge et al. 1999, 2004, 2008a and 2008b and Norman et al. 2013).
5) Attention is neither necessary nor sufficient for consciousness (Lamme, 2003, 2004, 2010a and 2010b; Koch and Tsuchiya, 2007 and Tsuchiya, Block and Koch, 2012).[11]

§1.4.3-An explanation of consciousness?

The position that attention is necessary *and* sufficient for phenomenal consciousness (position (1)) is one of the most controversial positions in logical space. One of the main driving forces behind this kind of view is the assumption that attention can be explained by the empirical sciences. A strong source of impetus for this belief has been Chalmers' division

[10] Mole gives a different view in his (2011a, ch.7).
[11] Block, Koch and Tsuchiya claim that consciousness is possible in the *near or total* absence of *top-down* attention. The distinction between 'top-down' and 'bottom-up' attention will be examined in chapter 2. Block also agrees that attention is insufficient for consciousness (2013a, p.182).

of the problem of consciousness into the 'hard' and 'easy' problems (1995, 1996 and 2010). Roughly speaking, the 'easy' problems are the problems that are tractable by the empirical sciences, whilst the 'hard' problems of consciousness are those of explaining why phenomenal consciousness should arise from the workings of a brain, and why it feels a certain way. These 'hard' problems are (according to Chalmers) unsolvable by the domains of psychology and neuroscience.

Many have disputed Chalmers' taxonomy.[12] However, what we should emphasise for present purposes is that Chalmers includes in his list of easy problems the ability to direct attention (1995, p.200). The inclusion of attention in the list of easy problems which are in principle solvable by the empirical sciences has fuelled interest in investigating consciousness in terms of attention, because it may be hoped that in explaining attention the empirical sciences may be able to explain phenomenal consciousness after all. Obviously, the prospects for this project are significantly improved if attention and consciousness co-occur, which is one key reason that the claim that attention is necessary and sufficient for consciousness has been so attractive.

Excitement over the prospect of using attention to explain consciousness is certainly not restricted to the philosophical literature, we also find it in the empirical literature. Consider the following quotation from the psychologist Michael Posner in an article strikingly entitled 'attention: the mechanisms of consciousness':

> 'I propose to discuss the issue of consciousness in light of recent findings about attentional networks of the human brain… I don't believe that any of these

[12] Lowe (1995) argues that all of the problems are hard and Dennett (1996) argues that all of the problems are easy.

> mechanisms are "consciousness" itself, just as DNA is not "life", but I do believe that an understanding of consciousness must rest on an appreciation of the brain networks that subserve attention' (1994, p.7398).

Posner's view has been recently counteracted within empirical psychology by the publication of Koch and Tsuchiya's provocatively titled 'attention and consciousness: two distinct brain processes' (2007). In this now classic review article, Koch and Tsuchiya survey a range of empirical data and attempt to argue that attention is neither necessary nor sufficient for consciousness thus standing in direct opposition to Posner's optimistic view. Koch and Tsuchiya do not stand alone in their scepticism, another prominent exponent of the view that attention and consciousness dissociate is the neuropsychologist Victor Lamme (e.g. 2010a). These 'dissociationist' views are represented as position 5 in the list of positions in §1.4.

Indeed, there is still a great deal of hope within cognitive science that an investigation of attention will help us to understand consciousness better. For example, in his (2012) work Posner says 'I still believe that much can be learned about consciousness from an understanding of attention' (p.1). The prospect of explaining consciousness in terms of attention is a very interesting one, and it is this prospect which much of this thesis will examine.

This concludes our brief history of these debates. In order to get into the details of assessing these claims, it will be helpful to turn to some of the empirical data that have been brought to bear on these questions. I will begin this in the next chapter.

Chapter 2
Attention and Consciousness

§2.0-Summary.

In this chapter, the debates over whether attention is necessary and sufficient for consciousness will be examined. It will be argued that these debates have suffered because of insufficient appreciation of the conceptual issues which affect how we interpret certain empirical results. Specifically, it will be argued that the major point of friction between different thinkers in this debate is disagreement over the meaning of the word 'attention', and not over some fact that can be revealed my experimental psychology or neuroscience. If the main argument of this chapter is correct, then much of the work currently carried out on the relationship of attention to consciousness is misguided and must be reassessed.

§2.1-Conceptual preliminaries.

In the previous chapter I gave a selective history of the debates surrounding the question of whether attention is necessary and/or sufficient for phenomenal consciousness. As I pointed out, these claims are somewhat unclear and could do with more precise formulations. I suggest we understand the central claims in this way:

i) *The necessity thesis (NT):*

Attention to some item(s) is necessary for a representation of that item (or those items) to be phenomenally conscious.

ii) *The sufficiency thesis (ST):*

Attention to some item(s) is sufficient for a representation of that item (or those items) to be phenomenally conscious.

I take phenomenal consciousness to be a term that applies to all and only those mental events that there is 'something it is like' to undergo. This understanding is in accordance with Nagel's characterisation (1974). Notice that this characterisation implies that only mental *events* are phenomenally conscious. I take this to be plausible (cf. Crane,

2001, §§31-32 and Soteriou, 2007) but readers who disagree with this claim can substitute another term if they wish. My characterisations of NT and ST are designed to accurately represent the views that are at issue in the debates in question. Some thinkers may wish to change the characterisations slightly to cohere with other beliefs about consciousness. For example, John Campbell (2002) is amongst those who would take issue with the claim that phenomenal consciousness 'represents' items in any important sense. Those who hold this view can replace the definitions as they see fit. Nonetheless, the characterisations I have given will serve for my purposes.

ST and NT allow for the possibility that the attention in question may be inward-directed at our own mental states and events or outward-directed at the world.[13] ST and NT also leave open the possibility that the item that the subject is attending to may be *self-representing*. That is to say, it may be the case that sometimes (for example, when we pay attention to perceptual events in our sensory system) those events represent *themselves* in a phenomenally conscious manner.[14]

An important issue that arises at this point concerns how we interpret the modal strength of 'necessary' and 'sufficient'. For example, it could be that attention is *nomologically* necessary or sufficient for representations of certain items to become phenomenally conscious. The claim that attention is necessary and sufficient for phenomenal consciousness would hold with nomological necessity if attention is always necessary and sufficient for consciousness *given the laws of nature that hold in the actual world*. Nomological necessity can

[13] As we saw in chapter 1, adherents of the HOT and HOP accounts would likely insist that attention *to one's own mental representations specifically* is necessary and sufficient for those representations to be conscious.

[14] Such a view would fit neatly into Uriah Kriegel's 'self representational' theory of consciousness (2009).

be contrasted with *metaphysical* necessity. Truths that hold with *metaphysical* necessity must always hold regardless of changes in laws of nature across different possible worlds. As I mentioned above, the primary focus of this chapter is a certain set of empirical results that have been thought to falsify the claim that attention is necessary and sufficient for consciousness. Since they are empirical results, they primarily bear upon how attention and consciousness interact *in the actual world.* For this reason, I suggest the claims in question be read as claims of *nomological* necessity and sufficiency.

This is not to say that such empirical results have *no* bearing on the claims when they are read in the stronger, metaphysical sense. If attention and consciousness can be dissociated in the actual world then this would also disprove the claim that attention is *metaphysically* necessary and sufficient for consciousness. This is because such a view would hold that attention is necessary and sufficient for consciousness in *all* possible worlds (including the actual one). I should mention that it is controversial whether nomological and metaphysical modality are really distinct (Shoemaker, 1998). This does not pose a problem for the discussion that I will be having because if it is the case that metaphysical and nomological modality are the same then the claims in question can be read either way without difference.

§2.2-Evidence against the sufficiency claim.

I turn first to ST. In this section I will examine some of the arguments *against* the claim that attention is sufficient for consciousness. In §2.3 I will consider an argument *in favour* of ST. In §2.4 I will delve more deeply into some other empirical evidence relevant to ST. I move on to NT in §2.5. Further issues related to these debates will be raised and discussed in §§2.6-2.8. My discussion will require quite an in depth examination of the

particular experiments that have been at the centre of these debates. This examination will sometimes become slightly complicated, but such complication is needed in order to bring to light many of the issues that are my main concern in this chapter, and which I will be discussing in much of the rest of the thesis.

§2.2.1-Blindsight and the sufficiency thesis.

In order to empirically establish that ST is *false* we need an example of a subject attending to an item, without having a phenomenally conscious representation of that item. There are data that have been taken by many to show this. We should begin by getting this information in focus. The results stem from experiments on the neurological patient whose pseudonym is GY. GY is an English adult male, who was 41 years old at the time that the experiments in question were carried out. At the age of eight, he was involved in a car accident in which he suffered severe damage to his left striate cortex (an area of the visual cortex at the back of the brain), which left him with the neurological condition of blindsight (Kentridge et al. 1999, p.1805 and Kentridge, 2011 pp.238-239). Blindsight subjects deny awareness of items which are placed within a certain area of the visual field. This is the area to which the subject is 'blind' and I shall (following convention) refer to it as the subject's 'blind field'. Some blindsight subjects lose sight in some section of the visual field, and some lose sight in their entire visual field. GY is one of the former.

What makes blindsight distinct from normal blindness is that blindsight subjects are capable of accurately detecting visual stimuli and making simple discriminations about the properties of entities that fall within their blind field. This is the case even though they continually deny awareness of the stimuli in question. The detection and discrimination abilities that blindsight subjects have with relation to stimuli in their blind fields fall well short of the abilities that non-neurologically impaired subjects have in relation to items that

fall within their visual fields, but it is controversial exactly *what* the differences are between the abilities of blindsight subjects and non-neurologically impaired subjects. Typically, blindsight subjects have to be prompted to respond to stimuli that fall within their blind field, but it is controversial whether blindsight subjects *always* require such prompting (Stoerig, 2010). The question of whether blindsight subjects always require prompting is interesting but it would carry us too far afield and I shall discuss it no further.

During experimental work on GY which was not specifically to do with attention and consciousness, GY mentioned that he had been trying to pay attention to things that fell within his blind field.[15] Such a strange remark prompted the experimenters to test whether GY really was capable of paying attention to items placed in his blind field (Kentridge et al. 1999). In order to do this, they used the Posner paradigm (figure 1).

[15]The original experiment was Kentridge et al. (1997).

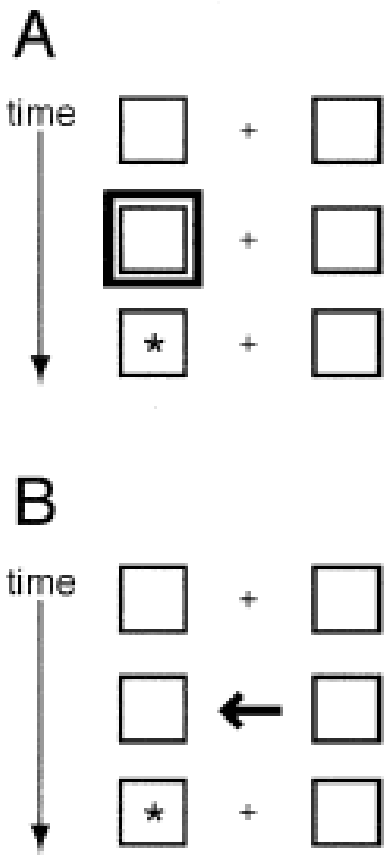

Figure 1: The Posner paradigm. The procedure labelled 'A' in the above diagram is the exogenous or 'direct' version. The procedure below (labelled 'B') is the endogenous or 'indirect' version of the paradigm. From Vecera, S. P. and Rizzo, M. (2003) "Spatial attention: normal processes and their breakdown." ***Neurologic Clinics*** **21, pp.575-607. Reproduced with permission from Elsevier.**

The Posner paradigm (e.g. Posner, 1980) is one of several procedures used to test for attention in empirical psychology. It is easiest to get a grip on how it works if we look at the procedure in figure 1(b). Subjects are given a fixation point which they are instructed to maintain their gaze upon (this is the cross in between the two boxes at the top of figure 1(b)).[16] Shortly after the subject maintains fixation, an arrow is presented at the fixation point, pointing to somewhere in the subject's visual field (this arrow is the 'cue'). There is then a brief pause after the cue (the arrow) is presented and before the target is presented. In figure

[16] Whether subjects obey the instruction to fixate on the cross can be verified using an eye scanner which tracks the movement of the pupil of the eye.

1(b), the target is the star in the bottom left box. The pause between the cue and the target is referred to as the cue target onset asynchrony. The target is the stimulus to which the subject is instructed to respond. In the first experiment, GY was simply instructed to press a button if the target had been presented at all, but sometimes the task can be more complex than that (later I will be looking at more complex experiments).

An important distinction here is between endogenous and exogenous attention. Endogenous attention is attention that is intentionally directed by the subject toward elements in the environment (endogenous attention is often called 'bottom-up' attention). The version of the Posner paradigm depicted in figure 1(b) is usually thought to track endogenous attention because the subject must see the arrow and then understand the meaning of the arrow and intentionally direct attention to where the arrow points. Because the cue (the arrow) is presented at a location other than the place where the target stimulus is subsequently presented, it is said to be 'indirect'.

The version depicted in figure 1(a) is similar in most respects, but this is the exogenous or 'direct' version of the paradigm. Exogenous (or involuntary) attention is attention which is grabbed beyond our control, rather than being intentionally directed by the subject. In this version of the paradigm, a subject is asked to fixate their gaze upon the fixation point once again, but here the cue (in this case the black box that appears on the left, rather than an arrow) appears at the location that the target stimulus (again, the star symbol) is expected to appear. So, in this version the cue appears at the same location as the target which is why it is referred to as a 'direct' cue. Because in this version the cue is direct, and does not rely upon the subject having an understanding of the meaning of the symbol, it is thought to test for exogenous attention.

Both the exogenous and endogenous versions of the paradigm depicted in figure 1 contain *valid* cues. Valid cues are cues which correctly indicate where the target will appear. So, in the version depicted in figure 1(a), this is a valid exogenous cue because the cue appears around where the target does indeed appear. In the endogenous condition depicted in figure 1(b), the cue is valid because it correctly points towards where the target will appear. Cues can also be invalid (or 'misleading'). A cue could be invalid if it incorrectly indicates where the target will appear. For example, in the version depicted in figure 1(b), if the target appeared on the right, then the cue would be invalid (because it points to the left). Cues can also be invalid if no target stimulus is present at all. These are called 'target-absent' trials, for obvious reasons. The time it takes subjects to respond to the target stimuli (the 'reaction time') is typically shorter in trials where the cues are valid, rather than invalid or where the target is absent entirely.

The experiment in question was carried out on GY using both endogenous and exogenous versions of the Posner paradigm. GY was given a fixation point, followed by a cue of the endogenous or exogenous kind.[17] Following this, the target was either presented where the cue had indicated (in the trials where the cues were valid), or it was presented at a place other than where the cue had indicated (an invalid trial) or no target was presented at all (the target-absent trials). Whenever the target was presented, it was presented in GY's blind field. This obviously had to be the case, because the point of the experiment was to ascertain whether GY could pay attention to something that fell within his blind field. In the trials that made use of the exogenous cues, *both the cue and the target* fell within his blind field.

[17] In the experiment in question, the exogenous cue was not a box, as depicted in figure 1, but rather a pair of horizontal bars that appeared around where the target was to be presented (Kentridge et al. 1999, p.1806).

So, in these trials GY was aware of *neither* the cue nor the target. As we shall see in §2.2.2, this is an important detail.

After the cue had been presented, an auditory tone was sounded. GY was instructed to press a button if he felt that a target had appeared in his blind field when the tone was sounded. The target in question was a small disc (Kentridge et al. 1999, p.1806). It was found that GY's reaction time was significantly improved when the cue had been valid, in both the exogenous and endogenous versions of the experiment. This increase in reaction time was *not* accompanied by a reduction in the accuracy of GY's guesses (Kentridge et al. 1999, p.1809). It is interesting to note that the fact that GY had increased reaction times in the *exogenous* versions of the paradigm ran counter to the expectations of the experimenters themselves (Kentridge et al. 1999, p.1810).

These experiments were followed by an interview with GY, where he was asked if he had had any awareness at all of the target stimuli that he had been looking for. The expectation was that he would deny any awareness because they fell within his blind field. This expectation was correct:

> 'Although GY made trial-by-trial commentary responses throughout the experiment, he did not report awareness of a single target.' (Kentridge et al. 1999, p.1808).

The experimenters used the results of this interview to draw the conclusion that GY was not conscious of the targets. Because GY's task performance was increased when the cues were valid, the experimenters drew the conclusion that the cues were facilitating attention to the targets. This led to the conclusion that the targets were unconscious but attended to, and this is what is required to disprove ST. The experimenters concluded that:

'attention could selectively modulate the processing of a target without that target's entering awareness' (Kentridge, 2011, p.240).

§2.2.2-Rival views on the data

These experiments have caused a great deal of debate and controversy, and it is far from universally accepted that they demonstrate attention in the absence of consciousness. In this section, I will consider four objections to the interpretation of the data on which they demonstrate attention without consciousness. Before I begin, I should say that of course there are myriad background methodological issues in the literature about experiments of this sort and there is no chance of me being able to thoroughly address all such issues here. Nonetheless, I hope to highlight several rival views and show why I am doubtful of them. After discussing these issues, I will raise some different issues in §2.2.3, which have not yet received the focus that they deserve. It is these latter problems that will occupy me mainly in this thesis.

§2.2.2.1-Space?

The first point to address is that there has been some debate about *what* exactly GY is attending to (if anything). Mole (2008) suggests that the GY results may only show the presence of attention to the *region of space* that the target was in rather than attention to the target itself. Mole argues that if this were the case then there would be nothing that is both attended to and unconscious, because GY is attending to the *space* and he is conscious of the space. He is not conscious of the *target* but that is okay, because he is not (strictly speaking) attending to it.

This issue is addressed at length by Kentridge et al. (2008a). Above I mentioned that in some versions of the Posner paradigm, the task is more complicated than simply guessing whether the target is present or not. One such experiment was Kentridge et al. (2004), which

again had GY as a subject. This experiment was also a version of the Posner paradigm. Cues (exogenous and endogenous) were presented, 80% of which were valid, and 20% of which were invalid (Kentridge et al. 2004, p.833). The target stimulus was in all cases a black bar, though in some cases it was horizontal, and sometimes vertical. GY was instructed to press a left button if the bar was horizontal, and a right button if it was vertical. In this experiment again, increased reaction speed was recorded in the trials that had valid cues. An additional point is that GY's ability to discriminate whether the bar was horizontal or vertical was above chance in *both* the valid and invalid cue condition, but his performance was *increased* in the valid condition (Kentridge et al. 2004, p.834). This is an important point: in this experiment, GY not only performed *faster* in the trials that had valid cues, but he was also slightly *more accurate.*

In these experiments (as in the previous ones) GY was asked if he was aware of the targets and he denied that he was (Kentridge et al. 2004, p.833). Interestingly, GY was so convinced that he was not seeing any targets that he at one point accused the experimenters of trying to deceive him:

> '[GY] steadfastly denied any knowledge of targets and, indeed, suggested at one point that there were no targets and that we were running some control condition' (Kentridge, 2011, p.239, cf. Kentridge et al. 2004, p.834).[18]

I find these experiments a convincing reply to the claim that GY could have been attending only to *space* rather than to the objects themselves. The main reason is this: consider that in order to succeed at these experiments (which GY was doing), GY would have had to be processing information about the *properties* of the target stimuli. So, in order

[18] Bob Kentridge has told me in personal communication that GY at one point spontaneously insisted that he could not see any of the targets, and said 'there's no point carrying on with this, I can't do it'.

to have increased performance in the valid conditions, the cue must have been facilitating processing of the *properties* of the targets. This strikes me as good evidence that (if GY was attending to anything at all) it was likely the *properties* of the targets (or the targets themselves) that he was attending to, and not simply the space that the targets occupied. Indeed, I must admit that Mole's suggestion that GY could be attending to *the space* that the target is in but not the target itself has always puzzled me: in order to succeed, GY needs to be processing information about the *target* better in the valid as opposed to invalid conditions. So (if we are willing to admit that GY was 'paying attention' at all in the valid condition) then why not accept that he was paying attention to the targets?[19]

These experiments constitute a (to my mind plausible) reply to the suggestion that GY could have been attending only to the *space* around the target rather than the target itself. There is also converging empirical evidence supporting the view that objects that are displayed but invisible to subjects can attract attention (esp. Norman, Heywood and Kentridge, 2013). Roughly, this line of empirical evidence involves an experiment in which a target stimulus (a green disc) is preceded by the presentation of a cue (a white disc). Sometimes the target appears within the same object as the cue, and sometimes the cue and the target fall in different objects. The crucial result is that subjects are faster to discriminate the presence of the target when it falls within the boundaries of the same object that the cue fell within.[20] Crucially, subjects are faster at detecting a target if it falls within the same

[19] Kentridge et al. also include further arguments against Mole's 'space' interpretation. One is based upon the claim that it is evolutionarily advantageous for attention to be sensitive to the *properties* of stimuli rather than merely to their location (2008a, p.110). The second draws on results from Remington and Folk (2001) to argue that attention preferentially selects task relevant properties of objects, it does not process information about all properties of objects indiscriminately (Kentridge et al, 2008a, pp.108-110). Plausible as these arguments are, I will not discuss them here.

[20] Egly et al. (1994) discovered this effect using objects which were visible to the subjects. The effect can also be observed with partially occluded objects (Moore et al. 1998).

object as the cue than they are at detecting a target which is *the same* distance from the cue, but which appears *outside* of the object that the cue appeared in. This occurs even when the object whose boundaries the cue and the target fell within (in this experiment, a rectangle) is invisible to the subjects (they were unable to discriminate the presence of the object when given a forced choice). This has been taken as good evidence that the cue was attracting attention to the *object* in which the target fell even though the object was indiscriminable to the subjects.

Taking all of these data together, I think they constitute convincing evidence against the 'space' interpretation. For what it is worth, Chris Mole has himself recently changed his mind on this issue due to this empirical evidence (forthcoming a). We could pursue this issue much more, but I will not do so here. This is partially because of space constraints, and partially because I think that there are other important issues which are much less appreciated. If readers are unconvinced by what I have said about Mole's interpretation then much of what I have to say about these experiments will still apply. I shall not discuss the space interpretation any more here.

§2.2.2.2-Target or Cue?

The second important question to ask is that *if* GY was attending at all, then was he paying attention to the *target* (the stimuli that he was instructed to respond to) or the *cue* itself (the arrow that pointed toward the target stimuli)? It may be said that if GY was only paying attention to the cue that was presented in his healthy field, and of which he was conscious then this is not an example of GY paying attention to something that he was not conscious of, so the experiments are not a counterexample to ST.[21]

[21] Thanks to Mette Hansen for raising this.

This issue can be neatly addressed by focussing on the exogenous versions of the Posner paradigm described in §2.2.1. Recall that in the exogenous versions, the cues were presented at the location at which the target would subsequently appear (the cues in this version were horizontal bars that appeared around the target). When GY was questioned about whether he saw the cues in this version, he denied seeing them. As Kentridge et al. say:

> '[GY] was also asked after each block of trials whether he had any experience of the cues; his responses indicated that he did not. For example: 'I would be none the wiser if you were not putting any cues up just to confuse me' and 'I just listen for the beep and press a button.'' [22]

In this version, GY denied awareness of *both* the cue and the target and he still showed increased performance. The question of whether GY was attending to the cue or the target does not arise here, because *both* of them fell within his blind-field. So, no matter which one was attended to this will still show attention in the absence of consciousness and so either one will be a counter-example to ST.

§2.2.2.3-Could GY have seen just a little?

The third question to ask is whether GY might have had some minimal level of phenomenal consciousness of the stimuli in question. Morten Overgaard (2006 and Overgaard et al. 2006) has stressed the need for psychologists to be sensitive to different possible levels of conscious awareness that may obtain in blindsight. Overgaard claims that there is good reason to think that many supposed blindsight patients *are* in fact consciously aware of the stimuli in question, if only peripherally or vaguely. Overgaard's evidence for

[22] The experimenters did run a control condition where they *intentionally* made GY aware of the exogenous cue (Kentridge et al. 1999, p.1809).

this claim is that, upon more careful and sensitive questioning of certain blindsight subjects, they have sometimes said things which imply that they *might* have had some kind of phenomenal representation of the stimuli that they are supposedly blind to. Occasionally, blindsight subjects will report feeling 'visual pin-pricks' or 'dark shadows' or 'white halos' (Overgaard, 2011, p.476). Needless to say, Overgaard's view is extremely controversial.

Could an opponent of Kentridge's view pick up on some of Overgaard's work here and claim that GY had a vague or peripheral conscious awareness of the target stimuli? I certainly cannot engage in a thorough philosophical analysis of the methodology of blindsight study in this thesis, so I cannot give these issues the dedicated discussion that they really deserve. However, I think that in these particular cases, this interpretation seems unlikely. Recall the resoluteness and certainty with which GY denied seeing the targets and the cues that fell within his blind field (he accused the experimenters of running a control condition). Importantly, when the experimenters interviewed GY they were not only probing whether GY had any *visual* awareness of the target. They also asked GY to tell them if he had *any* awareness of the target, including a 'feeling that something was there' (Kentridge et al. 1999, p.1806 and 2004, p.833). Even under these liberal instructions, GY still flatly denied having any awareness of the targets.

Since the *only* clear evidence that blindsight subjects may have some conscious awareness of stimuli presented in their supposedly blind fields is that they sometimes report some things which indicate they may do, we can take GY's steadfast denial of any knowledge of the stimuli in his blind field (and indeed, his belief that there weren't really any stimuli there) as good evidence that GY did *not* have any conscious representation of the stimuli.

Someone might object that in spite of this it could be the case that GY *did* have some phenomenally conscious representation of the stimuli, but that it was one he was unaware of, and one he denied having. The idea that there could be phenomenal experiences that the subjects that have them do not know about and *cannot* know about raises a host of complex issues (see Block (2001/2007, 2008), Cohen and Dennett (2011), Levine (2007), Papineau (2002, ch.7) and Taylor (2013d)). However, whether we are open to this suggestion or not, I think it is reasonable to claim that in the absence of *positive reason* to believe that GY was having phenomenal experiences that were inaccessible to him, then to insist that he *was* having such experiences seems to me to be unmotivated scepticism rather than a serious hypothesis.

§2.2.2.4-We should not trust blindsight!

These questions over the nature of vision in blindsight subjects feed into a much wider worry, which is that some thinkers have more general suspicions about the use of blindsight cases as a basis for drawing conclusions about consciousness in general. We can raise such worries by pointing out that subjects such as GY are brain damaged, and thus that (due to the phenomenon of neural plasticity) their brains are likely to have undergone radical rewiring. As a result GY's brain may not function the way that non-neurologically impaired subjects' brains do. An opponent might say that we should not extrapolate what GY can do to the rest of the population and as such we should not draw the conclusion that ST is false from a set of studies on a blindsight subject.

My view on this matter is that when it comes to asking whether attention is sufficient for consciousness, we do not need to be so trepidatious with the use of blindsight studies. Recall that the central claims are supposed to be ones which hold with (at least) nomological possibility. Those who accept ST and/or NT typically claim (at least) that attention is sufficient (and/or necessary) for consciousness given the laws of nature that obtain in the

actual world. For this reason, any one counterexample to ST which occurs in the actual world is sufficient to disprove ST, when interpreted as a claim about nomological possibility. This will be true whether or not the counterexample originates from a blindsight sufferer. It may be that GY's brain is unique in various ways (it undoubtedly is) but the mere presence of attention without consciousness in GY is sufficient to disprove ST by itself. After all, the point of the experiments was not to show that attention in the absence of consciousness is possible in non-neurologically impaired subjects; the point was simply to disprove the claim that attention is always sufficient for consciousness. When restricted to being about *this* claim, we need not concern ourselves with whether extrapolation to the general population is advisable.

So, I am not convinced by the worry that blindsight subjects' brains are so peculiar that we should not trust research performed on them. However, if readers are worried by these concerns then we can change the focus of our discussion away from the experiments on GY to studies which were performed on non-neurologically impaired subjects. These experiments form part of the core set of experiments that have been taken to be problematic for ST, and much of what I have to say about the GY experiments will apply to these experiments also. For these reasons, we should turn to examining these experiments now.

§2.2.3-The meta-contrast masking experiments.

The point of the experiments that I will now turn to was to demonstrate attention to a stimulus without a conscious representation of that stimulus in *non-neurologically* impaired subjects. Obviously these subjects do not suffer from blindsight so an alternative way of presenting stimuli unconsciously was required. The paradigm chosen was a meta-contrast masking paradigm (figure 2).

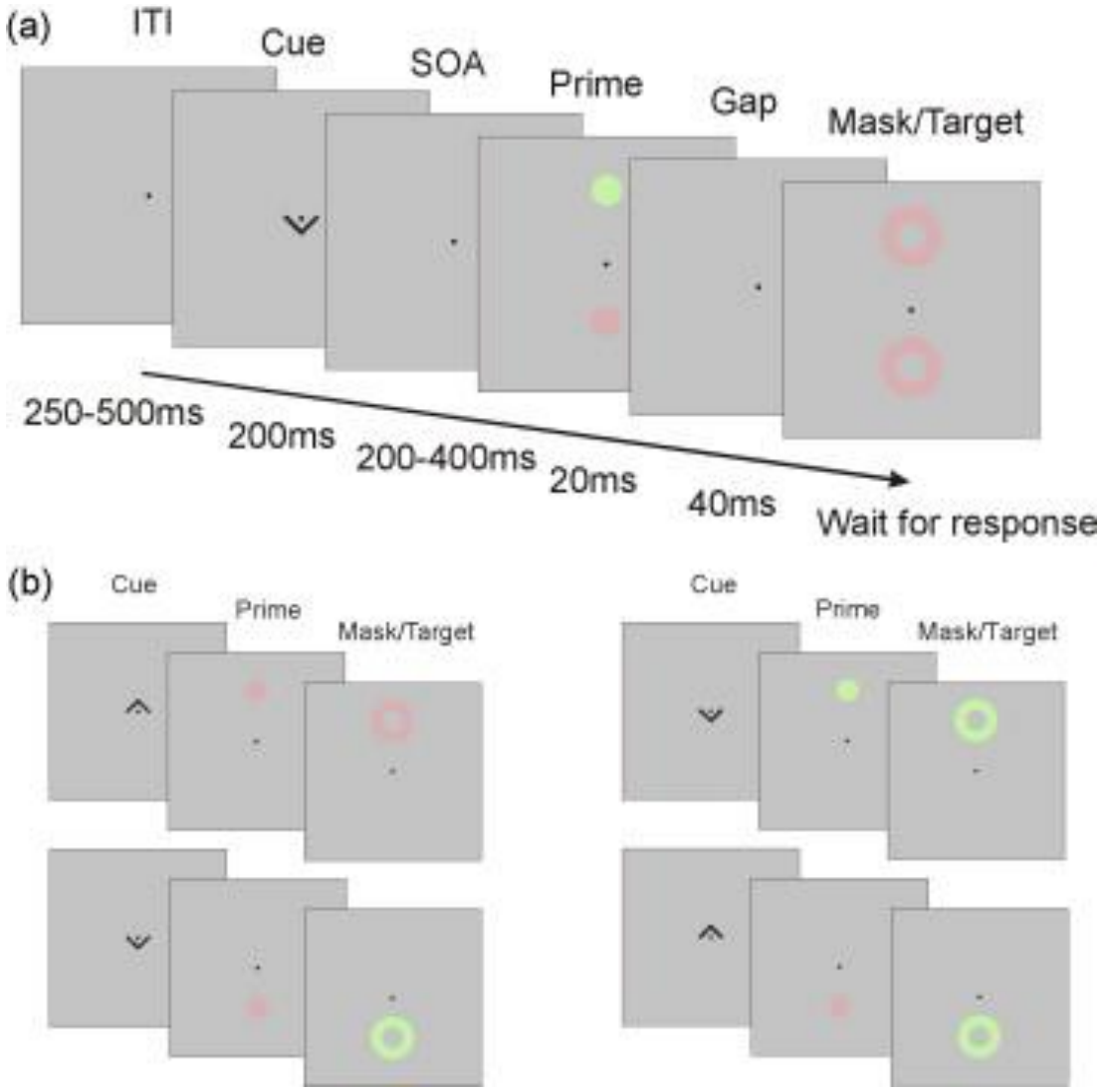

Figure 2: The meta-contrast masking paradigm used to test for attention without consciousness in non-neurologically impaired subjects. (a) depicts the dual-target trials, which are the most important for present purposes. The four diagrams in (b) depict different versions of the single-target trials. Notice that this is another modification of the Posner paradigm discussed above. In these diagrams, 'SOA' simply means 'stimulus onset asynchrony', which is just the interval between presentation of stimuli. From Kentridge, R., Nijboer, T. and Heywood, C. (2008b) "Attended but unseen: visual attention is not sufficient for visual awareness." ***Neuropsychologia.*** **46, pp.864-869. Reproduced with permission.**

Meta-contrast masking is a process where certain stimuli can be rendered invisible by the presentation of subsequent stimuli at a location around the location where the initial stimulus appeared. We can look at the top left diagram in figure 2(b) in order to demonstrate how meta-contrast masking works. In the diagrams in figure 2(b), time is going left to right (as is explicit in figure 2(a)). As in the GY studies, a fixation point is presented

(the dot in the centre of the display) and again success at maintaining this fixation point is monitored by the use of an eye scanner. An initial stimulus is presented (in the case of the diagram in the top left of figure 2(b), this is the small red disc above the fixation point). This stimulus disappears after a very brief exposure time. There is then a very brief pause. After the pause, a second stimulus is presented. In the diagram in the top left of figure 2(b), this is the red ring above the fixation point. The second stimulus (the ring) has inner boundaries that perfectly coincide with the outer boundary of the initial stimulus (the disc). The result is that signals from the second stimulus coincide with signals from the first stimulus in early stages of processing in the visual cortex, and the first stimulus (the disc) is not consciously perceived, though the second stimulus (the ring) is. For this reason, the second stimulus (the ring) is referred to as the 'mask' (it is also called the 'annulus'), and the first stimulus (the disc) is referred to as the 'masked' stimulus.

Though the masked stimulus is invisible to the subjects (because it is masked by the second stimulus), it is known that it can induce priming effects in subjects.[23] Specifically, it is known that if the masked stimulus and the mask are the same colour, subjects will be faster at discriminating the colour of the mask stimulus (Breitmeyer et al. 2004). When the masked stimulus and the masking stimulus are the same colour, they are said to be 'congruent'; when they differ in colour they are incongruent. A congruent version can be seen in the top left diagram in figure 2(b), and an incongruent version can be seen beneath it.

The experiment was designed to test if the priming effect brought on by presentation of masked stimuli that are the same colour as the masking stimuli could be exaggerated by the direction of attention. The subjects were four non-neurologically

[23] Priming occurs when invisible stimuli affect subjects' abilities in certain ways when the subject is unaware of this influence.

impaired participants (three female and one male) between the ages of 18 and 25 (Kentridge et al. 2008b, p.865). The participants were not told what the purpose of the experiment was and by the end of the experiment they *incorrectly* guessed what it might have been (Kentridge et al. 2008b, p.866). The specific version of the experiment that is most important for our discussion can be seen in figure 2(a). This is referred to as the dual-target trial, because there were two masking rings presented (Kentridge et al. 2008b, p.865). In this experiment, a cue (the arrow) appeared at the subjects' fixation point. Then two masked stimuli (the discs) appeared, one at the point at which the cue had pointed and one at another location. These two discs were of a different colour from each other (as can be seen in figure 2(a), one would be green, one red). Of course the subjects were unaware of these discs because they were masked. As can be seen in figure 2(a), the two discs remained on the display for only 20ms and then disappeared. This was followed by a pause of 40ms, and then two masking rings appeared. Recall that the masking rings are perfectly visible to the subjects. Importantly, *the two masking rings were always of the same colour*, either both of them were green or both red.

Subjects were asked to report on the colour of the rings as soon as they could (the error rate for reporting on the rings' colour was less than 1.6% on average across all of the trials (Kentridge et al. 2008b, p.866)). In some versions of the experiment (the 'congruent' ones), the masked disc that appeared at the location that the cue had pointed towards was the same colour as the rings that appeared subsequently (this is the version depicted in figure 2(a)). In other versions, the disc that the cue had pointed towards and the rings subsequently presented differed in colour (these were the 'incongruent' trials). The reaction times for subjects to correctly ascertain the colour of the rings was measured. The crucial result is that

subjects' reaction times were significantly lower (and thus their performance was *better*) when the cue pointed towards the disc that was of the same colour as the subsequently presented rings.

This may all seem a little detached from the issues we started with (whether attention is sufficient for consciousness). However, we can reconnect with our central questions once we get a grasp on the logic behind the experiment. The idea is this: in all of the trials, both a red masked disc and a green masked disc appeared on the display. So, at least one of them must be the same colour as the rings subsequently presented (because the rings were always either both green or both red). The difference is just that in some experiments the cue pointed towards the disc that *was* of the same colour as the ring, and in others it pointed towards the disc that *wasn't* the same colour as the ring. This means that *if* there is a drop in reaction time when the cue points toward the disc that *is* of the same colour when compared to the trials where the cue points toward a disc that is of a different colour from the rings, that must be because the cue is affecting the processing of colour properties *of the disc that it is pointing towards*, over the one that the cue is *not* pointing towards.

To put this another way: if the direction that the cue was pointing had *no effect* on the processing of the colour properties of the disc that appeared at the location that the cue pointed towards, then processing of the colour properties of both discs should be equal. But if they are both equal, then we would not expect a difference in reaction time when the cue was pointing *towards* the disc that had the same colour as the ring, as opposed to *away* from it. If the cue has no effect on processing of the disc, then surely we would expect the reaction time to be the same no matter whether the cue was pointing toward the disc of the correct colour, or toward the disc of the incorrect colour. Notice that in this experiment, the cue

must have been preferentially affecting processing of *properties* of the masked discs, rather than merely detecting for presence or absence of them or something like that.

The experimenters took this result to indicate that attention had been paid to the masked disc that the cue pointed towards, over the one that it did not point towards. However, in order to disprove ST we need to establish that the masked disc did not enter consciousness. One method used was an interview.[24] What the subjects said will be important later, so it is worth quoting this information in full. In the following quotation, the word 'prime' refers to the masked disc and the term 'annuli' refers to the rings:

> 'Following completion of testing subjects were systematically and individually debriefed. First, each was asked to describe everything they had seen on the display during the experiment. None mentioned the prime… They were then asked whether they might have seen anything else in addition to the fixation crosses, cues and rings (targets) that they had just described. Again, all four subjects maintained that they had seen nothing else displayed. They were then asked directly whether they had seen any coloured discs at the locations of the annuli centres just prior to the appearance of the annuli and again they denied seeing the primes. Finally, they were shown examples of the stimulus sequences slowed down by a factor of 10 so that the primes were clearly visible. The subjects registered astonishment that such primes had been present throughout the thousands of trials they had just completed' (Kentridge et al. 2008b, p.866).

[24] The experimenters also performed another experiment where they showed subjects the same experiments as they had just performed, but in some of which the masked discs were present, and in some of which they were not. Subjects were forced to guess whether they had been there or not, and they performed no better than chance (Kentridge et al. 2008b, pp.867-868).

The experimenters took such resolute denial of any awareness of the masked discs as good evidence that the discs had not been consciously perceived by the subjects. The claim that the masked discs were attended to but not conscious leads the experimenters naturally to the following conclusion:

> 'in normal observers, just as in GY, spatial attention can selectively facilitate the processing of unseen stimuli without those stimuli eliciting awareness. Attention cannot be a sufficient precondition for awareness.' (Kentridge et al. 2008b, p.866).

It has been helpful to explain the meta-contrast masking studies in some detail, for several reasons. Firstly, much of what I have to say about the GY studies also applies to them, and it will be instructive to see how the issues to do with GY carry over to the other experiments that have been done in this area. Secondly, as I have already said some thinkers are sceptical of studies involving blindsight patients, as they worry about extrapolating results about blindsight patients to the general population. I do not think we need to worry so much about this but if one does have these concerns, then one can simply take the meta-contrast masking studies as the main case study rather than the GY studies. We need not only look at blindsight if we do not want to.

§2.2.4-De Brigard and Prinz vs Kentridge.

Felipe De Brigard and Jesse Prinz occupy a particularly striking position in this debate. They say this: '[w]e claim that attention is necessary and sufficient for perceptual representations to become conscious' (2010, p.51).[25] De Brigard and Prinz are keen to show that the GY and meta-contrast masking results are not instances of attention in the absence of consciousness. The key reason they give for thinking that the results do not damage ST is

[25] Interestingly, De Brigard argues in his (2012) that attention is insufficient for conscious *recollection of memories*, so for De Brigard, it seems likely that ST will only apply to *perceptual* experiences.

that GY and the subjects in the other studies are not really *attending* to the items in question (De Brigard and Prinz, 2010 and Prinz, 2011, esp.pp.195ff.). De Brigard and Prinz claim that GY is not really attending to the stimuli in question, whereas Kentridge and his colleagues claim that he is. Whyis this? The answer is that the two sides of the debate are working with different definitions of attention.

Here is Prinz's definition of attention:

> 'attention can be identified with the processes that allow information to be encoded in working memory. When a stimulus is attended, it becomes available to working memory, and if it is unattended, it is unavailable.' (Prinz 2011, p. 184. Similar definitions can be found in Prinz 2012, p.93 and De Brigard and Prinz 2010, p. 52).

So for De Brigard and Prinz, attention is to be identified with the faculty that makes information available to working memory: all and only information that is available to working memory is attended to, on this definition. The concept of 'working memory' is a complex one, but luckily for us De Brigard and Prinz explain what they take working memory to be. To Prinz, working memory is:

> 'a short-term storage capacity, but one that allows for "executive control"… Once something is encoded in working memory, it becomes available to language systems for reporting, and with systems that allow effortful serial processing' (2011, p.184).

Prinz also expands the idea thus:

> '[t]he attended stimulus becomes available for processes that are controlled and deliberative. For example, we can *report* the stimulus that we consciously perceive, we can reason about it, we can keep it in our minds for a while, and we can wilfully choose to examine it further' (2012, p.92, original emphasis).

One important thing to note about these definitions of working memory is the presence of *reportability.* Notice particularly that when Prinz discusses reportability, he mentions language systems: for something to be encoded in working memory it must be available to a subject's language systems for *verbal report*.[26] Notice also that Prinz mentions that information available to working memory can (by definition) be used for *controlled and deliberative* action, so if a subject shows priming affects after being exposed to a stimulus, i.e. if she can perform above chance at discriminating certain stimuli when given a forced choice, but denies seeing it and claims only to be guessing what was there, then the stimulus does not count as having entered working memory, so long as we are working with Prinz's definition of working memory.

With this definition of attention in place, we can now see how the studies in question do not harm ST. The issue is this: in order to establish attention without consciousness, the experimenters had to establish that the subjects were not conscious of the stimuli in question. However, in order to do this they had to *ask* the subjects whether they saw the stimuli. We have already seen that the subjects denied seeing the stimuli in question. There is nothing dramatic or strange about this methodology: in normal experimental paradigms in psychology and neuroscience reportability is ubiquitously used to ascertain the presence or absence of consciousness.

The problem is that *if* the subjects do indeed deny seeing the stimulus then it will follow that the stimulus was not reportable, and thus not available to working memory, and (by De Brigard and Prinz's *definition* of attention) not attended to. So the data will not count

[26] Unless of course, the subject's language systems are somehow damaged or deficient. Prinz's idea seems to be that they are available to language systems *when those systems are present and behaving normally*.

as an instance of attention to something which was not phenomenally conscious. So the data in question will not count as a counterexample to ST. Indeed, Prinz seems to make just such an argument (though he does not state it in such stark terms) when he says the following:

> '[i]f my earlier analysis of attention is right, attention entails availability to working memory. Availability is clearly absent in blindsight, so GY cannot be instantiating *all* of the processes necessary for attention' (2012, p.115 cf. 2011, p.194).

Here Prinz relies upon his definition of attention in order to reject the claim that the GY studies demonstrate attention without consciousness. We can now contrast De Brigard and Prinz's definition with Kentridge's and his colleagues. In fact, Kentridge and colleagues supply at least two definitions of attention and it will be helpful to look at both:

> 'Visual spatial attention reflects the voluntary or involuntary prioritization of information in a selected part of a visual scene.' (Norman, Heywood and Kentridge 2013, p.836).
>
> And:
>
> 'The core of attention, as cognitive psychologists understand it, is the use of information to facilitate the execution of a task to which many stimuli might potentially provide the solution. The use of that information is facilitative… because it excludes some irrelevant stimuli from consideration.' (Kentridge, 2011, p.229).

Notice that these definitions of attention are far less demanding than Prinz's. For Kentridge, information need not be available to working memory for it to be attended. For Kentridge attention involves the 'prioritization' of information in certain parts of a visual scene; or the use of task relevant information to facilitate execution of the task when many stimuli could provide such a solution.

When we notice this difference in how Prinz and Kentridge define attention, the debate is cast in a new light. Take GY again. He is unable to report the presence of the stimuli in question (he denies awareness of them). As we have already seen, this unreportability disqualifies him from having paid attention to the stimuli, on Prinz's concept of attention. Conversely, when we use Kentridge's definitions of attention, we get a different verdict. To start with, we can take the first definition of attention given by Norman, Heywood and Kentridge above. GY is given a task which requires processing information about a specific target stimulus. He is superior at that task when certain cues are presented indicating where the targets will appear, and he is inferior at the task when the cue points away from where the target will appear. So it does seem that the cues are having the effect of prioritizing information processing to a certain part of the visual field. So, by the first definition of attention from Norman, Heywood and Kentridge given above, GY *will* certainly count as attending to the stimuli in question when the cue is valid.

Now take Kentridge's *second* definition of attention that I gave above (from his 2011). Things are a little less straightforward with this definition because it is not clear what Kentridge means when he says that information must be used to facilitate the execution of a task *that many stimuli might potentially provide the solution for.* I think we can get a handle on what Kentridge means by this italicised criterion by considering what he says by way of explanation of his definition later in the same passage. He asks us to consider subjects being given a task where they are given a series of pictures and asked to spot triangles in them. He then says this:

> 'being told in advance that any triangles in the next picture will be drawn in red could help subjects to solve this task… The use of this information to help in the performance of the task would be a paradigmatic instance of attention' (2011, p.229).

In this quotation, the crucial feature that Kentridge highlights is that attention is particularly attuned to certain properties (in this case redness) which will help to solve the task, and it preferentially selects redness over other properties that do not help to solve the task. For this reason I suggest that when Kentridge says that attention facilitates the execution of a task to which many stimuli might potentially provide the solution, he has in mind the idea that when we deploy attention in the hope of fulfilling a task, then the most efficient way to succeed is for attention to be particularly attuned to certain things to the detriment of others. As Kentridge himself says when elucidating his definition, if we want to detect red things then attention should be particularly attuned to red things, over things that are not red. I think that this attunement to certain properties, features or locations above others is all Kentridge has in mind when he says that attention should exclude some irrelevant stimuli from consideration. Notice that interpreting Kentridge in this way is plausible for another reason: because it makes his two definitions fit neatly with each other, as the first definition emphasises the idea of 'prioritization' of a certain area of the visual field over another area.

If we take this interpretation seriously then of course GY *will* count as attending to the stimuli in question. Valid cues have the effect of improving GY's performance, so it would seem to have the effect of preferentially boosting processing of task relevant properties at the location that they indicate, to the detriment of the locations to which they do not point. Similar things can be said about the meta-contrast masking studies: the subjects count as attending on Kentridge's, but *not* Prinz's definition.

At the heart of this debate is working memory. The stimuli in question are not available to the subjects' working memory systems, and the main evidence for this is that the subjects cannot *report* their presence. So once we have a definition of attention that implies

that stimuli unavailable to working memory are unattended, then the subjects will no longer count as having attended to the stimuli. If we go for a definition of attention that does not make this working memory demand (like Kentridge does), then we will likely conclude that the subjects *were* paying attention to the stimuli.

I thus claim that the disagreement really stems from different views on the meaning of the word 'attention'. I think this gives us good reason to think that (contrary to appearances) the real crux of the disagreement in this debate is a conceptual one over the meaning of 'attention' rather than an empirical one about some fact to do with GY, the meta-contrast masking subjects or what have you.

§2.2.5-Orienting?

Perhaps someone could respond to the problems that I have raised by saying that we should focus our efforts on assessing some of the *other* reasons that Prinz gives for thinking that GY is not really attending to the stimuli in question. Indeed, Prinz does argue that we have independent reasons for holding that the experiments in question do not show that GY was attending to the stimuli in question, based around the claim that GY was 'orienting' to the stimuli in question. I will briefly sketch out the 'orienting' response, and then show how similar problems arise here as well.

Here is what Prinz says about orienting:

> 'Orienting involves a shift in the allocation of input resources, either explicitly through hand and eye movements, or implicitly through receptive field changes. Informally, orienting alters what information gets in, and attention alters where it flows. (2011, pp.193-194).

Prinz elaborates upon this claim as follows:

> 'attention normally co-occurs with a shrinking of the receptive fields in the attended location in retinotopic neural areas such as V4 and V5. That means more cells respond to the stimulus, and the resulting representation has higher resolution. Such receptive shifts could explain GY's enhanced stimulus detection without assuming that GY also attends to the stimulus' (2011, p.193).

Prinz thus claims that this 'orienting' mechanism is sufficient to explain GY's performance, and so nowhere need we invoke 'attention' itself. There are several things that can be said to this response based around 'orienting'. One obvious problem is that Prinz relies upon the claim that GY was orienting to the stimuli *and that he was not attending to them,* (because, of course, GY's attending to the stimuli in question would be sufficient to disprove ST). The trouble is that (in the second quotation given above) Prinz himself admits that orienting and attention almost always go hand in hand. We normally orient to what we attend to, and vice versa. Prinz may be correct that orienting and attention *sometimes* are dissociable, but (given that Prinz admits that orienting and attention are *not normally* dissociated) the mere possibility of their being dissociated clearly does not show that they *were* dissociated in the case of GY. Indeed given that they *usually* go together, the default view should be that they were *not* dissociated in the case of GY.

However, even if GY's performance could be explained in terms of this kind of 'orienting' response the problem above would still remain. The issue is that if the orienting mechanism in question really *is* responsible for GY's increased ability in completing the task that he was set, then it seems clear that the orienting mechanism must be having the effect of selectively enhancing processing of task relevant information about the target stimuli, over other pieces of information, and thus facilitating the execution of the task. But if this is true, then by Kentridge's definition of attention, the 'orienting' mechanism will count as a variety

of attention. So Kentridge can say that even if the results are due to the fact that the stimuli were subject to orienting, they will still count as attended to (on Kentridge's definition of attention). So GY's case will still count as an example of attention in the absence of consciousness, and thus still a counterexample to ST. The basic issue here is not that the 'orienting' response is *wrong* per se but that it does not matter whether it is right or not, because either way Prinz will still insist that GY was *not* paying attention to the stimuli and Kentridge can still insist that he *was.*

§2.2.6-A reason to accept Prinz's definition of attention?

Prinz may reply to these worries by accepting that the crux of the disagreement is over the meaning of 'attention', but claim that *his* definition is *better* than Kentridge's. If this is true, then we should define attention the way that Prinz does. Some of the things Prinz says implies that he may take this line. Prinz does at one point accept that 'other researchers may choose to define attention differently' (2012, p.95). However, he also provides an argument *in favour* of his own definition of attention, and against other proposed definitions of attention, which it will be useful to examine in detail.

This raises an important question: can we expect one definition of attention to be 'better' than another one? There are a great many interesting issues connected to such a question, many of which will be examined over the course of the thesis. For now, it will be helpful to consider Prinz's argument in favour of his definition of attention, so that we can see how it connects with many issues considered over the course of the thesis.

Prinz proposes that we list 'paradigm' instances of attention, drawn from folk psychological discourse and then attempt to discover whether there is a common brain mechanism that underlies them all. If we find such a mechanism, we can identify it with

attention. Prinz then goes on to list some cases where attention seems to make information available to working memory (2012, pp.90-95). These include studies which link attention with short term memory retention (Rock and Gutman (1981)) as well as studies which have been thought to show that when working memory is full, it becomes harder to attend (Fougnie and Marois, 2007).

Prinz then makes the following claim:

> '[s]uch interactions between attention and working memory suggest an intimate relationship. The simplest explanation for this relationship is an identity claim: attention can be identified with the processes that allow information to be encoded in working memory ' (2012, p.93).

Prinz goes on to claim that:

> '[t]he idea of availability underlies all of the phenomena that we call attention … This account provides the only common denominator across the wide range of cases that we regard as examples of attention… the folk-psychological insight implicit in the range of phenomena that we call attention can map onto the empirical construct of availability to working memory' (2012, p.95).

Prinz's argument has the following structure:

1) In folk psychological discourse, what we refer to as 'attention' always correlates with availability to working memory.
2) The simplest explanation for this correlation is identity.

3) (Therefore) the folk use of the term 'attention' applies to availability to working memory.
4) We should identify attention with whatever the folk term 'attention' applies to.
5) (Therefore) attention *is* availability to working memory.

There are various places at which such an argument could be challenged. One of them is premise (1). Even before we examine the folk psychological usage of the word 'attention' it does seem extremely optimistic to expect it to map on perfectly to an already existing concept in empirical psychology such as availability to working memory. In chapter 3, I will provide general reasons for thinking that we should not expect the folk psychological usage of the word 'attention' to map onto one particular process or activity in the brain. Furthermore, in chapter 4 I shall present an example where the folk psychological use of the word 'attention' reliably departs from availability to working memory.

Premise (4) is also questionable. Needless to say (4) is unlikely to be accepted by many thinkers within empirical psychology, and it is also likely to prove unpopular with most philosophers. Even if it were the case that all folk usages of the term match up with availability to working memory, what reason do we have to prefer the *folk* usage over usages (such as Kentridge's) which deviate from this? I see no a priori reason to think that we should prefer the folk usage of attention to the usage of an empirical psychologist like Kentridge's, unless we were already convinced that the folk psychological use of the word was 'correct'. That is, unless we assumed that premise (4) was true. But obviously it will not do to assume premise (4) in an argument which will eventually lead to the conclusion that one's own use of the word 'attention' is 'better' than that of people like Kentridge. I conclude that Prinz's argument in favour of his definition of attention is unsuccessful

What of the empirical results that Prinz cites? Do they not show that there is an intimate relationship between working memory and attention? In response to this, I think we can accept that attention and working memory often interact closely (no one would deny this), but we need not commit to the conclusion that attention *must* be *identified* with availability to working memory.

I have examined the experimental data surrounding the GY experiments, as well as the meta-contrast experiments and the object-based attention experiments. I have considered certain interpretations of these experiments and rejected them as implausible. Then, I engaged an in depth examination of some of the conceptual background of these debates. I argued that the interlocutors in question are using the word 'attention' in ways different enough to deliver different answers to the main question in this debate, which is whether attention is sufficient for consciousness. Now, I will argue that these issues extend to other areas of the debate. In order to see this, I will turn to an argument *in favour* of ST, which comes from Declan Smithies (2011a and 2011b).

§2.3-Declan Smithies' argument in favour of ST.

In this section, I shall argue that Smithies' argument gains its force from the way that he defines 'attention'. This will serve to show how deeply these definitional problems run in this debate. In the course of his argument in favour of the sufficiency claim, Smithies supplies various considerations which bear on how we should define attention. He starts by providing two reasons for *not* defining attention in terms of its phenomenology. The first is this:

> '[i]f attention is defined in terms of its phenomenology, then we lose the theoretical significance of attention' (2011a, p.267).

By way of elaboration on this first point, Smithies says:

> '[i]f attention is just a distinctive kind of phenomenology, which plays no unified functional role… then why should we regard it as a central topic in philosophy and in cognitive science?' (2011a, p.252).

So, Smithies' first reason for resisting defining attention in terms of phenomenology is that we will then cease to be theoretically interested in it. I see no reason to think that if attention is defined in terms of its phenomenology, then we would not be interested in it from the point of view of cognitive science and philosophy and Smithies gives no argument to support this view. I shall set this argument aside here.

Smithies' other reason for not defining attention in terms of its phenomenology is more pertinent to the present discussion. As I mentioned in chapter 1, many thinkers (most notably William James) have defined attention directly in terms of its phenomenology. As I said, ST seems to simply fall out of such a definition of attention. Smithies is certainly aware of such a strategy for arguing in favour of ST, and this is his other reason for not defining attention in such a way: because then it would make ST trivially true, and would be a 'purely verbal maneuver' (2011a, p.259).

Even though Smithies does seem to be aware of these definitional issues, his overall argument for ST still seems to come down to the meaning that he assigns to 'attention'.

Smithies' argument has the following structure (Smithies, 2011a, p.248):

1) Attention is rational access-consciousness.
2) Rational access-consciousness is sufficient for phenomenal consciousness.
3) (Therefore) Attention is sufficient for phenomenal consciousness.

When Smithies says that attention is 'rational access-consciousness' he means that attention is the process that makes information 'fully accessible for use in the rational control

of action and thought' (2011a, p.248), and elsewhere he fleshes this out as the claim that attention makes information 'accessible to the subject as a reason that justifies the subject in forming a belief or performing an action' (2011a, p.262).[27]

Once Smithies has this definition of attention in hand, he uses a thought experiment to attempt to demonstrate the truth of (2) (2011a, p.261-63). The thought experiment in question is a variant on similar thought experiments used by Campbell (2002, 2004) and Dretske (2006). The basic idea is that non-neurologically impaired subjects can use information from their phenomenal experience to justify their beliefs and actions, whilst blindsight sufferers might form beliefs based on non-conscious visual perception from their blind field, but beliefs formed in this way *could not be justified.* If we are to have justification for our beliefs and actions (i.e. if we are to have rational access-consciousness) then we need phenomenal consciousness, or so Smithies claims. As he says:

> 'although unconscious information is sometimes accessible for spontaneous use in the control of action, it is not *rationally* accessible in the sense that it is accessible to the subject as a reason that justifies the subject in forming a belief or performing an action' (2011a, p.262).

With this argument in place, it is easy to see how Smithies can rebut the arguments against ST based on the empirical data that we have looked at. Smithies can claim that GY (who suffers from blindsight) cannot take information about the stimuli in question as a justifying reason for action and thought, and as such (by premise (1)) he cannot be attending to the stimuli in question. So, GY is not a counterexample to ST. The same applies *mutatis mutandis* to the case of the meta-contrast masking studies. The four subjects in those

[27] Smithies is borrowing the term 'access-consciousness' from Block (1995a). It is interesting to note that Block himself (in his original formulation of 'access-consciousness') mentions 'rationality' as well (Block, 1995a, p.382).

experiments could not take information about the masked discs as a justifying reason for action and thought and thus they cannot be attending to them. This applies to all of the experiments discussed above: Smithies' very demanding criteria for something to be attended are not fulfilled.

The problem with Smithies' argument is similar to the problem we found above with the Kentridge/Prinz debate. Smithies' thought experiment for premise (2) may go through, but (3) can only be established if we accept (1), which is itself simply a definition of attention. So again, it comes down to the meaning of 'attention'.[28] The crucial point that I am arguing for is that the debate turns on the issue of how different participants in the debate define their terms.

§2.4-More empirical data relevant to ST.

An opponent may at this point say that it could be that the debates over the studies that I outlined above do come down to how we define 'attention', but that my examination has been unnecessarily limited in scope. My opponent may say that I should concentrate on some of the *other* empirical data that are relevant to the issue in question. Such data may well settle the question of whether ST is true or not. I will now argue that exactly the same problem arises when we consider other prominent data which have been brought forward to

[28] Smithies does come close to giving an argument in favour of premise (1) when he claims that if we do not have a unified account of attention then we may become eliminativists about attention (2011a, pp.251-252). Smithies may turn this into an argument in favour of his definition of attention by saying that, if the choice is between accepting that attention is rational access-consciousness, and accepting that it does not exist, then we should accept the former option. Many readers may wonder what eliminativism about attention even amounts to, and may also be sceptical that eliminativism is a real threat. I will discuss these issues in chapter 5. For now, I will note simply that this argument cannot give us any *more* reason to accept Smithies' account of attention than, say, Prinz's or Kentridge's, so it cannot be used to support premise (1) of his argument here.

argue against ST. In this section, I shall discuss two possible sources of data that have been thought to be relevant to the issue.

§2.4.1-The erotic images studies.

Mole (2009) mentions a collection of experiments by Jiang et al. (2006), which have been thought to bear on the issue of ST. The experiment exploited the gender and sexual orientation of its subjects. The experiment is depicted in figure 3.

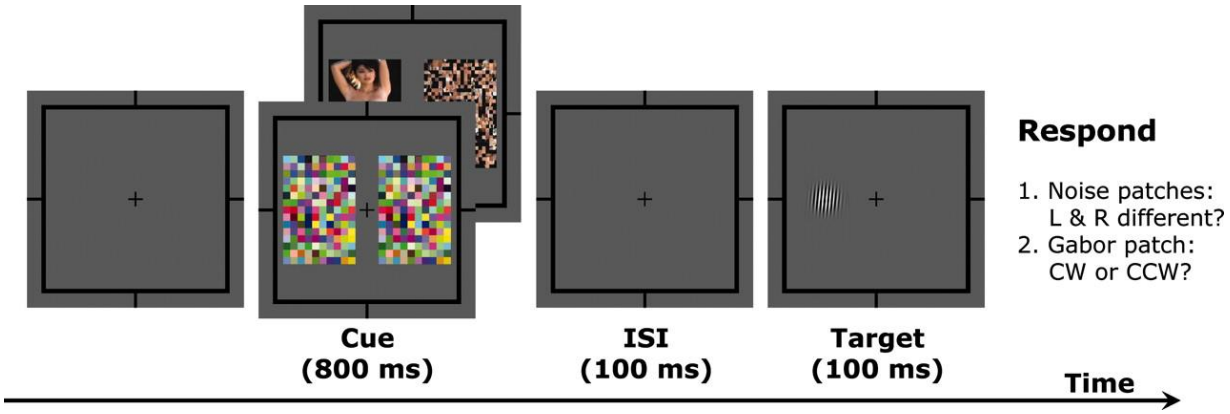

Figure 3: The sexual orientation based experiment, which makes use of visual 'noise'. From Jiang, Y., Costello, P., Fang, F., Huang, M., He, S., & Purves, D., 2006, "A gender- and sexual orientation-dependent spatial attentional effect of invisible images", *Proceedings of the National Academy of Sciences* **(USA), 1003(45): 17048–17052. Copyright (2006) National Academy of Sciences, USA.**

This experiment makes use of binocular rivalry and interocular suppression. Binocular rivalry occurs when different and mutually incompatible stimuli are presented to each eye. The information from the eyes then 'competes' and only the stimuli presented to one eye are consciously perceived (hence the term 'rivalry'). In some cases, the subject will perceive the stimuli presented to one eye, and then slowly this will fade, and they will begin to perceive only the stimuli presented to the other eye. This process will continue, with subjects perceiving first one set of stimuli, and then another. This effect has been reliably

reproduced in monkeys as well as humans and has been extensively used to identify the neural correlates of vision (Leopold and Logothetis, 1996).[29]

The experimental utility of binocular rivalry was increased by the development of interocular suppression (see Tsuchiya and Koch, 2005.The technique is sometimes called continuous flash suppression). In this technique, a 'noise patch' is presented to one eye. A noise patch is a high contrast image which is constantly flickering (typically at a rate of 10Hz). The colourful images in figure 3 are examples of such noise patches. The other eye is presented with a much lower contrast image, which is not flickering. The effect is that the noise patch reliably 'wins out' against the rival low contrast image, and the noise patch is consciously perceived, whilst the low contrast image is not (though information about the low contrast image is processed in the visual system (Fang and He, 2005)).

In the erotic images experiment, the subjects were ten heterosexual males, and ten heterosexual females (another experiment of the same type was carried out on ten homosexual males, and ten homosexual and bisexual females). Subjects themselves had to rate how homosexual or heterosexual they were on a scale from 0 to 6 (Jiang et al. 2006, p.17051). A fixation point was given, followed by the presentation of two different sets of stimuli, one set of stimuli to each eye. Each set contained two stimuli, so *two stimuli* were presented to each eye. One set of stimuli contained two identical noise patches. The other eye was presented with a pornographic image of either a male or a female, and also a scrambled version of that pornographic image (the pornographic image and its scrambled counterpart can be seen behind the colourful scrambled images in figure 3).

[29]I discuss binocular rivalry in more detail in Taylor (2013d). See also Chalmers (2010, ch.3).

As a result of interocular suppression, the colourful stimuli 'won out' and the pornographic image and the scrambled version of the pornographic image were not consciously perceived, but the two scrambled colourful images were. In order to test that the two colourful scrambled images really did 'win out', subjects were asked to abort the trial if they noticed any difference at all between the two stimuli that they could see. Since such a task would be easy if they could see the pornographic image and the scrambled version of the pornographic image, the fact that subjects did not abort the trials is good evidence that the pornographic image and the scrambled version *did not* win out over the two colour patches. Subjects aborted the trials less than 1% of the time (Jiang et al. 2006, p.17052).

Another piece of evidence for the claim that the pornographic images were not consciously seen is the control experiment that was carried out (Jiang et al. 2006, p.17052). In this control experiment the set up was precisely the same, except that subjects were *forced* to guess whether the (supposedly invisible) pornographic image was on the left or the right. This would be very easy if they could consciously *see* the image, but they did not perform above chance in the task, indicating that they couldn't consciously see it.

The two sets of images appeared, and then disappeared from the display. There followed a 100ms inter-stimulus interval with nothing on the display. Then a Gabor patch appeared on one side of the fixation point. A Gabor patch is a rippled texture, tilted to a specific orientation and can be seen in the box at the right of figure 3. The Gabor patch was tilted either clockwise or anti-clockwise, and subjects were asked to discriminate the orientation of the patch and report it by means of clicking either the left button of a mouse for one orientation, or the right button for another.

It was found that heterosexual males were better at discriminating the orientation of the Gabor patch when the Gabor patch appeared at the location that had previously been occupied by a pornographic image of a *female.* Heterosexual females were better at the orientation discrimination task when the image of the Gabor patch appeared at the location previously occupied by a pornographic image of a *male*. Homosexual males were superior at the discrimination task when the Gabor patch appeared at a location previously occupied with a pornographic image of a male. Bisexual and homosexual females were superior at the discrimination task when the patch appeared at the same location as a pornographic image of a female (Jiang et al. 2006, p.17049). These are fascinating results, because they show that whether the pornographic images impacted on increased performance in the discrimination task depended on the sexual orientation of the subject performing the task.

This experiment has been taken as evidence that the pornographic images attracted the attention of the subjects, even though the images were invisible to the subjects. The problem with these experiments is of course that it seems clear that the subjects were unaware of the pornographic images. As I have already said, the subjects were unable to perform above chance at a task which would have been easy if they did have access to information about the pornographic images. This failure to use information about the pornographic images in completion of these tasks can be taken as good evidence that information about the pornographic images was not available to the subjects' working memory systems at the time (because if information was so accessible, the subjects would have responded to the stimuli in the way that the control tasks were intended to probe).

However, if information about the erotic images was not accessible to the subjects' working memory systems, then (by De Brigard and Prinz's definition of attention) they will

not be attended to and *not* a counterexample to the view that attention is sufficient for consciousness. The same problems emerge for these results as all of the others. Similar things go when we apply these data to Smithies' view. Smithies can claim that information about the pornographic images was not available for use as a justifying reason for action and thought. Therefore (by Smithies' definition of attention) they were not attended to. Therefore there is no counterexample to ST here.

§2.4.2-The Soto et al. data.

Similar worries to those I have expressed above will apply to another suggestion which attempts to disprove De Brigard and Prinz's view. The relevant example was given to me by Bob Kentridge, and comes from Soto et al. (2011) (figure 4).

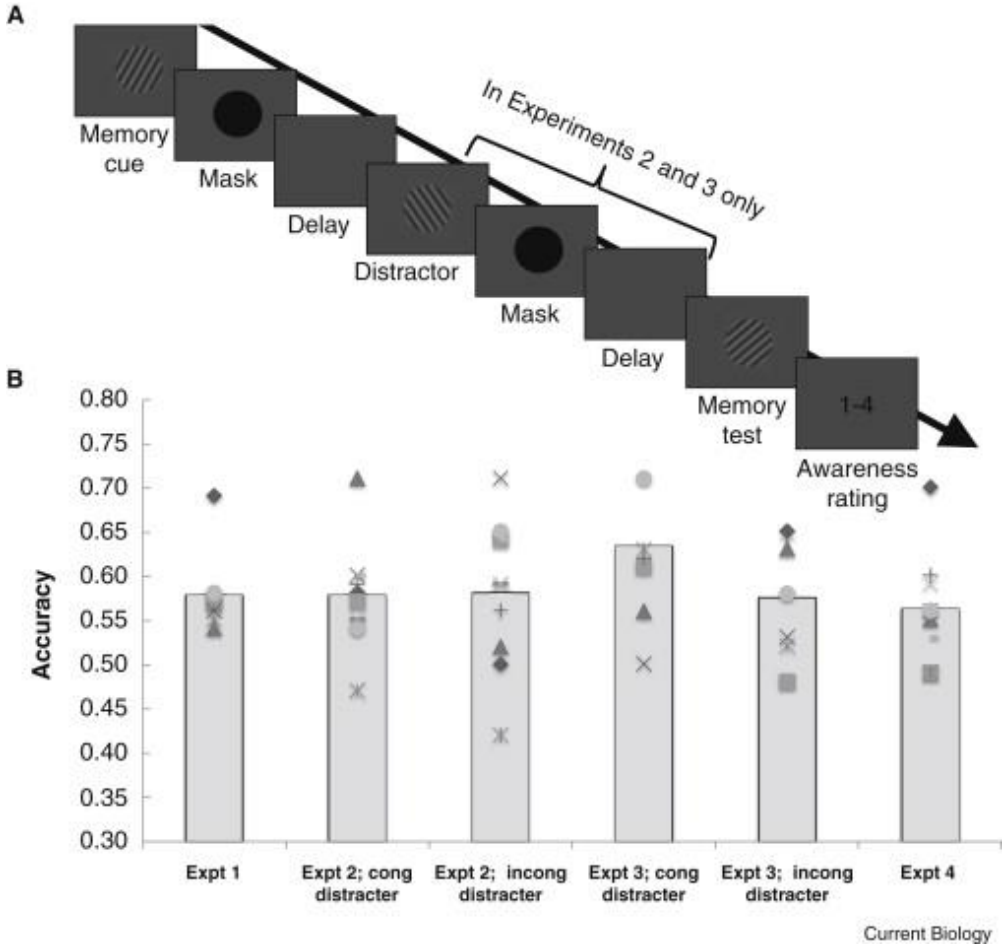

Figure 4: The Soto et al. experiment and results. Figure 4(a) will be most important for our purposes. From Soto, D., Mäntylän T. and Silvanto, J. 2011. "Working memory without consciousness." ***Current Biology.*** **Vol.21. pp.R912-R193. Reproduced with permission from Elsevier.**

What matters for our present purposes is the experimental procedure depicted in figure 4(a). As can be seen in the panel on the left of 4(a), subjects were presented with a Gabor patch, tilted to a specific orientation. This was followed by a mask, which prevented the initial Gabor patch from entering consciousness (the mask is depicted in the second panel from the left of 4(a)). Then there was a delay. We can ignore the section marked 'Experiments 2 and 3 only' as they will not matter for our current discussion (these did not appear at all in the version of the experiment we are interested in). After the initial Gabor patch and the mask were presented, a second Gabor patch was presented which subjects could clearly see. Subjects were asked to assess whether the second Gabor patch was tilted

to the same orientation as the first. Following this, subjects were asked to score their awareness of the *first* Gabor patch on a scale of 1-4 (1 being 'completely invisible' and 4 being 'completely visible'). This '1-4' awareness rating can be seen on the right hand panel of figure 4(a). Only experiments where subjects gave a rating of '1' were considered.

Subjects were well above chance at comparing the orientation of the first (invisible) Gabor patch and the second (visible) Gabor patch. The experimenters conclude that the *first* Gabor patch must have been encoded in working memory (and thus must have been *available* to working memory) even though it is unconscious. If this were true it would clearly be incompatible with De Brigard and Prinz's position because to them availability to working memory is identical with attention. So, if the first Gabor patch really were available to working memory but unconscious, then it will count as attended to but unconscious, even by De Brigard and Prinz's definition of 'attention', and so it should be a counterexample to ST. Are the Soto et al. results the counterexample we're looking for?

The answer is no, and the reason once again lies with definitions. This time it lies with Prinz's definition of working memory itself. Prinz's response to these data is to claim that the subliminal Gabor patch was not really encoded in working memory (2012, p.96). Prinz's reason for this is obvious. If the subjects deny seeing the first Gabor patch (as they did), then it will follow from Prinz's *definition* of working memory (which includes reportability) that it was not available to working memory. So the first Gabor patch will not count as an instance of an unconscious representation that is available to working memory, and thus not an example of something unconscious but attended. Notice that here, much of the work is being done by Prinz's definition of working memory, which in turn affects his definition of attention, since attention is defined in terms of working memory.

So by Prinz's definition of working memory, these data are not a counterexample to his claims. Let us now compare this to Soto et al.'s definition of working memory:

> 'Working memory allows individuals to maintain information in the focus of the mind's eye in the service of goal-directed behaviour' (2011, R912).

This definition is extremely vague, given that it includes mention of 'the mind's eye', which is difficult to make precise. Notice that on *some* interpretations of this definition of working memory, it will follow that the initial stimulus *was* available to working memory; on stricter definitions, it will follow that the stimulus *was not* available to working memory. So it is not even clear that Soto et al.'s data count the first Gabor patch as encoded in working memory *by their own definition* of working memory.

Indeed, we do not need to look far in the empirical literature to find different definitions of working memory, some of which would deliver the verdict that the first Gabor patch in the Soto et al. experiment *is* encoded in working memory, and others of which would deliver the verdict that it is *not.* For instance, Awh et al. (2006) define working memory as the 'temporary maintenance of information in a limited capacity system' (p.201). Such a definition of working memory is not very demanding, and as such would allow that the first Gabor patches in the Soto et al. experiment are in fact encoded in working memory. Conversely, when Victor Lamme (2003, p.13) discusses working memory, he (like Prinz) explicitly mentions *reportability* of the stimuli that are available to working memory. Lamme's definition would imply that the first Gabor patch in the Soto et al. experiment was *not* encoded in working memory (because it couldn't be reported). This problem is widespread, as Ned Block says:

> '[O]ne of the first things that strikes a reader of the "working memory" literature is that the term *working memory* is defined differently by different theorists' (2007b, p.539).

Once we link attention and working memory (as Prinz has) then how we understand working memory directly feeds into how we understand attention. Of course, it will also not follow that the Soto et al. data can be used to build a counterexample to Smithies' argument either. Smithies can perfectly consistently claim that information about the first Gabor patch cannot be taken as a justifying reason for action and thought, and thus that it does not count as attended to.

§2.4.3-A summary of the situation we are in.

I have argued that the empirical evidence marshalled against ST fails when we employ the definition of attention presented by those who defend ST. I have argued that the real point of friction between the two sides of this dispute is that they disagree over the meaning of the word 'attention'. I have argued that this problem arises with relation to the GY and meta-contrast masking studies and also applies to various other pieces of data that are presented in this debate.

Mole's (forthcoming a) argument is an excellent example of how these difficulties continue to recur time and again in this debate. Mole claims that the experiments we have mentioned demonstrate that attention can be unconscious. Mole takes this as a problem for theories such as Smithies' and Prinz's that predict that attention cannot be unconscious.[30] But Prinz and Smithies are free to claim that attention cannot be unconscious (based on their definitions of attention) and therefore that whatever the subjects in question were doing, it

[30] Mole also targets Sebastian Watzl's theory. We will examine Watzl's view in chapter 4.

was not *attending.* Each side of the debate relies on a view of 'attention' which their opponents will simply reject.

Before I go on to suggest some positive directions for resolving these problems, I should delve further into the problems that we are faced with here. In §2.5 I will say something about NT. In §2.6 I go on to examine the question of whether *any* empirical data could disprove De Brigard and Prinz's position.

§2.5-Is Attention Necessary for Consciousness?

The literature on whether attention is necessary for consciousness is even more vast and confusing than the literature on the sufficiency. I certainly do not intend to discuss it all here, especially given the time I have already dedicated to ST. However, I will discuss one particularly prominent piece of evidence that has been brought *against* NT, in order to highlight how what I have been saying about ST affects this debate as well.

Typically, the evidence that is marshalled *against* NT is not intended to demonstrate consciousness in the complete absence of attention.[31] Rather, the evidence is usually taken to demonstrate consciousness of a stimulus in the *near or complete* absence of attention. Sometimes further qualifications are added. For example, it is sometimes claimed that consciousness of a stimulus is possible in the near or complete absence of endogenous attention to that stimulus.[32] The reason that thinkers have included these qualifications is that it has proved extremely difficult to demonstrate conscious awareness of a stimulus without *any* attention to the stimulus. After all, for any stimulus at all that a subject is

[31] There are exceptions. One thinker who argues that consciousness is possible in the *complete* absence of attention is Victor Lamme (2003 and 2004).

[32] Another relevant distinction that is often emphasised in this context is between *focal* and *peripheral* attention. It is sometimes sad that it may be that consciousness of stimuli is possible in the absence of focal attention, but that the stimuli were subject to peripheral attention.

conscious of, one could always claim that the stimulus was subject to just a 'little bit' of attention.

Some thinkers despair at the prospect of ever experimentally dissociating the near absence of attention from the total absence of attention. However, because the problems with this dissociation are already well known (and have been discussed at length) I will not specifically discuss them here.[33] Rather, I wish to examine one prominent piece of evidence that has been brought against NT, in order to see how the issues that I have been raising in this chapter bear upon it.

The evidence that I will discuss comes from Li et al. (2002). It is depicted in figure 5.

[33] They are discussed in Li et al. (2002), Koch and Tsuchiya (2007), van Boxtel et al. (2010), Cohen and Dennett (2011), Cohen et al. (2012a and 2012b) Tsuchiya, Block and Koch, (2012) and Aru and Bachmann (2013).

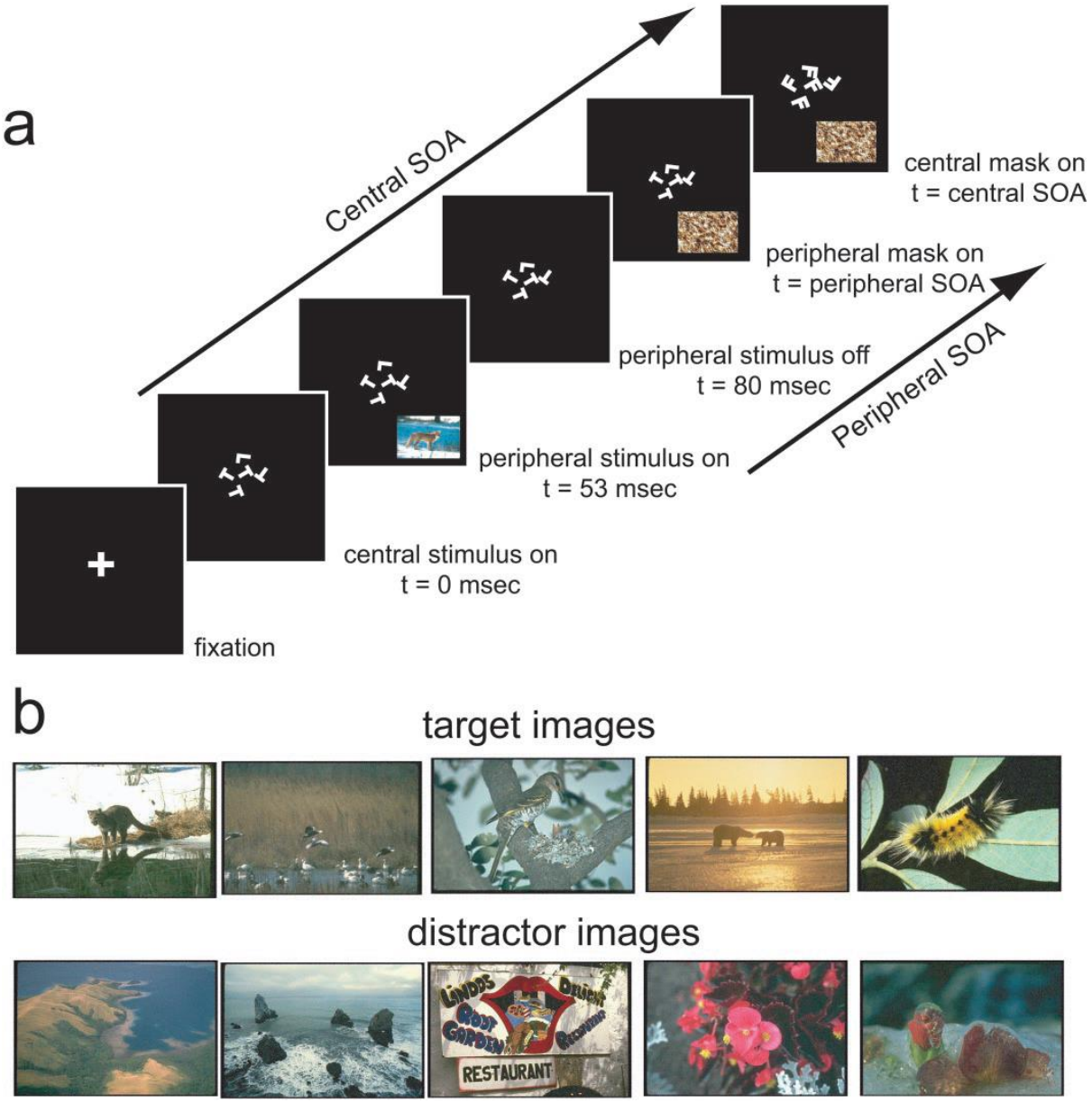

Figure 5: Figure 5(a) depicts the dual task of letter discrimination and peripheral picture discrimination. In this figure, the arrow depicts the direction of time. The letters at the centre of the display were the main task, the picture of the animal in the bottom right of the displaywas the peripheral task. In figure 4(b) we have some examples of the stimuli used in the peripheral discrimination task. From Li, F. F., VanRullen, R., Koch, C. and Perona, P. 2002. "Rapid natural scene categorization in the near absence of attention." *Proceedings of the National Academy of Sciences of the United States of America.* Vol.99, pp.9596-9601. Copyright (2002) National Academy of Sciences, USA.

We can start to understand the experiment by looking at figure 5(a). In this experiment, subjects were given a central fixation point. Then a collection of letters appeared at the centre of the display. Subjects were instructed to determine whether all five letters

were the same, or whether one of them was different from the others. Whilst they were performing this task, a picture was flashed up in the periphery of the display. Subjects were asked to determine whether the picture depicted an animal or not. Subjects were asked to keep a button pressed throughout the experiment. They were told to release it when an animal was present in the peripheral picture.[34] The top row of pictures in figure 5(b) are examples of the ones that subjects were instructed to respond to, and the bottom row are examples of the ones that subjects were instructed *not* respond to. The peripheral picture remained on the display for 53ms, and then disappeared. It was followed by a mask. In this experiment (unlike the above experiments) the mask was not there to prevent conscious perception of the image that it was masking. Rather, the animal *was* consciously perceived, but the mask ensured that it could no longer be seen after it disappeared from the display. Notice also (figure 5(a)) that at the end of the experiment the letters in the central task were replaced by other letters, which were not relevant to the task. This was also to control how much exposure subjects had to the central letter-discrimination task.

There were also two control experiments; one in which subjects had to perform *only* the letter recognition task, and one in which subjects had to *only* perform the peripheral picture identification task (Li et al., 2002, p.9596). The crucial result is that subjects performed *no worse* in the versions where they had to perform *both* tasks as when they had to perform just one of them.[35] The authors of the article themselves claim that (because the central task required so much attention) the picture used for the peripheral task must have been perceived in the 'near or total' absence of attention, even though it was consciously *seen.*

[34] In some versions of the experiment, subjects had to identify whether there was a vehicle present, rather than an animal.

[35] A similar experiment was carried out by Reddy et al. 2006.

Though the authors are careful to state that a *little* attention may be deployed to the peripheral task, these results have been taken by many as one of the clearest experiments opposing NT that we have (e.g. van Boxtel et al. 2010, p.4).[36]

I am less impressed with this kind of experiment than by the others discussed above, at least when it comes to its power to disprove NT. There are various points at which I think we can criticise it. However, what is most pertinent for our purposes is the claim that subjects were phenomenally conscious of the peripheral stimuli that they (supposedly) paid no attention to. Either this peripheral picture is reportable or it is not. If it is reportable then we can conclude that it is conscious but we can also conclude (by De Brigard and Prinz's definition of attention) that it is attended, so these results are not a counterexample to NT. If it is not reportable, then there is good reason to think that it is *not conscious* and once again, we have no counterexample to NT. The data can't touch the central claim.

§2.6-Could any empirical data refute De Brigard and Prinz's position?

We can now extend the observations made so far to develop an argument for a more general claim, which is that *no* empirical methods in which reportability is used to establish the existence of consciousness could possibly disprove De Brigard and Prinz's claim that attention is necessary and sufficient for phenomenal consciousness, or Smithies' claim that attention is sufficient for consciousness.

Here is my argument:

[36] See Lamme (2010a, 2010b) for a use of a partial report paradigm to disprove NT. Ned Block (2005, 2007a, 2007b, 2011 and 2012) has often been interpreted as using the partial report paradigm to argue against NT (Stazicker, 2011 interprets him this way). In fact, Block is careful to distinguish between attention and what he calls 'cognitive access', and it is the latter of these that Block takes to be unnecessary for consciousness. I cannot discuss the arguments that make use of partial report here, but I will touch upon them in chapter 4.

1) When reportability is used in order to establish phenomenal consciousness, reportability and phenomenal consciousness are understood to be correlates of each other. That is to say, if subjects can report the presence of a stimulus then this is taken as good evidence that they were conscious of it, and if they cannot report it, this is evidence that they are not conscious of it.
2) In order to disprove NT or ST, we would require a case where attention and phenomenal consciousness dissociate.
3) By defining attention in terms of working memory, and working memory in terms of reportability, De Brigard and Prinz have ensured that reportability will *always* correlate with attention.
4) (Therefore) anything which we can establish is conscious will also count as attended to, and anything that we can establish as unconscious will count as unattended to, as defined by De Brigard and Prinz.
5) (Therefore) Their position cannot be empirically disproved using experimental methods which use reportability to establish consciousness.

The basic point of this argument is just that the experimenters cannot win. The *very same* results that are used to establish the absence of consciousness can be used to establish the absence of attention; results which establish the presence of consciousness equally establish the presence of attention. There is no possibility for dissociation.

The meaning of 'reportability' in premise (1) is left deliberately vague. Of course it includes *verbal* report, but could also include broader concepts of reportability, as is used in the study of monkey consciousness (e.g. Leopold and Logothetis, 1996 and Logothetis and Schall, 1989). Paradigms involving monkeys use reportability of at least some kind, as they

rely on monkeys pressing bars at certain times to indicate what they can see. There are certain constraints on what can count as 'reportability': subjects who display priming effects (where subjects respond meaningfully to stimuli but deny awareness of them) do not count as 'reporting' the stimuli in question. This is consistent with how reportability is typically used in empirical psychology when establishing consciousness: priming effects are not normally taken to indicate consciousness of the stimulus that caused the prime; but the subject's wilful and voluntary report of a stimulus typically *is* taken as evidence of consciousness of that stimulus.

This argument can also be run *mutatis mutandis* to demonstrate that no evidence drawn from experimental paradigms that use reportability as a proxy for consciousness could disprove Smithies' claim either. It is hard to see how we might establish whether a subject was justified in taking information about something as a justifying reason for action and thought unless we *ask* her. If she does seem to be able to report information about the stimuli in question, and has reasonable knowledge about it, then we can conclude (by Smithies' account of attention) that she was attending to it. If she denies seeing the stimulus, then we will conclude that she not conscious of it or attending to it. Either way, a counterexample to ST is impossible.

Consider the way that De Brigard and Prinz motivate their position. As I outlined above, they use empirical data to support their view, and these data are drawn from experimental paradigms that use reportability of a stimulus as evidence of consciousness of that stimulus. However, given the argument above, it is not surprising that these empirical data support De Brigard and Prinz's conclusion, and that any suggested counter evidence

fails to disprove their view. Their definition of attention serves to *motivate* their view and *protect* it from falsification.

In my (2013a) I claimed that De Brigard and Prinz's position was immune to any empirical falsification of any kind whatsoever. Now I think this claim is somewhat quixotic. After all, we do not know how empirical science will develop in the coming years and it would be premature to claim that we will *never* be able to disprove their claim. Furthermore, there are several arguments that attempt to empirically separate out reportability and consciousness (e.g. Block, 2007a, 2007b and 2014). Nonetheless, this argument demonstrates that De Brigard and Prinz's position occupies a uniquely protected position in the debate. It also serves to explain *why* none of the empirical evidence so far marshalled against the position has succeeded in disproving the view.

§2.7-Why is this important?

Many thinkers assume that the debates over ST and NT are largely empirical debates. Philosophers partake in the debates but they mostly do so by engaging with the empirical literature. Different interlocutors' assertions of apparently contradictory statements such as 'attention is sufficient for consciousness' and 'attention is insufficient for consciousness' gives the impression that they are disagreeing about some empirical facts, which are subject to discovery by science. If what I have been saying in this chapter is correct, this is confused. If I am right, then the *real* point of friction between the interlocutors has been largely missed. Empirical results are obviously very prominent and important in this debate but (in my view) the *real* core of disagreement stems from usages of certain terms which are different enough on both sides of the debate to give different answers to the central questions. To put it in simple terms: all of the interlocutors in this debate can basically agree on the question of

what the subjects in the experiments were *doing*, but they disagree over whether what they were doing should be called 'attention' or not.

One important conclusion that we can draw at this point is this: it is misguided to attempt to look at the empirical data for an answer to the question of whether the subjects are attending to the stimuli in question. At best, examining the empirical data themselves is a highly incomplete method for establishing this. The data cannot themselves give us the answers we want, because the data cannot settle one way or the other the question of whether 'attention' should be identified with availability to working memory, rational access-consciousness or selection of action for facilitation of a task.

§2.8-Two possible reactions and the nature of attention.

There are at least two reactions one could have to the above observations that I have made about the debates in question, and these different reactions are often manifestations of much deeper views about the nature of attention itself. It will be helpful to highlight these issues now, as they will occupy me later in the thesis.

§2.8.1-The two reactions.

The first reaction (expressed to me by Sam Coleman amongst others) is to say something like 'it is indeed true that the interlocutors in this debate are disagreeing about the meaning of the word 'attention', but there is still a substantive debate to be had: the debate over which one is *really talking about attention*!' This kind of reaction fits well with much of the current philosophical literature on attention. One of the most prominent debates in this literature is over the question 'what is attention?' and there is much hope in this literature that we will arrive at a complete and correct theory of attention. It might be said

that we should hold off our evaluation of NT and ST until we have such a correct theory. When we do have such a theory, we can examine the data in question in light of it.

Another reaction to what I have been saying would be to claim that really, I have dissolved the debate over whether ST and NT are true or not. That is to say, if I have indeed shown that the debates come down to different interlocutors using the word 'attention' in different ways, then I have shown that the debate is merely verbal and not worth worrying too much about. Someone for whom this response is tempting would be likely to say that it does not make a great deal of difference whether we use the word 'attention' to refer to accessibility to working memory, or whatever. The natural conclusion to draw from this view is that no one side of the debate is more correct than the other, because their apparently conflicting views are really just a result of using the word 'attention' differently.

Someone who held this latter view would could then say that there is no one answer to the question of whether ST is true or not, and that there will in fact be multiple answers to the question of whether attention is sufficient for consciousness depending upon how one defines 'attention'. For example, it may be that rational access-consciousness *is* sufficient for phenomenal consciousness, whilst the Kentridgian understanding of attention (as increased task performance or prioritization of information in a part of the visual scene) is *insufficient* for consciousness. Someone who held this view would be likely to say that the residual question of which one *really is attention* is a misplaced question, as no single understanding of attention is in this sense privileged or 'more correct' than the others.

§2.8.2: Monism and pluralism.

It will be useful at this point to discuss some of the deep disagreements about the nature of attention that are plausibly fuelling the two different reactions that I discussed in

§2.8.1. The views I have in mind are monism and pluralism. We can understand pluralism as the view that there are several *importantly different* 'attention' concepts, no one of which is more privileged or 'correct' than the others.

Pluralism of at least some kind plausibly lies behind the second reaction that I mentioned in §2.8.1. Specifically, if someone thinks that the ST debates stem from the use of different concepts of 'attention', no one of which is 'correct' or 'privileged', then one will likely be attracted to the view that there are many equally legitimate attention concepts; no one of which is the 'correct' one. As we progress in this thesis, pluralism will be refined, expanded upon and the core claims of it will be made more specific.

The opposing view is monism. This is the view that there is one privileged 'attention' concept. To monists, there is *one* account of attention which is *better* than all of the others in some important sense. Monism is held, either implicitly or explicitly by many interlocutors in the debates in question, and we shall be encountering many such thinkers as we progress. Above, I mentioned that one could react to the problems that I have been pointing out with NT and ST by saying that we should work out which side of the debate is *correct* about what attention is and then judge the empirical data based upon that. This reaction is *not* flatly inconsistent with pluralism. Suppose for example that one believed that Smithies' and Prinz's concepts of 'attention' were 'correct' and that neither of them was privileged over the other, but for some reason one thought Kentridge's was *wrong* and that we should exclude it as somehow inferior to the other two. Both Smithies' and Prinz's concepts deliver the same verdict on whether GY was attending to the stimuli in question or not. So, in this case one would be a pluralist (because one accepts that at least two attention concepts should be

accepted, and that no one of them is privileged) but one would still think that there was an objectively correct answer to the question of whether GY was attending or not.

§2.9-What next?

I have argued that the debates over ST and NT are beset with conceptual difficulties that stem from different interlocutors disagreeing over the meaning of the word 'attention'. I have also mentioned one possible solution to this problem, which is to decide which use of 'attention' is *better* than the others, and interpret the evidence based upon that. This suggestion will be examined in the next chapter.

Chapter 3
The Road to Pluralism: first steps.

§3.0-Summary.

In this chapter, several strategies for dealing with the problems pointed out in the last chapter will be examined. These strategies revolve around the suggestion that we should decide which definition of attention is 'better'. The usage of the concept of 'attention' in folk psychology and empirical psychology will be examined, and it will be argued that these cannot help us to decide which use of 'attention' is 'better'. It will also be argued that we cannot decide which concept of attention is better based upon considerations such as theoretical usefulness. Independent arguments for pluralism about 'attention' will also be given.

§3.1-Solving the problems of chapter 2.

In the previous chapter, I argued that different interlocutors are using the term 'attention' in significantly different ways, such that the different uses of the term deliver different answers to philosophically and psychologically prominent questions about the interaction between attention and consciousness. I also pointed out two different reactions that one could have to this problem, one of which was to insist that *there is* an objectively correct answer to the question, because one concept of attention is privileged or correct or better than the others.

The basic idea behind this proposal is that if we are to make progress on questions such as the sufficiency or necessity question, we first need to examine the nature of attention itself. This is one kind of view we find expressed in the literature. Thus, Watzl:

> 'while the debate about the connections between attention and consciousness has been heated and fruitful it could benefit from a clearer conception of what attention actually is' (2013).

I will begin by unpacking the suggestion that one concept of 'attention' is 'better' than another. I shall then present several ways that we might decide which meaning of attention is 'better' and I shall argue that *none* of them can be used to evaluate the candidate definitions on the table in a way that will allow us to settle the problems I pointed out in the previous chapter. I will also use the arguments of the chapter to present some reasons to embrace pluralism about 'attention'. However, the *main* argument in favour of pluralism about 'attention' will come in the next chapter.

Some more preliminary points are necessary before I begin. Firstly, in the previous chapter I pointed out several candidate concepts of 'attention' that are causing trouble in the debates in question. My primary concern in this chapter will be with the question of whether we can pick one of *these particular* concepts of attention as superior to the others. However, it could be that there is some *other* concept of attention which is superior to *all* of these, and perhaps it is *this* concept that we should use when we are discussing the sufficiency and necessity questions. An opponent could push this challenge in at least two ways: she could say that there are a great many other candidate definitions of the word 'attention' in the literature, and that perhaps one of them is the 'best' one; or she could claim that perhaps there is a 'best' definition of the word 'attention' *that we have yet to think of.*

With respect to these suggestions, I should say that in the next chapter I will examine some of the accounts of 'attention' that have been most prominent in the literature, and see how consideration of the issues raised in this chapter apply to those accounts of attention specifically. Secondly, when it comes to the claim that there might be some other 'superior' concept of attention that we have yet to think of, I will say that the arguments of this chapter are intended to oppose this suggestion as well. The arguments of this chapter are also meant

to tell against the view that there are *several* different concepts of attention which are privileged or more correct than the others, but that they all deliver the same verdict on ST and NT. The main claim of this chapter is this: *any* general attempt at selecting a privileged concept (or concepts) of attention that are sharp enough to resolve the problems examined in chapter 2 must face the challenges that I lay down in this chapter.

There is also a more complex position in logical space, which would be that there *is* one objectively 'correct' concept of attention but that *we cannot know what it is.* Perhaps there is some epistemological barrier to us grasping the best concept of attention, but we should still believe that it exists. Such a position is interesting but it will not be the primary focus of this chapter. In this chapter I aim to examine the claim that we *can* decide which concept of 'attention' is superior.

§3.2-Unpacking the claim: Folk psychology.

How should we go about deciding whether a certain concept of attention is better than another? What does 'better' mean here? Better how? In order to get a handle on the ideas, I suggest that we should look at those thinkers who present reasons in favour of their own account of attention, and against others' accounts. By examination of their argumentation style we can gain an understanding of what the claim that one account is 'better' than another might mean. As we saw in §2.2.5, Jesse Prinz considers coherence with the *folk psychological use* of the term 'attention' to be one important advantage of a theory of attention. Indeed, Prinz argues that his own set of necessary and sufficient conditions for 'attention' gives a functional analysis of the *folk* concept of attention.

An examination of the literature shows that Prinz is not alone in considering coherence with the folk psychological use of 'attention' an advantage of a theory of attention. The thinkers in question often draw examples of normal day to day activities, and then rely on pretheoretical folk psychological intuition over whether the subjects in the examples are 'paying attention' or not, and then use this to motivate their own theories of attention. For example, Wu uses the example of kicking a basketball (2011a, p.99 and 2014, p.80); of picking up a hammer (2011b, p.53); of staring at a beautiful woman (2011a, p.107) and of being unable to get a tune out of one's head (2011a, p.107) as examples of attention which he then builds his theory around. Drawn as they are from folk psychological discourse, one would think that Wu's approach is to develop an analysis of the folk use of 'attention'.

Similar things go for Christopher Mole, who draws many of his examples of attention from normal everyday life, and relies on folk psychological intuitions about whether the cases in the examples are cases of 'attention' or not. Like Wu, Mole then takes them as the guidance for his discussion, around which he builds his theory of attention. At the start of his book, Mole sketches out three kinds of 'attention', drawn from folk psychological discourse, and claims that his theory can explain them all (2011a, pp.5-6). Mole also gives the example of 'daydreaming' (2011a, p.57) as a case of inattention, and argues that it is a virtue of his theory that 'daydreaming' is *not* counted as 'attention'. Mole's theory also excludes certain autonomic systems like the actions of the hypothalamus from counting as 'attention', and he justifies this exclusion by giving an example of a situation from everyday life (someone jogging) and claiming that intuitively we would not count the person's autonomic processes as 'paying attention' (2011a, p.59). Mole also argues that the claim that attention is insufficient for consciousness has the advantage of allowing us to

accept that groups such as football teams can pay attention, without having to claim they are conscious (forthcoming a, §1.4). Mole takes this to be an advantage because it coheres with normal folk usage of 'attention'. Mole also considers potential *counterexamples* to his theory drawn from folk psychological discourse (2011a, pp.76-79).

Indeed, Mole explicitly says the following:

> 'I suppose – as an axiom of philosophical methodology – that our theory of attention should, if possible, be a theory of the phenomenon referred to by the English word 'attention'.' (forthcoming a, §1.4).[1]

All of this shows that Mole considers folk psychological coherence to be an advantage of a philosophical theory of attention. The same can be said of Watzl (2011a) who gives certain examples of 'attention' drawn from normal everyday life (such as listening to a jazz band and reading a newspaper) which he uses to motivate his theory. Watzl also argues *against* various theories of attention by arguing that they *contradict* normal folk psychological usage of the word; he criticises Wu in this way (Watzl, 2011a, pp.154-155) and also Mole (Watzl, 2013).

There are various complications associated with these particular theories of attention, not least the fact that the thinkers do not *only* draw upon the folk usage of 'attention' in developing their theories. These complications will be examined later, but what is relevant for present purposes is that these examples show that it is widely assumed that coherence with the folk psychological term 'attention' is an advantage of a philosophical theory of attention. This suggests a strategy for deciding which account of attention in the previous

[1] It is obvious from context that Mole means the *folk* usage of the English word 'attention'.

chapter is to be preferred. The idea would be that we should 'attention', and then this one can be labelled the 'best' account of attention (or at least, the best that we have right now). I will address this suggestion in §3.3. Some thinkers may reject this suggestion, they may say that we should not rely on the folk psychological usage of 'attention' in order to decide which account is best. I will go on to address other ways of assessing which concept is superior in §§3.4-3.5.

§3.3-What the folk say about attention.

My argument concerning the folk use of 'attention' claim will have two main parts. Firstly I shall examine some experimental philosophy which can shed some light on how the word is used in normal discourse. I shall argue that these studies support the claim that people tend to have different intuitions about what does and does not count as 'attention'. The second part of my argument will be a wider examination of the folk psychological usage of the term, where I will argue that the word as used in folk psychology is ambiguous in various important ways. I use these claims to argue that it is unlikely that we can have a definitive answer to the question of which concept of attention is 'better' than another by examining folk psychological usage

§3.3.1-Experimental philosophy

I mentioned in chapter 1 that several thinkers hold the view that the claim that attention is sufficient for consciousness (ST) is part of our normal folk psychological or commonsense view of the mind.[2] As we saw at length in chapter 2, ST has itself been subject to a lot of dispute, but so has the claim that ST is part of commonsense or folk psychology. As we shall see, there is a small debate revolving purely around the question of whether 'the

[2] I take 'folk psychology' and 'commonsense psychology' to be synonymous.

folk' believe ST.[3] In this section, I will *not* address the question of whether the folk believe ST, but some of the experimental results that are part of this debate will be relevant to the questions that this chapter is concerned with.

In this debate, the following imagined case has been used to argue that the folk do not believe ST: suppose a mother is awakened by her crying child.[4] The normal folk psychological intuition when a mother is wakened in such a way is that the mother attends to the cry of the child *before* she is conscious of the cry (because it is attending to the cry that wakes the mother up). Therefore (goes the argument) normal folk psychological intuition is that attention is *not* sufficient for consciousness.

Felipe De Brigard (2010) has performed some experimental studies about this case. In order to establish what intuitions the folk really do have, he presented undergraduates at UNC, Chapel Hill, with a short vignette describing the mother being awakened by the child's scream. Then he devised three experiments to test their intuitions about the case. In each of the three experiments, the participants were given four sentences on a sheet of paper and asked to pick the one that best described the situation of the mother being awakened by the baby. The set of four sentences was different for each experiment. I will (following De Brigard) label these experiments 1, 2 and 3. I will discuss experiment 2 in the next section, but firstly, we can turn to experiment 3. In this experiment, participants were asked to decide which of the following sentences best describes the case of the mother and the child:

i) The mother notices the cry and attends to it.

ii) The mother notices the cry but does not attend to it.

[3] There is also a debate over whether the folk believe NT, which will be briefly examined in §3.3.3.
[4] Mole (2008) first discusses this case, where it is attributed to an anonymous referee.

iii) The mother attends to the cry but doesn't notice it.

iv) None of the above.

The results are striking. 46% of the subjects thought the case was best described in terms of (i), and 33% thought it was best described in terms of (ii). 13% went with (iii) and only 8% went with (iv).

The thing to focus upon for the purposes of my own argument is that the two most popular answers ((i) and (ii)) differ on whether they attribute attention to the mother. What this implies is that in normal folk psychological discourse, people can have different intuitions about what does and does not count as 'attention', and whether attention is or is not operative in a certain case. In this study, very significant proportions of the participants had the intuition that the mother *does* pay attention to the baby's scream, and another very significant proportion of the participants have the intuition that the mother *does not* pay attention to the baby's scream. So this study suggests that different people have different intuitions about the extension of concepts such as 'attention', and thus that the boundaries of the concept of 'attention' in folk psychology is not a sharp one.[5] A reader may have various worries about this experiment. I will address these concerns soon, but first I would like to discuss some more data.

§3.3.2-Comparing Mole and De Brigard.

In the previous section, we saw some data that suggested that different subjects *within one test group* can have different intuitions about whether a certain case is one of 'attention' or not. In this section, I will *compare* the results of another experiment that De Brigard performed with the results of an experiment of Chris Mole's. I shall argue that a

[5] De Brigard's conclusions are similar (2010, pp.199-200).

comparison of the two studies shows that there can be strong differences of opinion over what does and does not count as 'attention' *between* different groups of participants.

In his *second* experiment, De Brigard asked participants which of the following sentences best describes the situation of the mother waking up to the baby's scream:

i*) The cry wakes the mother because she is conscious of it and she is attending.

ii*) The cry wakes the sleeping mother because she is attending but she's not conscious of it.

iii*) The cry wakes the sleeping mother because she's conscious of it, but she's not attending.

iv*) None of the above.

Notice the subtle difference of phrasing from the experiment discussed in §.3.3.1. In this version of the experiment, it was found that 21.2% of participants found (i*) the most natural way to describe the situation. 60.6% found (ii*) the most natural, 12.1% found (iii*) to be the most natural, and only 6.1% opted for (iv*) (2010, pp.193-194). Notice what percentage of the subjects found it most natural to ascribe attention to the mother. Discounting (iv*), we find that both (i*) and (ii*) are responses that ascribe attention to the mother, so in total 81.2% of the participants in De Brigard's experiment 2 (unlike experiment 3) ascribed attention to the mother.

We can contrast the results that De Brigard obtained in experiment 2 with some results which Chris Mole got when he performed a very similar experiment on a group of

individuals.[6] Mole found that 12 out of 19 (63.16%) of responses counted the mother as *not* paying attention to the baby's cry (2008, p.91). In Mole's study, 63.16% of subjects thought the mother was *not* paying attention to the baby's cry, but 81.2% of participants in De Brigard's second study judged that she *was* paying attention to the baby's scream. The difference is striking.

What this at least suggests is that subjects can have different intuitions about what does and does not count as attention, and whether a certain case really *is* a case of attention or not. Differences of opinion can be found *within* one group of participants (as in De Brigard's third experiment, discussed in §3.3.1) as well as by comparing the answers received from one group with the answers received from another group (as when we compare De Brigard's results with Mole's).

If we take the results of these polls seriously (debunking explanations will be considered below) then this tells against the possibility of being able to use folk psychological intuition to decide whether the subjects in the studies discussed in the previous chapter are 'paying attention' or not. Recall that the candidate 'attention' concepts that we encountered in the last chapter all have subtle differences between them (to do with working memory, justification and verbal report), and these differences really start to matter when they are used in certain contexts, such as when we are examining data such as that which pertains to GY. Distinguishing between them by using folk psychology requires a very precise, fine grained and consistent use of the term 'attention' within folk psychology, which is capable of delineating the subtle differences between the different concepts. If the

[6] An important difference is that in the case of Mole's study, participants had no option to say 'none of the above' and they were allowed to tick *multiple* options.

folk have different intuitions even about quite quotidian and well known phenomena as the mother waking to the baby's scream, it seems unlikely that the use of 'attention' within folk psychology will be a precise and consistent enough tool to do this subtle work.

Of course, we should be cautious here. These data do go some way toward showing that intuitions about what is and is not 'attention' in folk psychology can be different for different people, but we should be hesitant before drawing the conclusion that this kind of difference of intuition is widespread throughout folk psychology generally. The reasons for this hesitation are obvious: both studies involved relatively small sample sizes (especially Mole's) and so it is possible that the results discussed above are simply anomalous. Perhaps there really is a widespread and unified usage of the term, but the samples taken here are just the outliers. This kind of caution is certainly advisable: we do not want to draw global conclusions from only a couple of polls.

Nonetheless, these data are the *only* empirical data that we have related this issue, so they should count for something in our considerations here. In the face of these data, it would seem ad hoc and unmotivated to insist that they were only anomalies, and that 'really' there is an accepted and sharply delineated usage of the term that is prevalent throughout folk psychology, which we should marry our concept of attention to. That would seem to run against the *only* empirical data that we have on what folk intuitions about the extension of 'attention' are. Things would be different if we had *good reason* to think that these results were just anomalies, but I know of no such good reason. In the absence of such opposing evidence, we should cautiously accept that we have defeasible reason to think that intuitions over what is and is not 'attention' can differ in folk psychological discourse within a community.

An objector could of course offer an alternative explanation of the data: they could claim that the subjects in question *do not* differ in their intuitions about what is and is not 'attention'. Rather an alternative explanation is that the subjects do not really have *any* strong intuitions when it comes to these unusual cases, and so pick answers randomly when asked. This alternative response has some plausibility to it, but I do not think it is a good route for my opponent to go down. If the (well known) phenomenon of a mother waking up to her baby's cry is a grey area, where subjects do not have strong intuitions either way, then I am doubtful that there would be a clear cut set of intuitions from folk psychology about whether the term 'attention' can apply to what the subjects are doing in the (extremely unusual) experimental paradigms examined in the previous chapter.

This is not to say that there could *never* be any hope of selecting between different 'attention' concepts by examining folk psychological usage. Some candidate meanings of 'attention' may be so ridiculous that they very clearly contravene folk psychological use. My point is only that, when we have a range of different candidate meanings which are only subtly distinct from each other, then (given the studies examined here) the chances of being able to use folk psychology to select between them is slim.

§3.3.3-More experimental philosophy.

Indeed, there are more reasons that can be given to think that different people will have different intuitions about issues regarding attention more generally (especially bearing on its relationship to consciousness) in folk psychology. In addition to the studies discussed above, there are additional studies which show that different people have different intuitions about whether attention is *necessary* for consciousness (Schwitzgebel, 2007 and 2011).

Schwitzgebel devised a study to test whether the folk have the intuition that attention is necessary for consciousness. He began by explaining the issues to his subjects, and polling their intuitions about whether attention is necessary for consciousness.[7] Each participant was then given a buzzer, which would go off randomly. When the buzzer went off, the subjects were asked to note down whether they had the intuition that they had had phenomenal experiences *before* the buzzer went off. They were asked whether they had experience *at all*, whether they could feel the pressure of their shoes on their feet, and whether they had any visual experience in the far right of their visual field (2007, pp.19-24).

As Schwitzgebel is aware, there are a great many problems with using these data to reach conclusions about whether attention *really is* necessary for consciousness. They can however tell us something about what people's *intuitions* are on the issue. Schwitzgebel found that not only is opinion divided roughly equally throughout folk psychology on whether attention is necessary for consciousness (2007, p.23); but that subjects frequently *changed their minds* throughout the experiment (2007, pp.14-24). As Schwitzgebel says:

> 'participants were impressively open-minded and often changed their views and expressed surprise over the course of the experiment' (2007, p.23).

What this suggests is that it is not just that people's intuitions on topics of attention and its relationship to consciousness are different from person to person, but that they are very *fickle* and *malleable.* As I emphasised above, we should treat these studies with caution (the sample may not have been representative of the general population) but it is worth

[7] In order to ensure that his own opinion did not sway them, they were later asked to guess what view he held (at the time he was more inclined toward the view that attention is unnecessary for consciousness). They were unable to guess correctly what his view was.

emphasising again that this is the *only* evidence we have, so it would be *ad hoc* to insist that we ignore it.

Furthermore, when we take a step back and examine some of the experimental philosophy that has been performed on folk intuitions about other folk psychological concepts, the prospects only get bleaker for the claim that folk psychological concepts in general are very sharply delineated, and provide a clear standard against which we can measure our own concepts. One striking study in the experimental philosophy literature tested subjects' intuitions about the folk psychological concept of 'knowledge' (Weinberg, Nichols and Stich, 2001, cf. Knobe, 2012). The experimenters asked participants whether 'Gettier' cases are cases of knowledge or not.[8] Interestingly, the experimenters found that white undergraduates at Rutgers University generally agreed that Gettier cases are not cases of knowledge, whilst the majority of *East Asian* students at the same university had the intuition that they *are* cases of knowledge. The extension of this suggestion to cases such as 'attention' should be tentative, but again it is suggestive. The study does suggest that different cultural background can result in different intuitions about the extension of folk psychological concepts.

To be clear: what I have said falls well short of a *proof* that the folk psychological use of the term is too poorly delineated to help us resolve the difficulties in question, or that there is widespread difference in intuitions about what is and is not 'attention'. Rather, the argument is meant as an inference to the best explanation: once we take account of all of the points that I have made about the folk psychological use of 'attention', we should conclude that it is unlikely that the folk use will be able to help resolve the problems encountered in

[8] See Gettier (1963).

the previous chapter. I think it is fair to say that the onus of proof is now on my opponent to show that there *is* a predominant meaning of the term in folk psychological discourse, *and also that it is sharp, well delineated and unambiguous enough for us to decide which concept of 'attention' in the debates in question is better.*

§3.3.4-What the folk say: more general concerns.

I will now present some different arguments concerning the folk use of 'attention', which are not based around experimental philosophy, but are based more directly upon everyday observation of the folk psychological usage of the term 'attention' in natural language. I argue that the term 'attention' as it is used in folk psychological discourse does not so much refer to one 'core' entity or process in particular, but that the predicate can be used in different ways to convey different meanings in different scenarios, and understanding such utterances relies heavily on conversational pragmatics. If what I say in this section is plausible, then there *is no* one 'core' use of the folk psychological concept of 'attention', but rather 'attention' serves as a placeholder that can serve many different roles in different aspects of folk psychological discourse.

We can begin by noting that it is plausible from normal everyday conversation that the predicate 'attention' is used in a heavily context dependent way, and can be used by different speakers at different times in different ways. Consider a case where one is reading a novel on a plane. It may be that one naturally describes oneself as 'paying attention' to the novel, when one interprets 'attention' to mean something with very minor epistemological requirements. For example, perhaps when one naturally describes oneself as 'paying attention' to the novel, perhaps one simply means that they were reading it as opposed to listening to the safety announcement. In this case, it may well be natural to say that you

were 'paying attention' to the novel, but suppose someone sitting next to you quizzed you about the novel and asked you if you were 'really' paying attention to the novel. In this case, you may interpret 'attention' as something which designates much higher epistemological demands. For example, you may now come to interpret 'paying attention' as keeping in mind characters' backstories, noticing subtle plot devices, thinking about the social themes of the novel or whatever. You may then change the ascription of 'attention' to oneself and admit that one was *not* paying attention to the novel, in this sense. This kind of shift in ascriptions of attention is commonplace throughout folk psychological discourse.

Often the way that the term 'attention' is used in folk psychological discourse is far more complex than this example shows. In what follows, I will give some examples of cases where the term 'attention' is used to convey very complex and abstract meanings, where the term does *not* function to refer to a mental state or process. If this is correct, then it is a mistake to attempt to use the folk psychological use of the term to adjudicate between the rival definitions of the previous chapter, as a great many folk psychological uses of the term do not even function to refer to a mental faculty or process *at all.*

In normal folk psychological discourse, it makes sense to talk of *group attention*. These are instances where 'attention' is predicated of *groups* of people, rather than single subjects.[9] Consider the claim that 'the Liberal Democrats need to pay more attention to the needs of the students, if they are to have any hope of beating UKIP'. In this case we see that the folk psychological term 'attention' can be used to apply to whole groups of people, not single subjects. These uses of the word 'attention' are commonplace. We may talk of a

[9] Ratcliffe (2007 and 2008) and Needham (1972) discuss the similar phenomenon of 'group beliefs' (esp Ratcliffe, 2007, p.207).

particular rugby team ‘paying attention’ to the opposing team’s flankers. It also makes perfect sense to say something like ‘the disaster at Isandlwana made the British army pay more attention to the Zulus’ battle tactics’. The ascription of ‘attention’ to groups, organisations and whole nations is commonplace in folk psychological usage of the term.

It is important to notice that in the cases of the claims about the Liberal Democrats and the British army, subjects are not using the term ‘attention’ to refer to some cognitive process or faculty of mental focalisation, but rather they are using the term to convey some wider, more abstract meaning. In the first case they are expressing something about the state of politics, and the way that certain political parties interact with each other. In the latter case they are saying something about the interaction of the British and the Zulu armies, and perhaps making a comment about the arrogance of Imperialism.

When subjects use expressions like this, they rely on a variety of factors to aid their audience in understanding their intended meaning. In the case of the comment about the Liberal Democrats, such factors will typically include the context in which the utterance is made; background beliefs about what the Liberal Democrats have done, and how that is viewed by the student population. Perhaps the audience will also have to make certain assumptions about the attitudes and viewpoint of the speaker who makes the utterance (most likely the speaker will be making such assumptions about the audience as well). It is only within the net of all of these factors that the speaker’s intended meaning will be correctly conveyed to the listener.

There are many more vivid examples of ‘attention’ being used in this way. Suppose that in the last stages of a dying relationship someone says to their partner, ‘you just don’t

pay attention to me any more'.[10] In this case, the speaker does not seem to be predicating any one mental state or process of her partner, rather she is making an extremely complex claim about their relationship together, its history, the way that they interact with each other and so on. Indeed, she is not only offering a description of the situation, but also expressing her own attitude towards it, and effectively making claims about the amount of effort she is willing to put into the relationship in future.

These are all ubiquitous features of the folk psychological use of the word 'attention'. When a sleep deprived new parent says 'I just can't believe how much attention young children require' he is not talking about how much he 'highlights' his son in experience, or saying that information about his child is easily accessible to his reasoning systems or anything like that. Rather he is making a more general claim about the physical and mental strain that he is suffering in order to look after his child, and making more general reports about the turn that his life has taken since having a child.

One of the most common uses of the term 'attention' in folk psychological discourse (at least in British English) is when we describe someone as an 'attention-seeker'. When we make such an utterance we do not mean that the person in question likes information about himself to be available to other people's working memory systems, or that he likes to have information about himself prioritised by others' visual systems. Rather, what we mean is that he takes active steps to occupy certain positions in social situations, that he is self-obsessed, that he will cut others off mid-speech and perhaps that he is likely to exaggerate

[10] This example is modified from one given in Ratcliffe (2007, ch.7) when he discusses *beliefs*. The example is 'I just don't believe you, I've heard it all before'.

certain stories about himself. Perhaps we are also making a comment about the way that he has been overindulged by those around him.

A further complication is that in folk psychological discourse, ascriptions of attention are often intimately entwined with *normative* claims. When a teacher tells a student 'you're not paying attention!' or when she says to his parents 'he just refuses to pay attention in lessons' she is not (only) making some kind of ascription about the mental states of the student, she is also rebuking him. What she is saying has normative consequences. She is not only describing how things are, but she is also expressing what he *should* be doing, and how she wants him to change his behaviour. Maybe she is also expressing the thought that his parents should be worried, or should perform some kind of activity to change states of affairs in certain ways. Similarly, when people say 'Barack Obama is just not paying attention to the needs of single mothers', it seems clear from the quotation that the speaker thinks that the President *should* be paying more attention to the needs of single mothers. What we see here is that 'attention' is (at least in some contexts) a heavily *normative* concept when employed in folk psychological discourse.

Nothing I have said here should be very surprising. We competently use the word 'attention' in normal folk psychological discourse frequently, and we easily grasp the meanings of different utterances that make use of the term, even though the term can be used to communicate a wide variety of different meanings, dependent upon a great many different factors.

The different ways that the folk psychological concept of 'attention' is used make it apt for us to slide from one meaning of the term to another, even in philosophical discourse. Consider the view of Nicholas Bommarito (2013), who argues that attention can be

employed in ethical philosophy to explain the virtue of modesty. Bommarito says the following:

> 'Modesty involves certain patterns of conscious attention, which are characterised by an *inattentiveness* to good qualities that reflect well on oneself…. [m]odest patterns of attention also involve a positive attentiveness to the role of external causes and conditions in producing the good qualities. These patterns of attention must happen for the right reasons' (2013, p.111).

When Bommarito is establishing what he means by 'attention' he uses the following example:

> 'Emma is a philosopher walking through the park and thinking very hard about the main argument of a paper she is writing…. she walks around a tree that was in her path. She was not ignorant of the tree. After all, she managed to avoid it. At the same time, she did not pay attention to it either; her attention was entirely on the paper' (2013, p.99).

With the example of Emma the philosopher, Bommarito's use of the word 'attention' here seems to be a kind of attention that is involved in *thought* (Emma is concentrating on an argument she is constructing) and Bommarito also links the kind of attention he is discussing with *perceptual* attention, by way of using a contrast class (Emma was *not* paying attention to the tree). So Bommarito's use of the word in this quotation is the kind of attention that is involved in attentive thought and perception.

However, when Bommarito first uses the concept of attention in order to establish that attention can account for modesty, using the example of a great architect (whom he calls David), he says this:

> 'Though he attends to his skills and their importance, he more often considers how fortunate he has been to have the opportunities that led him to where he is. "Sure, I've walked a lot of roads," he muses, "but nearly all of those roads were already paved when I got there." He often reflects on how lucky he has been to have a supportive and stable family, to be born in a place with access to good education, and how patient and encouraging his teachers have been' (2013, p.102).

The difference between these two quotations is striking. When attention was introduced, we were given the example of Emma paying attention in thought to an argument she is constructing, and we had perceptual inattention as a contrast class, but when 'attention' is first put to work in the theory, it seems like what is happening here is that David is expressing an extremely complex set of mental states, opinions, and reflections on the structure and development of his life.

I will not criticise Bommarito's actual proposal here, rather I introduce it only as an example of how easy it is to slide between uses of the term 'attention' which seem substantively different. It is small wonder that this occurs, if the folk psychological usage of the term is as wide ranging as I have been arguing.

Indeed, it is plausible that context dependency and conversational pragmatics are a general feature throughout folk psychological usage, not just in regard to 'attention' but with a great many folk psychological terms. After all, it is the context dependency of the folk psychological term 'knowledge' which fuels contextualism in epistemology (e.g. Lewis, 1996). It is well known that people will be less likely to ascribe knowledge to someone in contexts such as those involved in discussions of external world scepticism, whilst people are far more likely to ascribe knowledge to people in normal situations as when one says 'I know that this river flows past the church'. Folk psychological concepts are often used in

different and context dependent ways to communicate different meanings depending upon the scenario.

Everything I have been saying above spells trouble for the idea that we should marry our concept of 'attention' in the debates over ST and NT to the one that best coheres with folk psychological use. This is because there is no *one* folk psychological use of the term, against which we can measure our concept of attention, but rather the term is ambiguous between a range of different meanings and different uses, which we are heavily dependent upon conversational pragmatics to disentangle. Indeed, if what I have said is correct then often the use of the term in folk psychology does not even function to refer to a mental state or process. The use of the folk psychological term is far wider and less neat than that.

§3.3.5-An objection and a reply.

Perhaps an objector will accept much of what I have said about the folk psychological usage of the term but say that there is still some 'core' usage of the term within folk psychology that we should ally our account of attention to. The claim may be that there are all sorts of different folk psychological uses of the term, but that some of them are *literal* and some of them merely *metaphorical*, and we should only take account of the literal ones when we are deciding how to tell between different proposed meanings of 'attention'.

For example, it is commonplace for someone to say something like 'I can't believe that she's been spreading such lies about me!' In these sentences, subjects use the word 'belief' to communicate a certain meaning (again, reliant upon certain contextual factors). However (says my opponent) we should not take these statements to express *literal* meanings of the word 'belief', but only metaphorical ones. It may be said that many of the

uses that I have elucidated are in fact just such metaphorical uses. Subjects do not *really* mean that the Liberal Democrats need to pay attention to students, they are speaking metaphorically. The objector may then claim that when we attempt to develop a concept of 'attention' that can do the work that we need it to do in these debates, we should use one that only uses the literal uses of 'attention' within folk psychology. Of course, this will not directly address the points I raised about the experimental results from Mole and De Brigard, but it may go some way towards addressing what I have said about the different roles that the folk psychological predicate plays in §3.3.4.

I think this is a reasonable point, but I have several reservations about it. The first reservation I have is that I do not think that the uses listed above really *are* metaphorical uses of the term. As I said, the kinds of uses that I have given examples of are not outliers or exceptions to the rule, rather they are among the most common ways that the term 'attention' is used in folk psychological discourse. I think it is quite a stretch to say that the term 'attention-seeker' is metaphorical. Certainly, to call someone an 'attention-seeker' would not strike a normal listener as an indirect use of language: rather it is a standard and pervasive use of the term in folk psychology. This point applies *mutatis mutandis* to the other examples given above: given their prevalence and the natural way that they are used in folk psychology, to say that they are *all* metaphorical strains credulity.

An objector may say that my claims here rely on a particular view of what does and does not count as a metaphorical use of the term 'attention', and that my intuitions on these issues might not be widely shared. In response to this, I say that of course we would expect intuitions to vary over the question of whether a certain use of the term is metaphorical. But this fact itself is problematic for my opponent. Precisely *because* intuitions over whether a

certain term is metaphorical will be divided, we should expect the boundary between what is the 'core' (or literal) meaning of the term and the metaphorical use of the term to be a fuzzy one. This means that the 'core' usage that we end up with will correspondingly be fuzzy. Once again, the fact that it is fuzzy makes it unlikely that we will be able to use it to make the extremely fine distinctions required to resolve the problems of the previous section.

Another worry is that one cannot hope to divide up the literal and metaphorical folk psychological uses of 'attention' without begging important questions. So, consider the phenomenon of 'group' attention examined above. It may seem that this is a good candidate of a use of the folk psychological term 'attention' which is non-literal. The problem is that some thinkers in the debate treat group attention as a real and literal kind of attention. Mole (2011a, pp.166-167) is perfectly happy to allow that groups can pay attention to things. He uses the example of Tottenham Hotspur Football club defeating Wigan. Mole claims that understanding of how to perform tasks (such as winning at football) can be ascribed to group agents, and that as a result this task can be performed more or less attentively. Mole also takes claims such as 'she started attending to her career I her twenties' as a real and literal kind of attention (forthcoming a, §1.4).[11]

So, we cannot insist that group attention is not a literal use of 'attention' unless we insist that Mole is wrong to treat it as a genuine and literal kind of attention. But what grounds can we have to claim that Mole is wrong to say that this kind of attention is *literally* a kind of attention, other than mere insistence? We cannot dismiss Mole's view that this kind of attention is literally a kind of attention without some criteria that establish what is

[11] Mole insists that the case of a soldier 'standing to attention' does not literally count as a variety of attention. He does not argue for this claim, however (forthcoming, §1.4).

and is not a literal usage of the term, which is exactly what we wanted to establish by examining folk psychology in the first place.

§3.3.6-Folk psychology: Conclusions.

I have argued that the concept 'attention' in folk psychology is used in different ways by different people, such that people simply have different intuitions about what the extension of the concept is. This gives us reason to be sceptical of the strategy of deciding which account of attention is 'better' by consulting folk psychological usage, because folk psychological usage is most likely not delineated in a fine-grained enough way to perform this task. I have also argued that the term in folk psychology is often used to communicate very abstract meanings (dependent upon conversational pragmatics and context) rather than to refer to one mental faculty or process.

An important point: I do not mean any of this in any way as a critique of folk psychology (cf. Churchland, 1981). My main conclusion is simply that it is a mistake to look to folk psychology to guide our definition of attention because the folk use of the psychological predicate 'attention' is not suited to fulfil this role. Whatever function the folk psychological term has, *this* one is simply not it.

§3.4-Empirical Psychology.

In §3.3, I gave several examples of cases where some of the main thinkers in the debate use the folk psychological use of 'attention' to motivate their accounts of attention. The most natural conclusion to draw from reflection on this method is that the thinkers are trying to develop an account of the *folk* usage of 'attention'. However, things are not so simple. Many of the thinkers in the field *also* draw on data from empirical psychology and neuroscience, implying that they intend their use of 'attention' to cohere with the use of the

term in this field (Mole, 2011b, p.75);[12] Watzl (e.g. 2011a, pp.159-161); Wu (2014, esp. chs.1-2) and Prinz (e.g. 2011, pp.182-187). Typically, thinkers do not explicitly state whether they intend their theories as a theory of the *folk psychological* concept of attention, or whether they take the term 'attention' in empirical psychology as their starting point, or whether they intend their theory as a theory of *both.*

Some of the difficulties to do with lack of clarity in this respect will be examined in the next chapter. For now however, the use of empirical psychology suggests another way of deciding which account of 'attention' is 'better'. The idea would be that we should look at how the word 'attention' is used in *empirical* psychology and neuroscience, rather than folk psychology, and then see which usage in the debates best coheres with this usage. The use that best coheres with what empirical psychologists and neuroscientists mean is the account that we should advocate.

This response contains an empirical commitment, which is that the concept of 'attention' is well delineated in empirical psychology and neuroscience, and that there is a consensus in the way that the term should be used across empirical psychology. Let us see whether that is true. I shall argue that it is not. I shall argue that the use of the term in empirical psychology is insufficiently sharp to settle the disputes over ST, and that a wider analysis of the use of the term in empirical psychology reveals that it is used in a great many different ways in empirical psychology, just as was the case in folk psychology.

§3.4.1-Back to the debates over ST.

Recall that one of the problems in the previous chapter was that we were unsure whether or not to accept that the subjects (such as GY) in the studies from the previous

[12] Though Mole also *disagrees* with how many psychologists have viewed attention (e.g. 2011b and 2012).

chapter were attending to the relevant stimuli or not. The supporters of ST claim that they are *not,* and their opponents claim they *are.* What we would need to solve these problems in the suggested way is for the use of the term 'attention' in empirical psychology and neuroscience to be sharp enough to help us decide one way or the other. However, what find when we look at the wider empirical literature is that many of the accounts of attention on the table would count the subjects as attending, and many of them would count them as *not* attending.

To see this, consider the following comments about attention from neuropsychologist Victor Lamme:

> "attention… determines whether a (conscious) report about a stimulus is possible. Likewise, attention determines whether items are stored in a sufficiently stable manner (working memory) to allow report at a later time or to allow a comparison" (2003, p.13)
>
> And this one:
>
> "it seems like attention guards the gate towards a conscious representation that can be consciously reported or remembered." (2004, p.862).

In these quotations, Lamme explicitly allies attention with 'conscious report', just as De Brigard and Prinz do. Since the subjects cannot consciously report the stimuli in question, it will follow from Lamme's definition of attention that they were not attending to them. It is interesting to note that Lamme distinguishes what he calls 'attention' (which is similar to Prinz's definition of attention) from what he calls 'non-attentional selection systems', which he claims do not count as 'attention' because they are not 'under voluntary control' (2003, p.12).

Lamme is not alone amongst empirical psychologists in this regard. Bernard Baars (1997) similarly claims that attention is a subcomponent of working memory. Baars says this: 'selective attention can be defined as selection among potential conscious contents' (2002, p.50). In this quotation, Baars is explicitly defining attention as something that acts upon 'potential conscious contents'. He does not explain what 'potential' might mean, but it seems unlikely that such a definition could count subjects like GY as paying attention to the stimuli in question, as it seems hard to see how we could say that the stimuli in question were 'potentially conscious'.

Baars' definition delivers the verdict that the subjects were not paying attention to the stimuli in question, as do Prinz's and Lamme's definitions. Unlike Prinz's definition, Baars' and Lamme's definitions do this by explicitly defining attention in terms of consciousness, thus establishing the link between consciousness and attention by stipulative fiat. We can find more definitions of attention in the literature that do the same, and thus lead us to the conclusion that the subjects were *not* paying attention to the stimuli.[13] For example, Huang and Pashler say this:

> 'Visual attention, in its most fundamental sense, is a selective visual process that governs access to consciousness' (2007, p.599).

So Prinz, Lamme, Baars and Huang and Pashler use accounts of attention which would deliver the verdict that GY is *not* attending to the stimuli in question.

However, compare these definitions of attention to the following:

> "'Attention' refers to the cognitive mechanism that allows certain information to be more thoroughly processed in the cortex than non-selected information." (Cohen et al. 2012, p.411).

[13] As I pointed out in §1.4.2, William James' definition does this.

In contrast to the definitions of attention given above, Cohen et al. put no constraints of 'conscious reportability' upon their definition of attention, and as such their definition of attention leaves open the possibility that GY *is* attending to the stimulus in question (which is precisely the conclusion they draw).

Cohen et al. have introduced an understanding of attention which is not very demanding in contrast to those who define it in terms of consciousness, reportability or working memory. Recall from chapter 2 that Kentridge does the same. Several empirical psychologists are in agreement with Kentridge and Cohen et al. in defining attention in a non-demanding manner. Consider Katherine Armstrong's definition:

> '*Spatial* attention can be operationally defined as the enhanced processing of visual signals at a particular location in space' (2011, p.79).

On Armstrong's definition, then, anything which results in 'enhanced processing' of 'visual signals' can count as attention. It seems likely that the subjects' activity counts as enhanced processing of visual signals (after all, their response times and/or accuracy rates are improved in valid conditions). So by Armstrong's definition of attention, the subjects *are* attending to the stimuli in question.

In summary, the whole problem that we started with was that different interlocutors disagreed on the nature of attention to such an extent as to give different answers to the question of whether the subjects in the experiments described in the previous chapter were attending to the stimuli or not. The response currently under scrutiny was that we should decide this issue by deciding which definition of 'attention' most closely mirrors the one used by the experts. However, the experts' definitions themselves seem to differ on whether or not they entail that the subjects were attending to the relevant stimuli.

§3.4.2-Triviality.

Above I mentioned that some psychologists define attention in such a way as to give the conclusion that GY *was* attending to the stimuli in question, whilst others defined it in a way that entailed that he was *not*. There is a related problem here as well, which is that on many accounts of attention, the claims in question are likely to become entirely trivial. Take for example, Armstrong's definition as enhanced processing of visual signals at a particular location in space, or one of Kentridge's own definitions of attention that I mentioned in §2.2.4 as 'the voluntary or involuntary prioritization of information in a selected part of a visual scene' (Norman et al. 2013, p.836). The trouble is that if we take these permissive definitions seriously, then the claim that attention is insufficient for consciousness is likely to become entirely trivial, as it is well known that there are many mechanisms of prioritization and selective enhancement in the visual system, not all of which give rise to consciousness.

Alan Allport summarises this situation well:

> '[f]or some authors, practically any kind of modulation (enhancement, suppression, etc.) of neuronal response is considered to be the expression of attention' (2011, pp.25-26).

If we take seriously the permissive accounts of attention given just now, which would count enhanced processing of visual information, or indeed any 'prioritization' of information of any kind to count as attention, then almost *any* of the selection and enhancement mechanisms in the visual system will count as attention, and since we can be reasonably sure that not all of them give rise to consciousness,[14] the claim that attention is insufficient for consciousness will become true, but trivial.

[14] I suppose that panpsychists (e.g. Strawson, 2006) would say that all such processes give rise to consciousness. However, Strawson's claim is not specifically about attention, but about all of matter,

We need not delve too deeply into the empirical literature in order to see this. Take, for example, the fovea. The fovea is an area of the retina which contains a very densely packed collection of rods and cones, allowing for accurate and detailed information about the objects that fall within the visual angle of the fovea to be gathered by the retina and transmitted along the optic nerve into the visual system. Objects that fall within the fovea's visual angle are said to be *foveated.*

When objects are foveated, the eye receives much more information about those objects. So, if we define attention simply in terms of prioritization of certain kinds of information in the visual field, then the eye will count as instantiating attention. But it would be peculiar to hold that the process of foveation is sufficient for phenomenal consciousness. So it is hard to see how, on these accounts of attention, the claim that attention is insufficient for consciousness is anything more than trivially true.

The issue is that if we take a more permissive account of attention which allows GY to count as attending to the relevant stimuli, we may end up making the claim that we wanted to uphold (that attention is insufficient for consciousness) hopelessly trivial, and thus lose any grip on why the GY experiments are of independent importance or interest.

§3.4.3-Empirical psychology and pluralism.

There is another important point to make here regarding how attention is understood in the empirical literature. Recall that one reason that we began our examination of the empirical literature was because we wanted to uncover a common usage against which we could decide which account of attention is 'better' than the others in various debates, by comparing it with various uses in the empirical literature.

---

and it is clear that his concerns are radically different from the concerns of those in this debate. I shall discuss panpsychism no more here.

However, this is a risky strategy for my opponents to engage in from the beginning, because there are many empirical psychologists who would reject the claim that there is one 'correct' definition or understanding of attention in the first place. Many psychologists endorse something like pluralism. Consider the following quotation from Elizabeth Styles:

> 'attention is not a single concept, but the name for a variety of psychological phenomena' (Styles, 1997, p.1).

Finally, in her review of the last 25 years of research on visual attention, Marisa Carrasco makes the following comment:

> 'The field has developed a consensus that attention is not a unitary construct' (2011, p.1517).

As will become obvious over the course of this thesis, my own view is in sympathy with that of Styles and Carrasco. However, I am *not* saying that we should accept these views simply because some of 'the experts' hold them. All I am saying is that my opponent is ill advised to defer to the experts in order to assess which account of attention is 'better' because many of the experts would agree with me that to ask which account of attention is 'better' is somehow a bad question. I suggest, then, that we have good reason to accept that the use of the word in empirical psychology is unable to allow us to settle which concept is 'better' than the others in the debates over ST and NT. In §3.7

§3.4.4-An objection and a reply.

My opponent may at this point insist that all I have been doing is cherrypicking the uses of the word 'attention' that most suit my case. Perhaps really there is some core usage in empirical psychology that most scientists can agree on, which is sharp enough for us to be

able to decide which concept of 'attention' is better in the debates over ST and NT. Perhaps I have just been picking the outliers and the voices of dissent to make my case.

Ultimately, I have to confess I find it difficult to see how I could respond to an objection like this. It is as clear to me as anything can be in philosophy that the last thing we want to do is start to argue over who does and does not count as an 'expert' in empirical psychology. I do not think that arguing over who has authority to tell us how we should and shouldn't use the concept 'attention' and whose opinion we do and do not want to take on board will be very philosophically fruitful. I will say only that I have demonstrated that there *is* widespread difference in how the concept is used throughout empirical psychology and folk psychology, and if there is something wrong with my examples, I will need to be told *what it is.*

An important point to make is that if my opponent wishes to resist pluralism, then of course it will not be enough to show that there are a certain small number of core uses of the word in empirical psychology (though I doubt that even this is the case). Rather, in order for the opponent to resist pluralism about attention they must show that there is only *one* core usage of the term. Such an attempt has, I think, slim prospects of success. My opponent may be unimpressed by me shifting the burden of proof at this stage, but we should remember the place in the dialectic that this objection comes in. I have been arguing that the word is used in different ways, and that there are importantly different concepts of 'attention' in the empirical literature. My opponent's objection at this point amounts to simply insisting that this is not the case, and that 'really' there is some core use here. So there is very little I can say other than repeat my argument. If my opponent denies this, then I think it is fair to

demand a counterargument, or more in depth analysis, or something similar which can show *why* (given everything that I have said) we should accept her view rather than mine.

§3.5-Theoretical Use.

I began this chapter by noting that several of the thinkers in these debates seem to consider coherence with the folk psychological use of 'attention' a virtue of a theory of attention. As we progressed, I mentioned that these thinkers also include discussion of the *empirical* work on attention, and this implies that they intend their usage of the term to cohere with its use in empirical psychology. We can now add another layer of complication here, as many of the thinkers *also* attempt to put their concepts of 'attention' *to theoretical use.* Wu at one point explicitly says of his own theory of attention:

> 'My goal is to identify a theoretical conception that is psychologically and philosophically useful.' (2011a, p.97).

Similarly, Christopher Mole puts his concept of 'attention' to work in several debates in philosophy regarding reference and consciousness (2011a, ch.7).[15] This seems to imply that (these thinkers at least) are not *only* aiming to capture the folk psychological use of attention, or its use in the empirical sciences; but are also attempting to give an analysis of some theoretically useful concept of 'attention'. This suggests a third way of deciding which definition of 'attention' is 'better', which is to select the concept of attention which is the most *theoretically useful.*[16]

[15] Though Mole is in general pessimistic about the prospects of explaining reference or consciousness in terms of attention.

[16] Mette Hansen is among those who have suggested this to me.

Notice that this suggestion departs significantly in spirit from the two suggestions already examined. The two previous suggestions take as their assumption that we should ally our account of 'attention' with how the term is currently used in some community. The current suggestion is more revisionary in spirit, because the claim here is not that our preferred concept of 'attention' should *mirror* current usage, but should *improve* on it.

What might it mean for a concept to be 'theoretically useful'? This is a complex question, subject to several different interpretations. I will not be addressing *all* of these issues in this chapter, because we will be returning to the idea of a 'theoretically useful' concept of attention over the course of the thesis. However, at this point it will be useful to provide a map of the terrain and explain where I discuss each particular issue.

In this section, I will provide an argument based on inference to the best explanation for the claim that we should not expect to be able to decide which definition of 'attention' is 'better' based on theoretical use. This argument will be quite general. I will return to the issues with greater specificity in chapter 4 (§4.6), where I shall pinpoint *several specific different* concepts of 'attention' in the literature and give a detailed argument for the claim that each one is theoretically useful, but for different purposes. I also give a thorough analysis of the different roles that they are useful for. In chapter 4 I also aim to explain *why* they are useful for these purposes *specifically.*

Importantly, there is a very specific version of the claim that we should decide which concept of attention is the most 'theoretically useful', which also invokes metaphysical considerations. This is the claim that we should pick the concept which refers to a *natural kind.* The mere mention of natural kinds raises a host of complications and complexities, which will require a chapter of their own to pick over. This is chapter 6. Finally in chapter 7,

I will connect many of the issues to do with the theoretical use of certain attention concepts with some of the wider literature on classification and usefulness in the philosophy of science literature.

§3.5.1-Plethora of theoretical uses.

The idea that we should select which concept of attention is theoretically useful immediately raises the obvious question: useful for *what*? The suggestion that we should select which account of attention is the most theoretically useful will only bear fruit if we can decide *what we want our concept of attention to do.* However, when we look at the sheer breadth of different theoretical roles that 'attention' has been expected to fulfil, we find a wide range, with no obvious unifying characteristics.[17] As I pointed out extensively in chapters 1 and 2, attention has been used by many thinkers to explain phenomenal consciousness (Prinz, 2012; Rosenthal, 1997 and Lycan 1987 and 2004). It has been used to explain action (Wu, 2008, 2011a, 2011b and 2014) and to account for the structure of our phenomenology (Watzl, 2011a). The concept has been used to criticise various theories of consciousness such as global workspace theory (Block 2007a, 2007b, 2011, 2012 and 2014b). Attention has also been used to criticise certain kinds of representationalism about phenomenal consciousness (Block, 2010 and Wu, 2010). It has been used to argue in favour of certain views about the phenomenology of peripheral visual experiences (Block, 2013a, 2013b and 2014a), and has been linked to memory retrieval (De Brigard, 2012). It has been used to discuss what properties can be represented in consciousness (Prinz, 2013a); has been deployed in debates about the unity of consciousness (Prinz, 2013b); as well as debates centring on the nature of cognitive penetration (Mole, forthcoming b).

[17] Mole refers to this as the 'overburdening' problem (2011b).

The use of the concept of attention has also featured in philosophy of language and epistemology, where it has been invoked to explain our ability to make demonstrative reference and to have *knowledge* of demonstrative reference (e.g. Campbell, 2002, 2004, 2011; Levine, 2010; Matthen, 2006; Clark, 2006 and Raftopoulos, 2009). Descartes used attention in an attempt to eradicate doubt about what he called 'clear and distinct ideas' (1988, p.309). The concept has been used to draw links between epistemology and consciousness (Smithies, 2011a, 2011b and Siegel and Silins, forthcoming; and Silins and Siegel, forthcoming). The concept of 'attention' has been used further in epistemology to account for how we can have certainty within a contextualist framework (Lewis, 1996).

The concept has also been assigned work elsewhere in philosophy. In metaphysics, attention has been used to explain how certain properties could be *both* dispositional and categorical (Heil, 2003, ch.11 and Martin, 1993), and has further been used in metaphysics to defend a trope ontology (Gibb, 2012). As we have seen, it has also been employed in ethical philosophy to explain modesty (Bommarito, 2013). Iris Murdoch uses the concept of 'attention' in her explanation of what 'the good' is (1970, ch.1).

And that was just from philosophy! In empirical psychology, the number of different roles that 'attention' has been assigned is at least as dizzying. A particularly clear example of the amount of work that 'attention' has been assigned comes from Ronald Rensink's (2013 and 2014) discussion of the relationship of attention to visual consciousness. Rensink taxonomises attention into five distinct *varieties* of 'attention', which are assigned different roles within the visual system. He labels these subcomponents 'sampling', 'filtering', 'binding', 'holding' and 'indexing'.

'Sampling' is the variety of attention which is responsible for collection of information by the eye for transmission along the optic nerve (Rensink, 2013, §0.1). Sampling dictates what information gets into the visual system in the first place. This is the variety of attention that I have discussed above, and which would render the conclusion that attention is sufficient for consciousness trivially true. Filtering, by contrast, is the process which *selects* from information that has entered the visual system based on what properties or locations the subject is actively searching for (Rensink, 2014, §3.1).

The remaining subvarieties of attention that Rensink discusses have strikingly little in common with the two just illustrated. He labels the third variety of attention 'binding', which is the process by which different properties are understood as belonging to one object (as when we see a ball as both red *and* spherical).[18] The fourth ('holding') Rensink uses to explain how we can track the identity of an object through a change in its superficial properties (2014, §3.3). The final variety of attention, which Rensink labels 'indexing' is the process by which we individuate a physical object and see it is separate from other objects in the environment.[19] Furthermore, each one of these five jobs subdivides into several further sub-roles, which further varieties of attention take up the job of performing (Rensink, 2013 and 2014).

It is obvious that Rensink assigns 'attention' a huge number of major jobs in the visual system, from selecting the information that is allowed into the brain via the optic nerve, to the construction of a representation of an external world object which can be tracked through changes in its properties and individuated from the environment around it.

[18] Here Rensink is clearly heavily inspired by Anne Treisman's 'feature binding' theory of attention (Treisman and Gelade, 1980 and Treisman, 1988 and 1996).
[19] Rensink uses the term 'indexing' in his (2013) and 'individuating' in his (2014).

Indeed, it is not too much of an exaggeration to say that Rensink gives almost *every* major job in the visual system to some variety of attention.

Notice not only that 'attention' has been assigned many different jobs in the visual system by Rensink, but that the jobs seem extremely different from each other, and take place at different stages in perceptual processing. Two varieties of attention that Rensink discusses seem to be involved in the *selection* of information for consumption by the visual system, whilst the other three seem primarily to be concerned with using this information to *construct* a representation of the outside world.

Rensink is not alone within contemporary psychology in assigning attention a lot of diverse roles which lack any apparent unifying characteristics. Posner and Rothbart divide attention into three main components, labelled 'alerting', 'orienting' and 'executive attention'. They summarise them thus:

> 'Alerting is defined as achieving and maintaining a state of high sensitivity to incoming stimuli; orienting is the selection of information from sensory input; and executive attention involves mechanisms for monitoring and resolving conflicts among thoughts, feelings and responses' (2007, p.7).

Here (as with Rensink) we see several significantly different functions assigned to attention. It is interesting to note how different Posner and Rothbart's view is from Rensink's: both seem to assign attention various jobs, but each assigns it *different* ones.

The concept of attention has spread throughout perceptual psychology, but its use is certainly not restricted to this, it is also used in developmental psychology to account for knowledge of other minds (Reddy, 2010. Cf. Eilan et al. 2005 for philosophical perspectives on this work). Above I mentioned that attention has been used to attack global workspace

theory, but it is certainly not only an enemy of the view; it is used in various ways by those within psychology and philosophy who *support* global workspace theory (Dehaene, 2014; Dehaene and Naccache, 2001, Dehaene et al, 2006 and Dennett, 2001b). We will be discussing the use of attention in global workspace theory in the next chapter. Chun et al. neatly summarises the situation of attention research in contemporary psychology:

> 'The concept of attention has now permeated most aspects of perception and cognition research' (Chun et al. 2011, p.74).

What we see from this brief analysis of the philosophical and psychological literature on attention is that attention has been put to work in a great many different theoretical roles, and has been expected to serve a great many different purposes throughout both disciplines. Given the sheer range of different theoretical roles that 'attention' has been put to, do we really expect *one* concept of attention to be the most useful for all of them? Indeed, what reason have we even to think that a *small number* of 'attention' concepts could be useful for all of these roles? Do we really expect that the *same* concept of attention can help us to explain how we can track the identity of an object across a change in its visible properties *and* explain the virtue of modesty? Can we explain all of these things using the same concept that we employ when testing whether GY has increased task performance in relation to certain stimuli that have fallen in his blind field, or when asking whether phenomenal consciousness is a necessary precondition of knowledge of demonstrative reference?

What this taps into is that we don't really have a clear idea exactly *what we want our concept of attention to do.* If there were one particular well understood theoretical role that we wanted our 'attention' concept to fulfil, then things would be simpler. We could state the desiderata of our theory of attention, and then we could ask which concept(s) we need in

order to perform this work. We do not, however, even know where to begin. In my view, there is a wedge to be driven here between 'attention' and another concept where similar issues apply, which is the concept of 'species'. 'Species' has been defined in importantly different ways throughout the history of biology, and the acceptance of several different species concepts which are useful for different purposes is a common view, in biology and philosophy of biology alike.[20] Philip Kitcher (1984) is one such pluralist about the concept of 'species', but he offers criteria by which we can judge the theoretical use of the 'species' concept. He suggests that any useful 'species' concept must *not* be 'redundant', 'boring' or 'wrongheaded' for the purposes of evolutionary biology. With these criteria in hand, Kitcher is able to accept that several species concepts fulfil the criteria in question, but reject other species concepts as failing by this criterion (such as the creationist one or the phenetic one).[21]

The relevance of this to my own argument is this: in the species case, there is plausibly a set of criteria that can be given (derivative on the interests of evolutionary biology) by which we can assess the theoretical use of a proposed species concept. The problem with 'attention' is more fundamental than this: it has been used so much in empirical psychology and in so many areas of philosophy that we don't even know how to *begin assessing* the theoretical use of a proposed 'attention' concept.

I am perfectly willing to accept that some of the theoretical roles that 'attention' has been put to can be fulfilled by just one concept. Perhaps some of them can. But this will not carry my opponents very far. If my opponent is attempting to use these observations to argue for *monism* (recall that monism is the view that there is *one* concept of attention which

[20] I will have more to say about the links between 'attention' and 'species' in chapter 5.
[21] The creationist concept is couched in terms of divine creation. The phenetic concept divides organisms into species based upon overall similarity .

is superior to all of the others) then it is not enough that there be a concept that can fulfil a *small* number of the theoretical roles that have been associated with 'attention', there must be *one* concept that is *more* theoretically useful than all of the others, at *all* (or at least most) of the roles that 'attention' has been given.

My opponent could hold the position that there are a variety of 'attention' concepts, useful for different roles, but which *all deliver the same verdict on whether GY was attending or not.* As I have emphasises, this is certainly a possible position, but given the large number of roles that we have found above, some of which are intimately related to consciousness, and some of which are completely divorced from it, it is likely that a list of the concepts that can fulfil all of the theoretical roles in question would contain concepts that are so disparate from each other that we will receive a range of different answers to the question of ST and NT. Return to one of Rensink's subconcepts of attention, which has the theoretical use of explaining how information enters the optic nerve. It seems likely that on any 'attention' concept that is well set to fulfil *this* theoretical role, GY will count as attending to the stimuli in question (since information is being transmitted from his retina along his optic nerve). Conversely, when we look at concepts of attention based around the theoretical roles involved in drawing links between consciousness and rational belief formation (as we find in Smithies and Campbell) then it is unlikely that GY (as well as the subjects in the other experiments) will qualify.

The argument of this section is an inference to the best explanation: since there are so many different theoretical roles for 'attention', we should not expect *one* concept to fulfil them all, and we should not expect all of the useful 'attention' concepts to deliver the same verdict on whether the subjects in the experiments discussed in chapter 2 are attending to

the stimuli or not. Readers may think that it is incumbent on me to be more precise and say *which* concepts are useful for *which* particular roles, and *why* they are so useful. As I mentioned above, I will be doing this in chapter 4, as it will be one of the main premises in one of my arguments for pluralism. For now, the argument of this section is sufficient to make the general points relevant to this chapter.

§3.5.2-Objection: the future of attention research.

My opponent could say that it may be that the theoretical uses that we wish to put 'attention' to will narrow in the future. Perhaps we will decide that we do not wish to put attention to a certain theoretical use, and thus give up employing the concept that fulfils this use. If the uses narrow down far enough, then in the end there may be only one theoretical role that we expect 'attention' to fulfil. Correlatively, the concept that fulfils this role may be the only concept of attention that we are interested in retaining. So if the theoretical roles narrow down far enough, we will have monism after all.

This argument is slightly off target. It will be obvious that my main emphasis has been on contemporary attention research. My claim is that given the *current* state of attention research, finding one concept that can fulfil all of the roles in question is unlikely. If the state of research changes a great deal in years to come, then perhaps there will be one concept which can do all of the work we want to be done. I have no problem with that. Indeed, there would be something fishy about an argument which derived much of its force from an examination of the theoretical weight assigned to 'attention', but which applied regardless of the state of attention research. So, this objection is of limited force.

Nonetheless, I think that there is good reason to think that attention research will *not* narrow down on one theoretical role for attention in the near future (I obviously cannot

make any predictions about the distant future). To see this, consider the recent explosion of interest in attention in the philosophical literature. In addition to the *current* interest in attention, there are new research programs which are just starting up. In his 2014 book, Wayne Wu lists several areas that could benefit from further discussion of attention. These include attention in schizophrenia, attention in non-visual perception, attention and computational approaches to the mind, attention to time, attention and veridicality, attention and linguistic pragmatics and attention and testimony (p.9).

This recent proliferation of theoretical roles for attention (and different concepts of 'attention') can serve as the basis for an inductive argument for the claim that the number of theoretical roles assigned to 'attention' will *increase* over the coming years. Predictions about the future of philosophy and psychology should be treated with care, but I think that this is a reasonable one. Indeed, I am inclined to think that proliferation of certain key concepts should be expected to be the norm in most fields, not merely those that that concern 'attention'. That is, the more theoretical weight one concept is expected to take, the more likely that the concept will fragment into a variety of distinct concepts.[22] This is what we saw in the 'attention' case and it is also what we see in a great many others. If 'attention' research began to narrow down on one theoretical role in the future then it would be the exception and we would rightfully be surprised by this.

§3.6-The meta-question.

I have been discussing the use of 'attention' in folk psychology, empirical psychology and I have also been looking at the idea of 'theoretical use'. There is however, a more fundamental question lurking in the background. Let us suppose that there were one

[22] We shall see several examples of this in chapter 5.

concept of 'attention' which perfectly captured the use of 'attention' in folk psychology; as well as one concept which perfectly captured its use in empirical psychology, and one which was the most 'theoretically useful'. An immediate question that we could then ask is this: why should we prefer any one of these three concepts above the other two? This is what I call the 'meta-question': what independent reason could we have for preferring, say the concept as used in empirical psychology, to the one as used in folk psychology, or the one that is most theoretically useful in philosophy and/or psychology? Since monists argue that there is one concept of 'attention' that is privileged above all others, this question must be addressed.

In many ways, this is the most fundamental question of all, and it has particular traction for the problems that we encountered in the previous chapter, with the debates over ST. To see this, recall that Kentridge, Prinz and Smithies were all working with different concepts of 'attention', and that this is what the debate over ST comes down to (at least in large part). If we set aside the individual concepts that they have used, and focus instead upon the *motivations* that each thinker gives for their own account of attention, then the meta-question will come to the foreground.

As we saw, Kentridge defines attention in terms of *task performance.* This is unsurprising, as Kentridge is an empirical psychologist, and defining attention in terms of task performance is useful for making attention empirically *testable.* This would fit with the way that Kentridge characterises his definition of attention, when he says that 'the core of attention as cognitive psychologists understand it…' (2011, p.229). Unsurprisingly, Kentridge also supplies an operationalist definition of 'consciousness' in terms of reportability (2011, p.230).

The motivations behind Prinz's definition are very different. One of Prinz's explicitly stated aims in his definition of attention is to find a functional analysis of the *folk psychological* concept of attention (2012, pp.90-95). When we turn to Smithies, we find his reasons are also different. One of the core things motivating Smithies' account seems to be finding 'a theoretically significant distinction… between conscious attention and its unconscious functional analogs' (2011, p.259). Smithies' other motivations also involve drawing links between consciousness, knowledge, justification and demonstrative thought (see esp. 2011a, pp.261-268).

Kentridge's definition is one useful for experimental psychology, Prinz's is (among other things) designed to mirror folk psychological usage, and Smithies' seems to be designed to fulfil certain theoretical roles within philosophy. When we take account of these differing motivations, it is unsurprising that the resultant definitions differ from each other so much. So, even if we could agree that there was one 'core' usage in folk psychology, and one in empirical psychology, and also one that was the most 'theoretically useful' then the issues raised in the previous chapter will simply re-emerge, because there will be a question raised as to why we should even pick one *set of motivations* over others.

This is another version of a difficulty that I pointed out in §3.5. The problem is that it is not clear what the desiderata for a concept of attention even are. Things would be simpler if all of the interlocutors in the debate agreed on what a definition of attention should and should not do. Then we could decide which definition fulfils these criteria best and judge the issue based upon that. However, there is not even agreement in this regard.

§3.6.1-From Folk Psychology to Empirical Psychology and back again?

There is a possible response to the worries raised by the meta-question, which would rely upon some controversial theses about semantic content. The idea would be that when

certain laypeople use certain terms, they in some sense 'defer to the experts' to set the meanings of their words. So, on a crude version of this theory, when a layperson uses the term 'arthritis', they simply mean 'whatever the doctors mean by arthritis'. Similarly, many philosophers have thought it plausible that when a layperson uses terms such as 'electron' or 'Higgs-Boson particle', even though they don't know that much about electrons or the Higgs-Boson, they just mean 'whatever the physicists mean by 'electron' or 'Higgs-Boson''. The crucial claim is that the semantic content of (at least some of) my words is fixed by certain experts in the fields in which the word is used. This kind of view is most famously associated with Burge (1979).[23]

Suppose we accepted this social externalist account of semantic content. My objector could then claim that when normal people use the folk psychological predicate 'attention' all they mean is 'whatever the experts mean by this'. So, the meaning of the folk psychological predicate is fixed by the experts (presumably psychologists and neuroscientists) in the field. For this reason, the question of whether the folk psychological use is to be preferred to the use in empirical psychology does not really arise, because in a way, they are *the same* use. This is because the content of one is inherited from the content of the other. This would go some way towards addressing some of my arguments in this chapter (though obviously not all).

I cannot delve deeply into the issues that this objection raises here, but I think that it faces various problems. The first is obvious, which is that the response is a hostage to fortune, because it assumes a very particular model of externalism about semantic content,

[23] Burge himself goes further, and argues that not only is *semantic* content sometimes fixed by experts in the fields, but so too is (at least some) *mental* content. However, this is more than my objector requires, and I set it aside here.

which is highly controversial. A second (and related) point is that, when it comes to certain terms, semantic externalism of this flavour can appear very plausible. The most obvious cases are neologisms in science. When a certain term is a neologism, invented for the purposes of science, it is plausible that (when that term enters folk usage) its meaning is inherited from the science that produced it.[24] This is why terms such as 'arthritis' and 'electron' are reasonably plausible examples of terms whose content is (at least in part) derived from deference to the practices and disciplines that invented the term. However, for terms that were *not* neologisms in science, the issues are much more murky. It is much less clear whether the content of terms such as 'lily', 'fish' or 'fruit' is fixed in part by deference to the experts in the field, or whether the folk usage of these terms is just *different* from the scientific usage (cf. Dupré, 1993). Of course, this is the case with 'attention', because it is not a scientific neologism. The case of 'attention' is thus far less clear cut than cases such as 'electron' or 'Higgs-Boson'.

§3.7-Some reasons to accept pluralism.

I have argued that the chances of being able to defer to folk psychology, empirical psychology or 'theoretical use' in order to decide which concept in the ST or NT debates is better are slim. In my view, we can go further at this point, and mount an argument for pluralism about the 'attention' concept. To see this, consider what we have encountered in our examination of the use of 'attention' in folk psychology and empirical psychology. In folk psychology we have found that the concept is used in different ways, by different speakers, to convey different meanings at different times. We have good reason to think that subjects' intuitions about what is and is not 'attention' are malleable and non-uniform, and

[24] Thanks to Robin Hendry for pointing this out to me.

are context dependent. In empirical psychology, we have a plethora of different accounts, which serve different theoretical purposes. We have found that some empirical psychologists define the term so liberally as to make certain claims regarding attention totally trivial, and others define it so restrictively as to make the same claims obviously false.

The argument at this point is simple: the term 'attention' is used in many different ways throughout folk psychological and empirical psychological discourse. Therefore, *if* we wish to develop an account of 'attention' which coheres with the folk or empirical use of the term, then we should embrace pluralism: we should embrace a range of different concepts of 'attention', each tailored to particular ways that 'attention' is used in folk and empirical discourse.

I do not claim that by embracing pluralism, we will be able to come up with a list of 'attention' concepts which together perfectly capture *all* of the different uses of the term in folk psychology and empirical psychology. I think that the use of the term is far too malleable for that. Rather, my claim is that the term is *so* malleable and context dependent that the best way of capturing the meaning of particular utterances of 'attention' is to tailor a particular 'attention' concept to fit what is being expressed in that utterance. Such a method will result in a wide variety of different 'attention' concepts, none of which capture *all* of the folk or empirical uses of 'attention', but which (when taken together) can capture a large number of them.

I call this preliminary argument for pluralism the 'argument from use'. Notice that an analogue of the meta-question cannot be raised for the argument from use. The meta-question asks why we should prefer the use of 'attention' in folk psychology to that in empirical psychology, or to the most theoretically useful concept of 'attention'. Such a

problem does not arise for the pluralist, because the pluralist is happy to accept a range of different alternative meanings of the term 'attention', some of which may be drawn from folk psychology, others from empirical psychology, and others that are theoretically useful. The need to select which of these motivations is 'better' does not arise: we simply accept them all as equal.

Readers who have a more revisionary mindset may be unimpressed with this argument. Some readers will claim that we *should not* be in the business of developing an 'attention' concept which just mirrors folk psychological or empirical psychological use of the term, and thus the argument from use will have limited force. Such readers are deferred to my *main* argument for pluralism, which will come in the next chapter. However, for those who *are* interested in developing a concept of 'attention' which mirrors folk and empirical use of the term (at least to some extent) then this argument will be a strong motivation towards pluralism.

We can reinforce this argument for pluralism by posing an important question to my opponents. The question will be examined and expanded upon in chapters 4 and 5, but it will help if I bring it up now. In this chapter, we have been examining the possibility of deciding which account of 'attention' is *better* by consulting various practices. The question is this: what good reason do we have to think that there *must* be one concept that is better than the others in the first place? Correlatively, what good reason do we have to expect that there must be one objectively correct answer to the question of whether ST or NT are true or not? Why would we think that there just has to be one concept (or a small number of concepts) which all deliver the same answers to these questions? What motivation could we

have for these claims? It is important to raise this question now, as it is unclear precisely what the motivation for monism is.

§3.8-Conclusion.

This chapter can be read in two ways. Firstly, one can read it as a continuation of the last chapter, and an examination of certain proposed solutions to the problems pointed out in that chapter. In this sense the chapter serves to argue that the concept 'attention' as used in folk psychology, empirical psychology, and in terms of 'theoretical use' cannot serve to help us to decide which account of 'attention' is *better* in the debates over ST and NT and thus cannot help solve the problems that were pointed out in chapter 2.

This chapter can also be read in another way, as an independent examination of the use of 'attention' in folk psychology, empirical psychology and the theoretical uses of 'attention'. The chapter also contains (§3.7) an argument for pluralism about 'attention' which can stand independently of the main argument for pluralism which will be developed in the next chapter. For this reason, I would hope that even readers unconvinced of the arguments of chapter 2 (and much of the rest of the thesis) can still accept many of my claims in this chapter.

My arguments of this chapter have been quite abstract, so will it not help to see how they pan out *when they are applied to the philosophical theories that already exist?* In particular, there are thinkers who have tried to give *necessary and sufficient* conditions for attention, is it not important that I examine these views? Indeed, I think it is important and I shall do so in the next chapter.

Chapter 4
The Attention Essentialist Programme.

§4.0-Summary.

This chapter will concern certain philosophical attempts to give a reductive analysis of attention. I focus particularly on several thinkers who try to give necessary and sufficient conditions for attention in a non-circular way. I call this approach 'attention essentialism'. It will be argued that several attention essentialist accounts fail when they are read as theories of the folk or empirical concept of 'attention'. It will also be argued that attention essentialism as a whole should be rejected, when read this way. However, it will be argued that some of the proposed accounts of 'attention' put forward within the attention essentialist programme can be retained, provided they are theoretically useful. In the latter half of the chapter, the main argument for pluralism about attention will be given, and this pluralist argument will be defended from criticisms.

§4.1-Attention essentialism.

The arguments of the previous chapter were quite general. Much of this chapter will be dedicated to the question of *how* these problems apply to individual theories of attention that have been put forward in the philosophical literature. It will also be helpful to draw out some of the wider consequences of our assessment of these theories, as some of these considerations will lead me on to my main argument for pluralism about 'attention'. The discussion of this chapter will at times be quite complex, as all of the theories that I will examine can be interpreted in multiple different ways. However, I think this complexity is worth it, in the service of raising and discussing issues that have not yet received the focus they deserve.

The theories that I shall be examining take as their starting point what Wayne Wu (2014) dubs the 'metaphysical question' which is 'what is attention?' In answer to this question, many thinkers have attempted to give what can be called a *reductive analysis* of attention, where that phrase is understood as giving necessary and sufficient conditions for

attention in non-circular terms. In the literature so far, individual positions couched within this project have been examined, but the present chapter and the paper that is its distant ancestor (Taylor, 2014) are the first attempts to give a serious philosophical analysis of the project taken as a whole.

The general project is one that I call 'attention essentialism' and it can be characterised thus:

> **Attention Essentialism:** There exists a set of conditions α such that fulfilling the conditions in α is necessary and sufficient for something to be attention. All and only those entities that fulfil the conditions in α are instances of attention, and the conditions in α give a non-circular explanatory analysis of 'attention'.

It is important to notice that in giving necessary and sufficient conditions for attention, attention essentialist theories are not only theories of what attention *is* but also theories of what attention *is not*. To see this, consider the fact that attention essentialists claim that all *and only* those entities that are captured by their criteria count as 'attention'. In this way, attention essentialist theories carry with them a commitment to the view that there is one privileged and 'correct' analysis of attention, and that if something fails to meet the criteria in the analysis then it does not deserve the name 'attention'. For these reasons, I shall interpret attention essentialism as a variety of monism. The monist attitude is evident in the ways that attention essentialists describe their own project. Thus, Mole:

> 'This book presents a single unified theory of attention, intended to apply to attention in all its forms. According to this theory, 'attention' is not a family

> resemblance term, nor is it ambiguous, or folksy, or ill-defined, or nonreferring' (Mole, 2011a, vii).

§4.1.1-Attention essentialism: what are we trying to do?

As I said, the attention essentialist theories that I will examine arise out of a desire to answer the question: 'what is attention?' However, as soon as we reflect on what this question amounts to, the issues already get extremely murky, as the question is itself open to various interpretations. Which interpretation we opt for will dictate how we view each theory. As I have already highlighted in chapter 3, the thinkers in question sometimes assume that coherence with the folk psychological use of 'attention' is a virtue of a theory of attention. Recall from §3.2 that Mole takes it as 'an axiom of philosophical theorising' that a theory of attention should 'be a theory of the phenomenon referred to by the English word 'attention'' (forthcoming a, §1.4), where he means the *folk* use of the English word 'attention'. Indeed, it would be difficult to understand the methodology of using examples and counterexamples from normal folk psychological discourse, unless the thinkers in question were aiming to give an analysis of the *folk* concept of 'attention'.

As I also pointed out in chapter 3, many of the thinkers *also* draw on the use of 'attention' in empirical psychology, as well as using arguments to do with theoretical use. This gives us at least three different ways of reading each attention essentialist theory: as an analysis of the folk psychological concept of attention; as an analysis of the empirical psychological concept or as the development of a concept which is useful for certain theoretical purposes. These desiderata are rarely explicitly separated, so perhaps the theories should be read as having some *mixture* of these aims.

It is a pervasive and pernicious feature of the literature that precisely *what* a theory of attention is aiming to explain is almost never explicitly addressed, even though this will affect how we view the entire project in question. For this reason, it is difficult to examine the viability of attention essentialism in a general way, because what the theory is intended to be a theory *of* will affect how we assess its success or failure. It may be that a theory succeeds when judged by one set of criteria, but fails when judged by another set. For this reason, the best way to examine these theories is divide and conquer: I shall see how the theories fare when they are read one way, then assess their success when they are read another way.

Firstly, I shall consider the theories when they are read as giving necessary and sufficient conditions for the folk or empirical use of 'attention'. In the previous chapter, I argued that the folk usage of 'attention' is extremely heterogenous and steeped in conversational pragmatics. If I am correct about this, then we would not expect to be able to give necessary and sufficient conditions for the folk usage of 'attention'. I shall argue that this is indeed the case. I shall also argue that the theories in question deviate from the use of 'attention' in empirical psychology as well (§§4.2-4.4). I will then turn to considering whether we should expect *any* theory to give necessary and sufficient conditions for the folk or empirical use of 'attention' (§4.5) and I argue that the project as a whole should be abandoned. My arguments in this section are intended to supplement the arguments of the previous chapter.

I shall then assess the theories when they are read as attempting to develop some theoretically useful concept of 'attention' (§4.6). I shall argue that *several* of the accounts of 'attention' succeed at being theoretically useful in some way, and give specific examples.

This will lead me on to my main argument for pluralism, which will draw upon the current chapter and chapters 2 and 3 (§4.7). I then refine this argument further, consider some objections and offer replies (§§4.8-4.9). I will conclude by agreeing with some aspects of the attention essentialist programme (because I am willing to accept several of the concepts put forward within this programme as theoretically useful). However, I reject the monist background on which attention essentialism is based.

§4.2-Judging attention essentialist theories.

We can begin by making explicit some criteria by which to judge attention essentialist theories, when they are read as theories of the *folk* or the *empirical* concept of 'attention'. I suggest the following two:

1) *The extensional adequacy criterion:* The account should not be so lenient as to include any cases which clearly are *not* cases of attention, or be so parsimonious as to exclude any cases which clearly *are* cases of attention.
2) *The non-circularity criterion:* The account should not be circular, either obviously or non-obviously and should not make use of terms that are as obscure or more obscure than the term to be explained.

Criterion (1) will strike many as a strong constraint to fulfil, but note that criterion (1) is not too much to ask of someone who claims to be able to give necessary and sufficient conditions for attention, as attention essentialists attempt to do. Different readings of attention essentialism will yield different versions of criterion (1). For example, if we are trying to explain the folk use of 'attention' then criterion (1) would be the criterion that we

should not deviate from the folk use of 'attention'. *Mutatis mutandis* for attention essentialism as applied to the use of the term in empirical psychology.

Of course, the analyses that are offered by attention essentialists should not be circular, and this is the basis for criterion (2). Analyses can be circular in an obvious or a non-obvious manner. A clearly circular analysis of attention would simply be: 'something is attention iff. it is attention'. Analyses can be circular in less obvious ways than this of course. Suppose I purported to analyse 'napkin' as 'serviette'. The problem here would be that 'serviette' and 'napkin' are simply synonyms, so by invoking one to explain the other, we fail to give an adequate analysis of 'napkin'.[1] One good example of a non-obviously circular analysis is Donald Davidson's (1969, 1970) criterion for the individuation of events. Davidson's claim was that two events are identical iff. they have the same causes and same effects. The problem with this was that Davidson's ontology dictated that all causes and effects were themselves events, so to defer to them in explaining the individuation conditions on events was to assume that they were themselves well individuated, which was the point at issue.

Closely related to the issue of circularity is the demand that the criteria given in the analysis should not be as obscure or more obscure than the term to be explained. This is connected to the requirement that attention essentialist theories should be explanatory: if we make use of terms in the explanation that are just as obscure as those in the term to be explained then we will not have made much progress in explaining the term in question. Obviously, two thinkers can disagree on what counts as 'obscure'. In such cases, I have no

[1] Of course, I don't want to get dragged into a depressingly middle-class debate about whether 'serviette' really is synonymous with 'napkin'. This was the most plausible example I could think of, and the point is clear.

general method for resolving such difficulties, and I would advocate treating them on a case by case basis. I think that these criteria fit with how the attention essentialists understand their own project.

§4.2.1-Attention essentialism and pluralism.

Before we begin assessing the attention essentialist theories in question, it will be helpful to see how attention essentialism relates to pluralism, so that we can have a general idea of how the two projects stand in relation to each other. In chapter 2 I made a distinction between monism and pluralism about 'attention', but (as I urged in that chapter) they need not necessarily be mutually exclusive options.

Recall some of the problems I pointed out in chapter 2, which gave rise to the central discussions of this thesis. In that chapter, problems arose because interlocutors were using the word 'attention' in significantly different ways, and giving different answers to the question of whether certain subjects in certain experimental paradigms were paying attention to the stimuli in question. Now, suppose we have an attention essentialist account which we are happy with, and which we are convinced is superior to the other accounts but *which is itself too vague to settle the dispute of whether the subjects in question are 'really' attending or not*. So, we would then have a set of necessary and sufficient conditions for attention, but we would still not be able to use this to settle the questions that arose in chapter 2. The same problems would then arise: different interlocutors could sharpen the concept in ways different enough to give different answers to important questions, and we would not be able to decide which one of these accounts was 'correct' (because *ex hypothesi* the attention essentialist account we would have on the table would be too vague).

This point may seem abstract, so a concrete example will help. Suppose that we are happy that that we could give a reductive analysis of 'bachelor' as 'unmarried male'.[2] We would then think that we had necessary and sufficient conditions for something to be a bachelor: all and only unmarried males are bachelors. Suppose that even amongst those who accept this reductive analysis, there is a dispute about whether a certain male is a bachelor or not. Suppose that some people insist that he fulfils the requirements to count as 'married' whilst others insist he does not (maybe the vicar pronounced him married, but he never signed the registry book or something like that).

In this case, even though everyone accepts the essentialist analysis of 'bachelor' as 'unmarried male' it will be indeterminate whether he is a bachelor or not, because it is indeterminate whether he is 'married' or not. There would be different ways of sharpening the concept 'married', which will result in some people claiming that he is married and others insisting that he is not. Correlatively, he will be a bachelor when employing one sharpened concept of marriage, and not when employing another concept. Different participants in the debate could disagree in this way, *even though* they all agree that 'bachelor' can be given necessary and sufficient conditions in a reductive manner in terms of 'unmarried male'.

The point of this example is to show that a unifying analysis of a concept at one level (of 'bachelor' as 'unmarried male') is compatible with a pluralism of that concept at a more fine grained level (when it is indeterminate whether or not a subject fulfils the criteria laid down in the analysis). So a fine grained pluralism about a concept (of a level required to

[2] This is obviously not correct (a five year old boy is not a bachelor) but we can assume it for the sake of the example.

make distinctions between different interpretations of 'married') can be compatible with a more abstract and vague monism about that very same concept.

If something analogous were the case with 'attention' then what would arise is that at one level we could embrace monism about attention (because we would have our attention essentialist account that we are happy with). But at a more fine grained level of analysis, we would still have to make divisions between different subconcepts of attention because there will be different ways of sharpening the monistic account that we have on the table, and these different ways of sharpening the concept will give different answers to important questions. What would then arise is a combination of pluralism at a fine grained level, and monism at a more abstract level: we would have a unifying analysis of attention, but we would still need to accept pluralism to ask questions that this unifying analysis of attention is too vague to resolve.

So, attention essentialism is not flatly incompatible with what I have been arguing for in this thesis. We can merge a 'soft' variety of monism (where we have a unifying account of attention at a very abstract and vague level) with a more fine grained version of pluralism (where certain sharper distinctions will have to be made at a finer level in order to have certain debates).

§4.3-Jesse Prinz.

We are finally in a position to address the attention essentialist theories specifically. As we saw in chapter 2, Prinz (2010a, 2011 and 2012) has advanced his own account of attention as part of his theory of phenomenal consciousness. We can helpfully remind ourselves of Prinz's theory:

> 'attention can be identified with the processes that allow information to become encoded in working memory. When a stimulus is attended, it becomes available to working memory, and if it is unattended, it is unavailable' (2011, p. 184).

Prinz also says that:

> 'this account provides the only common denominator across the range of cases that we regard as examples of attention' (2012, p.95).

It is clear that, to Prinz, being available to working memory is necessary and sufficient for something to be attended. Recall that working memory (as Prinz understands it) is a particular memory system where information is encoded and stored. Stimuli encoded in working memory become available for reasoning systems in the brain to access, and can also be used in certain kinds of action control and are available for verbal report. I think that Prinz's account of attention fails to fulfil criterion (1). Specifically, Prinz's account seems to encompass *too much:* it includes things which we do not normally think of as attended to in folk and empirical psychological discourse.

One of the central concepts in this account, that of 'availability' to working memory is a dispositional concept. For something to be 'available' to working memory is for it to be *able* to be encoded in working memory *if* it is required for such encoding. For something to be *available* to working memory, it need not actually be encoded, just as money in the bank can be available even if it is not withdrawn. The distinction between availability to working memory and encoding in working memory is of central importance in Prinz's theory of consciousness (see esp. 2012, pp.99-106) and we will be returning to it below.

The problem is that there seem to be a lot of mental states that are *available* in this way to working memory, but which are clearly not attended to. I am thinking of what Ned Block (1995a) calls 'quiescent beliefs', which are beliefs that we have, but of which we are not thinking at this particular time (see also Chalmers, 1997 and Crane, 2013). One example is the belief that Canberra is the capital of Australia. This belief is one that you will already have had, and you have likely had it for several years. This belief is certainly available to working memory. You can report the content of the belief, and you can use the belief in reasoning, for example, you could infer from the fact that John is going to Canberra, and that Canberra is the capital of Australia to the conclusion that John is going to Australia. The belief that Canberra is the capital of Australia is available to working memory, and it has been available to working memory for the whole time that you have been reading this, because it *could* have been encoded in working memory *if* it was required.

By Prinz's definition of attention then, this belief must have been attended to all along (because it was available to working memory all along). However, this seems wrong. It is clear that the belief was not *attended to* at all (at least, not until you read this paragraph). You are very unlikely to have been attending to the fact that Canberra is the capital of Australia until I brought it up just now, but all along it was available to working memory for you. So here we seem to have an example of something which is available to working memory, but which is not attended to. What this objection highlights is the fact that mere *availability* to working memory is too broad a notion to be an accurate account of attention.

It seems absurd to deny that this belief was available to working memory all along, and it seems equally absurd to claim that it was attended to all along. So here we have an

example where attention and availability to working memory dissociate. For this reason, we have good *prima facie* reason to reject Prinz's account of attention, as it fails by criterion (1).

Notice that the fact that Prinz's concept includes quiescent beliefs is a problem when his theory is read as an analysis of the *folk* use of 'attention' *or* the use of it in empirical psychology. I know of no empirical psychologist who would want to claim that quiescent beliefs are *always* subject to attention, even when the subject is not thinking about them.

§4.3.1-Replies.

Prinz may argue that quiescent beliefs are not really 'available' to working memory in the sense relevant for his theory. How might he do this? One likely route will be to invoke his account of the neural *realisers* of availability to working memory. In his (2012, ch.4), Prinz claims that availability to working memory is realised by 'gamma synchrony'. Prinz could then say that (if we assume that availability to working memory is realised by gamma synchrony) it will be an empirical question whether quiescent beliefs are available to working memory because it is an empirical question whether they involve gamma synchrony.[3]

One obvious issue with this response is that it will not so much show that Prinz's account is right, as it will leave it an open empirical question whether it is right or not. However, a deeper problem with this response is that it relies on identifying availability to working memory with gamma synchrony, and to rely on this identification in response to my concern is to beg the question. To see this, let's examine how the identification between availability to working memory and gamma synchrony is motivated in the first place, for

[3] Prinz does not consider my objection specifically, but some of the things he says hint at the fact that he might give this kind of response (e.g. 2011, p.187).

Prinz (see Prinz, 2012, ch.4). The argument draws on Lewis' (1996 and 1970) method for identifying mental faculties with brain events.

Prinz's argument has the following structure:

1) Define 'attention' functionally (as availability to working memory).
2) Look at those areas of the brain that fulfil this functional role (i.e. look at those areas of the brain that do support availability to working memory).
3) Find the properties of those brain areas in virtue of which they fulfil this functional role.
4) Claim that these properties (in this case gamma synchrony) realise the role of availability to working memory (this delivers a neural definition of the functional role specified in step (2)).

Notice that the definition of 'availability to working memory' used in step (2) cannot be a *neural* definition, because the point of this whole argument is to help us *find* such a neural definition. The problem is this. *Before* we can make the identification between gamma synchrony and availability to working memory, we must decide which areas of the brain fulfil the functional role of availability to working memory, defined non-neurally (this is step (2)). Only once we have already decided which areas of the brain *do* in fact fulfil the functional role in question can we then go about making the kind of theoretical identification between this functional role (availability to working memory) and a certain neural property (gamma synchrony). But when we look at the areas of the brain that realise the role of 'availability to working memory' (defined non-neurally as in step (2)) we find that quiescent beliefs fulfil this role perfectly (before we even look to the neural details).

So, when we perform step (2) in the above argument, we conclude that quiescent beliefs *do* fulfil the role of availability to working memory. It would be a mistake at this point to attempt to claim that they do not count as available to working memory because they do not involve gamma synchrony, because that assumes that we have good reason to identify availability to working memory with gamma synchrony in the first place, which is the question at issue. If it does turn out that quiescent beliefs are not correlated with gamma synchrony, then we would effectively have undercut the motivation to accept the identification of availability to working memory with gamma synchrony in the first place. This is because *before* we make a neural identification of this kind, we must already be convinced that the neural property in question (gamma synchrony) is found in those areas of the brain that fulfils the role of being 'available to working memory' (defined without reference to the neural details, in step (2)), and it does seem as though quiescent beliefs fulfil this role.

The problem falls out of the motivation for identifying gamma synchrony with availability to working memory in the first place. Perhaps a simpler way of putting it is this: the only reason we had to identify gamma synchrony with availability to working memory was that we find gamma synchrony in those areas of the brain that are available to working memory (defined functionally and non-neurally). But quiescent beliefs *do* fulfil this functional role perfectly well. So if it does turn out that gamma synchrony is *not* found in the areas of the brain that correlate with quiescent beliefs, then all that will happen is that we will have lost the motivation to identify availability to working memory with gamma synchrony.

§4.3.2-Extending the criticism to other accounts: Wayne Wu.

My focus so far has been on Jesse Prinz's theory of attention, but I think that quiescent beliefs also cause trouble for other attention essentialist accounts, notably that of Wayne Wu (2008, 2010, 2011a, 2011b and 2014).

At the core of Wu's account is *action*. Wu claims that attention is involved in solving the 'Many-Many problem' which (roughly) is the problem faced by a system which has many potential inputs, and many potential actions it could perform, and needs to choose between them. As I mentioned in chapter 3, one of Wu's examples (2011b, p.53) is an agent looking at a tool bench. There are many potential objects that the agent could act upon (the different tools) and many different actions that she could perform with each tool. This generates what Wu (2011a, p.100) calls a 'behavioral space': there are many potential 'inputs' (in our case, perceptual information about the tools), and many potential 'outputs' (actions that could be performed, such as picking up a hammer) so selection must occur which links a particular input to a particular output. Wu claims that attention *is* an appropriate 'linkage', it is a selection of a certain input which guides a specific output.

Wu uses these ideas to provide a set of necessary and sufficient conditions for attention, thus:

> "S's attention to *X* at *t* is S's selection of *X* so as to solve the Many-Many Problem present to *S* at *t*-namely, selection of *X* inherent in *S's* traversing a specific path in the available behavioural space at *t*." (2011a, p.109).

Or, more informally:

> "If there is to be action, a specific link must be selected, one that is constituted by an input-output connection *where the former guides the latter…* Once the structure of behavioural space is thus characterised, a solution to the Many-Many Problem… plausibly identifies a form of attention-namely, the subject's selection of information (input) that guides or otherwise informs his or her response" (2011a, p.101).

Wu develops his account to include exogenous (or involuntary) attention as well, but my focus will be on the endogenous, voluntary forms that we find in the 'hammer' example. As we can see from the above quotation, the notion of 'guidance' is crucial for Wu's account, the input must *guide* the output. Wu also cashes out this idea by saying that the input must *inform* the output (2011a, p.93) and by saying that the subject must be *attuned* to certain inputs in order to respond to them (2011a, p.111). Wu intends these notions to rule out the familiar problem of deviant causal chains, as when an intentional state causes a certain bodily response apparently without the agent's participation.

The issue arises when we take seriously the idea that attention selects inputs which *guide* and *inform* a response. Return to the example of selecting a hammer to hit a nail, say. It is true that the perceptual experience of the hammer will be one appropriate input to guide the response, but on its own this is certainly not sufficient. A great many other inputs must also be guiding the response. For example, the agent's beliefs that the hammer is an appropriate tool for the task at hand, and also her belief that the hammer can be manipulated in certain ways by the agent's hand. This general feature of a great many actions has been emphasied by Burge (1997) amongst others. When I go to the shop to buy soup, my action is guided and informed (in part) by beliefs such as that the shop will sell me soup in exchange for money, that soup is nourishing, that I have the means to open the can

when I get home, and so on. Without these beliefs (or some beliefs like them) my action of buying the soup would not make sense.

Given this plausible feature of much purposive intentional action, it would seem that a great many quiescent beliefs are being selected to inform and guide a particular action, and that they are being used to guide the agent in traversing a particular behavioural space at a particular time. They are being selected to build appropriate linkages between input and output in a way conducive to creating purposive action. So, by Wu's account, these quiescent beliefs will presumably count as attended to. However (as was the case with Prinz) this seems to be the wrong result: these beliefs are surely *not* attended to when we perform these actions, rather they are entirely *ignored.* When I reach for a can of soup, I am not paying attention to the can-opener I have at home, in any way at all.[4] For this reason, quiescent beliefs cause trouble for *both* Prinz and Wu, and make their accounts stumble by criterion (1). As I emphasised above, their theories do not seem to capture the folk psychological *or* the empirical psychological use of the term. So far I have suggested problems for two prominent attention essentialist accounts, those of Prinz and Wu. I will now press on to examining another attention essentialist account, which is Sebastian Watzl's.

[4]A possible response from Wu may be to attempt to sharpen the terms 'control' and 'guidance' in order to exclude the quiescent beliefs from counting as attended to. I obviously cannot hope to argue that *no* such response could work, but I will say that these terms are notoriously hard to get a grip on, despite their frequent use in the philosophy of action. For this reason, the task ahead of Wu is extremely large and I hold out little hope for its success. Another response from Wu would be to claim that quiescent beliefs are relevant to the formation of *intention* rather than the deployment of physical *action.* However, this would merely force the problem back a stage: suppose an intention were formed by *deliberation* (and that deliberation counts as a kind of mental action). Then Wu's theory would count *all* the beliefs that went into the creation of this intention as attended to (because they guide the mental action of *deliberation*). However, clearly these need not always be attended to, at least not in all instances of deliberation. Thanks to an anonymous referee of *Erkenntnis* for discussion of this issue.

§4.4-Sebastian Watzl.

Sebastian Watzl (2010, 2011a, 2011b and 2011c) has put forth a phenomenologically oriented theory of attention, which attempts to give necessary and sufficient conditions for attention.[5] I criticised Watzl's view in my (2014). I now think that my discussion there was misleading and uncharitable, so I would like to take the opportunity here to give a better explanation of the issues I have with Watzl's theory.

Watzl proceeds by stating several properties that he thinks attention plausibly has (2011a, pp.147-148). One of these is that the phenomenology of attending to something is intimately related to what you are *not* attending to. Watzl says:

> 'consciously attending to something in part consists in consciously experiencing what is unattended in characteristic ways' (2011a, p.155. (italicised in original)).

Watzl also claims that:

> 'consciously attending to something consists in the conscious mental process of structuring one's stream of consciousness so that some parts of it are more central than others' (2011a, p.158).

This is Watzl's main claim about what (conscious) attention is: to consciously attend to something is to make it more central in one's consciousness.

Of course, it is well known that attention can change how we represent the properties of the attended object in phenomenal experience. There is reasonable evidence that attention can make objects appear to have a higher contrast than when the same objects

[5] In his 2011a Watzl puts the view forward as a view only about *conscious* attention rather than attention *tout court.* His views on this are less clear in his 2011b and 2011c. In any case, my discussion will apply whether or not his theory is read as a theory of conscious attention or attention tout court.

are unattended (Carrasco et al. 2004); can make a certain gap appear bigger (Gobell and Carrasco, 2005) and can make an object appear darker than when that same object is unattended (Tse, 2005). A popular (though more controversial) idea is that attention to something makes our phenomenal representation of it more *determinate* (Grush, 2007; Cohen and Dennett, 2011; Stazicker, 2011 and Nanay, 2010. See also Block, 2010).

So, we can accept that attention can lead to a change in how the properties of external objects are represented in phenomenal experience. However, this kind of phenomenal change is not what Watzl has in mind when he says that attention involves making some items 'more central' than others; he argues that the phenomenology of attention cannot be captured in this way (2011a, pp.151-153. See also Wu (2010) for a similar view). Rather, Watzl takes the relevant relation as a primitive:

> 'The relevant structure has as its primitive the phenomenal *peripherality* relation "x is peripheral to y"… Consider the case where you are focussing your attention only on the sound of the piano. In the corresponding attentional structure, all other parts of your experience are peripheral to your experience of that sound. It is helpful to also have a name for the converse of this relation-that x is central to y, just in case y is peripheral to x.' (2011a, p.160).

Watzl also says:

> 'Peripherality is the basic notion. I make no attempts at defining it reductively. It is a relation between experiences. Maybe there is, at the end of the day, a reductive definition of peripherality in more fundamental terms, maybe there is not (I doubt that there is one' (2014, p.66).

Though the peripherality relation is primitive, Watzl does give us a metaphor to aid with understanding:

> 'Here is an intuitive gloss on peripherality: attentional organization is like prioritizing books you want to read. You stack them up with the one that is most important to you on top, and the others further down. Peripherality, then, is like being lower in the stack. The attentional organization of consciousness is such a prioritizing.' (2014, p.66).

We can summarise Watzl's view thus: x is attended to over y iff. x is more central than y in consciousness. And x is more central than y iff. y is peripheral to x in consciousness. The peripherality relation is itself taken as primitive.

In the interests of clarity, it is important to emphasise the points about Watzl's view that I am *not* disagreeing with. I do not know whether Watzl is correct to claim that the phenomenology of conscious attention cannot be explicated entirely in terms of the represented properties of external objects. I also do not know whether or not understanding the phenomenology of attention must involve thinking about how *unattended* objects are represented in phenomenal experience.[6] I do not know whether either of these claims are true, but I will grant them both to Watzl for the sake of argument.

The issue I have is that the peripherality and centrality relations are inappropriate for the task of giving a reductive analysis of conscious attention. We can look at Watzl's overall argumentative structure to see this: attention is explained in terms of the centrality relation, and the centrality relation is explained in terms of peripherality, which is itself left as

[6] Notice that this claim implies that there is consciousness outside of attention (2011a, pp.149-150). I will not dispute this claim here.

primitive. But it is hard to see what something being 'central' in one's consciousness really *could* mean other than just that it is attended to over something else. Equally, it is hard to really understand what it might mean for something to be 'peripheral' other than to say that other things are attended to over it. That is to say, when we ask what 'peripherality' means, we are in danger of having no clear answer other than to say that y is peripheral to x iff. x is attended to over y. Recall Moliere's doctor, who claims that a certain potion causes someone to sleep because it has 'dormitive virtue'. The problem here is that to say that something has dormitive virtue is simply to say that it causes one to sleep. In a similar way, to say that something is more central than something else seems to say little more than that it is attended to over something else.

Watzl does give various *examples* of the peripherality and centrality relations (in the quotations given above he uses the example of listening to a piano piece). However, it is hard to see how this does any more good than presenting examples of something that is attended to over other things and then leaving it at that. Of course, Watzl himself will insist that the notion of 'peripherality' should *not* be explicated in terms of 'attention', but in this case the relation seems quite obscure. It is unclear how much progress we would have made in explaining attention by invoking such a relation. Either the notion of peripherality is itself explicated in terms of attention, in which case we have a circular account; or the notion is not explicated at all, in which case it seems very obscure.

Importantly, I am *not* saying that we *cannot understand* the relations in question. I am willing to accept that by giving examples of attention, we can understand these relations in some way. Perhaps we could learn to recognise the relations, and speak meaningfully about them. Rather, I claim that the relations are unable to do the work required for a reductive

analysis of attention, because it is unclear how they are really different from the concept of 'attention' itself.

For comparison, suppose we wanted a reductive analysis of what a 'belief' is. We could give various examples of beliefs. We could also come to recognise when we do and do not have a belief (we do this all the time). We could also come to speak meaningfully about 'beliefs'. However, this would not add up to a reductive analysis of 'belief': it would not be a set of necessary and sufficient conditions for beliefs that explain what a belief is in other terms. We would be fixing the target phenomenon which we want to explain; but we would not be explaining it. The case is similar with the strategy that Watzl uses. We may come to be able to recognise the centrality relation, and its converse relation, as well as argue against certain ways of analysing these relations, but they do not add up to a reductive analysis of 'attention'. It seems to label the thing which is meant to be explained, rather than providing a solution. For this reason, I think Watzl fails to offer a reductive analysis of 'attention'.

We have now examined three attention essentialist accounts. There is another important account on the table, which is that of Christopher Mole. However, Mole's theory is complex and intricate and will take time to examine properly. For these reasons, I have placed my examination of his theory in an appendix to this chapter. Interested readers can consult that appendix; those less interested in Mole's theory can skip it.

§4.5-The prospects of attention essentialism.

I have examined some attention essentialist accounts and argued that they fail to give a reductive analysis of the folk or empirical use of 'attention'. But do the problems pointed out above apply to *particular* attention essentialist accounts that have so far been put forward,

or are the problems more general? Should we say that a reductive analysis of the folk or empirical use of 'attention' has yet to be found, or should we give up on giving such an analysis at all? These questions are most welcome at this point, and they tap into issues that have not yet received much discussion. I will now stop analysing particular attention essentialist accounts, and turn instead to these more general questions.[7]

Many of the arguments I gave in the previous chapter tell against the possibility of being able to offer necessary and sufficient conditions for the folk or empirical concept of 'attention'. I stand by those arguments, but can we say something more? Is there another reason to be sceptical of attention essentialism? One problem here is that when we discuss the prospects of the overall project of giving necessary and sufficient conditions for 'attention' itself the issues become much more ephemeral, because we are no longer examining a particular proposal, but rather a goal or a general method for approaching the questions. Nonetheless, I think there are two good points that should give us good reason to reject attention essentialism when it is read as a theory of the folk or empirical concept of attention. These arguments will serve to bolster the arguments already given.

§4.5.1-Failure of individual accounts and paucity of argument.

The first is that the failure of *individual* attention essentialist accounts to give us a good analysis of the folk or empirical concept of 'attention' at least gives us some good reason to question the background project on which the theories are based. It is a reasonable point against an overall project that the individual accounts put forward within that approach do not work.

[7] Attention essentialist theories which I cannot discuss for reasons of space are the theories of Carolyn Dicey-Jennings (2012) and Philip Koralus (2014a and 2014b). See De Brigard (2014) for criticisms of the latter.

In response to this, my opponent could say that there are lots of theories which people accept even though no one particular version of that theory is without problems. An example is physicalism about the mind. Physicalism of one form or another is still the dominant position in the metaphysics of mind today, despite the fact that individual physicalist accounts are all extremely controversial, and that all of the major physicalist views have garnered serious criticism from both physicalist and non-physicalist philosophers. However throughout all of this, many have remained committed to physicalism of at least some kind. So (says my opponent) it could be that individual accounts face problems, but the failure of individual accounts need not imply that the overall approach is wrong. We could hold out the hope for necessary and sufficient conditions for the folk or empirical use of 'attention' while acknowledging that each individual theory has problems, just as we can remain physicalists whilst acknowledging that each version of physicalism faces problems.

My counter-response is that the reason that people remain physicalists is because they are convinced of the arguments in favour of physicalism itself. They accept that there are good reasons to be a physicalist, and that the view is overall better than its rivals. Usually, the arguments given in favour of physicalism revolve around the twin claims that it helps to preserve the causal efficacy of the mental (Papineau, 2002 and Levine, 2001) and that it offers a more simple and more parsimonious ontology than its best contender, which is dualism (Block and Stalnaker, 1999 and Hill, 2009). Conversely, what arguments are there for the claim that we should expect to be able to give necessary and sufficient conditions for attention?

The best example of an argument in favour of attention essentialism that I know of is from Brian O'Shaughnessey. Amongst a long and interesting discussion about attention and perception, O'Shaughnessey notes that there are several different tasks that a subject could be involved in, and that attending to some of them makes it harder to attend to others. O'Shaughnessey then says this:

> 'The attention of which *in fact* we speak is such that it can be variously distributed, and such that there is so much to go around and no more. The latter quantitative truth, and the interchangeability of the items that use and need parts of the quotum, show that we are dealing with a unitary phenomenon' (2002, p.282).

Notice that O'Shaughnessey specifically puts his argument forward as a claim about how we 'in fact' speak about attention, indicating that he has the folk concept in mind.[8] There are two ways O'Shaughnessey's argument fails. The first is that the claim that there is a single finite amount of attention which *all* attention-demanding tasks must partake of is an empirical claim which is likely false. One good line of empirical evidence against this view is the classic experiment that shows that dividing attention between sight-reading music whilst reciting a text delivers a performance that is just as good as if one were performing either one of the tasks, and not the other (Allport et al. 1972). This seems at least *prima facie* to tell against the idea that all tasks that require large amounts of attention draw upon some one common resource in order to perform the task attentively.[9] At the very least, O'Shaughnessey's claim is (at least partially) empirical, and he provides no evidence for it.

[8] This is further backed up by the general folk psychological approach of O'Shaughnessey's work.
[9] Rejection of the 'common resource' view of attention is argued for by Mole (2011a and 2012).

Secondly, even if we waive this point, O'Shaughnessey's argument still does not go through. It does not follow from the fact that many tasks that require attention all draw upon the same finite resource to the fact that attention is a 'unitary phenomenon'. Consider an electric car, which has no internal combustion engine. In this car the battery would be required in order to provide energy to the motor, the radio, the brakes, the steering, the air conditioning and so on. However, it clearly does not follow that all of these mechanisms form a 'unitary phenomenon' and we would not wish to claim that this gave us reason to embrace monism or essentialism about all of these mechanisms taken together.

Setting aside O'Shaughnessey's argument, more arguments for expecting necessary and sufficient conditions for 'attention' are very difficult to find. Wu (2014, p.5) says that it is the most 'optimistic' option, but even if this is true it is hardly an argument in favour of it. Mole (2012) notes that some of the arguments *against* it have failed, but again this does not constitute an argument *in favour* of attention essentialism or monism in general. Since the attempt to give necessary and sufficient conditions for attention is a popular position in the philosophy of attention, one would expect there to be several arguments in favour of this project, but really the question of *why* we should even expect such an analysis has been largely unanswered.

My argument invites a question: if there really are no good arguments in favour of this view, then why is it a popular one? Answering such a question in full would require an extended investigation into the sociology of philosophy, which I certainly will not attempt. However, my speculation is that the popularity of attention essentialism (as a theory of the folk or empirical use of 'attention') can largely be attributed to the fact that it is often an insidious motivation which operates behind the scenes, rather than being explicitly stated,

argued for or defended. This attitude toward attention essentialism is both very popular and completely wrong. Attention essentialism should *not* be taken as a background assumption, or starting point. Rather, it is an extremely committal philosophical position, which may well be false, and which requires very substantial arguments in its favour before we should accept it.

There is a simple and plausible explanation for why the assumption is so popular: the fact that the *same* word has been used in various different situations and contexts leads thinkers to assume that there must be some common unity to these uses. It is natural to attempt to *state* what this unity consists in; and the most natural way of doing this is to attempt to say what it is that all of the things in question have in common. This leads us straight to attention essentialism.

§4.5.2-The nature of concepts.

The second argument against attention essentialism applies to attention essentialism when it is read as applying to the folk concept of attention specifically. This stems from the fact that there are good reasons to think that necessary and sufficient conditions cannot be given for concepts in natural language in general.

The evidence for this comes from the cognitive science work on concepts: the view that natural language concepts can be defined (in the sense of being given non-circular necessary and sufficient conditions) is now largely dead throughout both the philosophy and psychology of concepts. It is generally accepted that there are (at best) hardly any concepts in natural language that can be defined, and certainly there are no proposed definitions that are uncontroversial.

Several things have contributed to this general view on the status of definitions. A strong one is the failure of traditional attempts to give non-circular necessary and sufficient conditions for any concepts at all that have been of interest within philosophy. One prominent example of this is the case of 'knowledge'. It is well known that the attempt to analyse knowledge as 'justified true belief' failed because of so called 'Gettier cases' (Gettier, 1963). Attempts to revise the project, and offer new necessary and sufficient conditions for 'knowledge' have met with more counterexamples (Zagzebski, 1994 and Kripke, 2011, ch.7) and many thinkers have now entirely given up on the project of attempting to give a reductive analysis of knowledge at all (e.g. Williamson, 2000). I need not belabour the failure of other reductive analyses of philosophically important concepts here; examples are very easy to find ('art', 'science', 'value', 'mental', 'physical', 'causation'). Indeed, even those words which have often been held up as paradigm examples of definable concepts such as 'bachelor' and 'grandmother' have been subject to counterexamples (see Laurence and Margolis, 1999).

The claim that natural language concepts cannot be defined will require some clarification. Of course, one can *stipulatively* define any concept one wants in terms of any other concepts that one wants. One could stipulate that by 'bachelor' one means 'unmarried man' and (assuming we allow the stipulation) this would be acceptable. Similarly, one can offer the kind of definitions that are usually given in good dictionaries. These definitions almost always (if not always) fail to offer non-circular necessary and sufficient conditions for the concept in question, but rather they serve to give the reader a general idea of what the concept normally applies to, assuming that the reader already has a large background of relevant knowledge and makes many background assumptions about how the concept

should be applied. I accept that concepts can be defined in *these senses* but these kinds of definitions are not what lie at the heart of the attention essentialist programme. The attention essentialists are not *stipulating* what they mean by attention (this would make their claims lack any substance) and they are not simply trying to offer a dictionary definition of attention.

The thinker who is probably most well-known for pressing the claim that almost all concepts are undefinable is Jerry Fodor (esp. 1981, 1998 and 2008).[10] Many of Fodor's arguments have the same structure as the ones given above: almost any proposed definitions one could give seem to be subject to counterexamples.

It will be helpful to quote Fodor on this matter:

> 'these days almost nobody thinks that concepts are definitions… There are practically no defensible examples of definitions; for all the examples we've got, practically all words (/concepts) are undefinable' (1998, pp.44-45).

In addition to the fact that there are (at best) hardly any natural language concepts that can be defined, there was another factor that contributed to the death of the definitionalist view of concepts which was in many ways even more important. This was the emergence of several other theories of concepts which were more explanatorily successful and which fit the data better. I am here thinking of nativism (Fodor, 2008), prototype theory (Rosch, 1978), theory theory (Carey, 2009), exemplar theory (Medin and Schaffer, 1978), neo-empiricism (Prinz, 2002) or some mixture of these theories (Machery, 2009). Obviously, all of these positions differ in many important ways, and I certainly cannot settle the question of

[10]The view is certainly not original to Fodor.

which we should opt for, but it is fair to say that they are all unanimous in their rejection of the definitionalist view.[11]

This allows us to raise the following challenge to attention essentialism:

1) There are (at best) hardly any natural language concepts for which we can give necessary and sufficient conditions, which give a reductive analysis of the concept.
2) Attention essentialism attempts to give necessary and sufficient conditions for the natural language concept 'attention' in a way that offers a reductive analysis of the concept.
3) (Therefore) attention essentialism as a theory of the folk term 'attention' is likely false.

Notice that this argument works against varieties of attention essentialism that are directed at giving necessary and sufficient conditions for the *natural language* concept of 'attention' (this is encapsulated in premise 2). As already argued, the support for premise (1) comes from the dominant theories of concepts in contemporary cognitive science, and we get to (3) from (1-2) by inference to the best explanation.

Someone could resist (1), by claiming that it does not follow from the fact that no good definitions or reductive analyses of a certain concept have *so far* been put forth to the claim that there is *no possible* definition or reductive analysis of that concept (E. J. Lowe is among those who make this point (2012)). This claim is certainly true. Fodor and everyone else have not *proved* that we cannot have definitions of most of our concepts. It could simply

[11] I do not wish to claim that absolutely *no one* holds the definitionalist view. Jackendoff (1992) holds a version of it. However, he is very much in the minority and I will not address his view here.

be that our concepts are *hard* to define and require a great deal of effort and thought to define them. However, we are not looking for proof here. The argument is that the failure of most (if not all) attempts at reductive definitions over the last 2,500 years or so gives us *good reason* to think that most (if not all) concepts cannot be defined in the sense relevant for our discussion.

It is helpful here to consider another of Fodor's points:

> 'Oh well, maybe there's one definition. Maybe BACHELOR has the content *unmarried man*. Maybe there are even six or seven definitions; why should I quibble? If there are six or seven definitions, or sixty or seventy, that still leaves a lot of words/concepts undefined… The OED lists half a million words, plus or minus a few' (1998, p.45).

Fodor's point here is important. Even if we had some definitions which we were convinced were good ones, it would still be true that only a *very tiny percentage* of concepts can be defined. The claim that almost all concepts cannot be defined is all I need for premise (1) to hold up. A few good definitions will not swing the balance here, because the overwhelming majority of concepts would remain undefined. For this reason, I think we have good reason to reject the hunt for necessary and sufficient conditions for the folk concept of 'attention'.

Certainly, the attention essentialist *could* continue to hold on to her position. However, in the light of the paucity of arguments in its favour, the problems with the accounts so far put forth and the fact that it relies on a view of concepts now heavily unpopular in cognitive science, I think it is fair to demand some good arguments for the

position before we can consider it again. We obviously shouldn't hang on to a position just because it is *possibly* true, even though we have good reason to prefer an alternative position.

If what I have been arguing is correct, then contrary to what thinkers such as Mole claim, it was a mistake ever to even *attempt* to capture the folk psychological use of 'attention' in a set of necessary and sufficient conditions, because such a project is simply quixotic. Ultimately, it is true that I am pessimistic about the prospects for attention essentialism to give us necessary and sufficient conditions for the use of 'attention' in folk psychology and/or empirical psychology. But I am pessimistic because I think there are g*ood reasons* to be so.

§4.5.3-Is 'attention' a special case?

Readers may be wondering at this point whether 'attention' is a special case. One of my two arguments in favour of the rejection of the attempt to give a reductive analysis of the folk or empirical concept of 'attention' drew on very general claims about the nature of concepts in natural language. Wouldn't my concerns equally apply to almost *any* term in the English language (or any language)? Is there something particular about 'attention' here?

The answer is yes and no. On one hand, I am convinced by the arguments put forward by Fodor and others that (in general) reductive definitions of concepts in natural language are not to be had. In this respect, 'attention' is not an exception. However, there are at least two important differences between issues surrounding 'attention' and issues over most other concepts in natural language. Firstly, unlike most concepts used in natural language, there has been a great deal of effort dedicated to giving a reductive analysis of 'attention', and as such it is especially appropriate that these concerns be raised here. After all, to say that my arguments could apply to a *great many* different concepts does not make

the arguments any weaker when they are applied to a case where the project of trying to offer a reductive analysis of a certain concept is still alive and well.

A second difference between 'attention' and most other concepts is that most concepts in the English language are not causing quite as many methodological problems as 'attention' is.[12] A large part of this thesis has been attempting to pick apart some of these problems, and (as we shall see below) recognition of these problems forms part of the impetus for accepting pluralism about 'attention'. This momentum is lacking when it comes to most of the words in the English language.

§4.6-Theoretical use.

I have argued that we should not expect a reductive analysis of the folk or empirical use of 'attention', and I have applied this argument to several particular theories of attention in the literature. At this point, a reader could be forgiven for thinking that I simply advocate a wholesale rejection of all of the theories of attention that have been put forward within the attention essentialist programme. Indeed, if the views do face problems, then why shouldn't we just get rid of them all and start from scratch? Why not simply reject them all, based on the fact that they fail to give a good analysis of the folk or empirical use of 'attention'? These are reasonable questions. However, I think that dismissing all of the views entirely would be an overreaction.

In order to see this, suppose that we reject the claim that we should assess a proposed account of attention in terms of whether it successfully gives necessary and sufficient conditions for 'attention', as used in folk or empirical discourse. Instead, suppose

[12] Of course, *some* of them are. Several will be examined in chapter 5.

we assess certain 'attention' concepts based simply upon whether or not they are *theoretically useful*. In a nutshell, this is the approach that I recommend. Importantly, any 'attention' concept could prove to be theoretically useful, not only those that have been proposed as part of an attempt to give necessary and sufficient conditions for attention. On the approach I recommend, we should start assessing the proposed concepts in terms of whether they are theoretically useful, and then welcome *a variety of different 'attention' concepts* for different purposes.

Suppose we take this line of thought seriously. Where will this leave the attention essentialist programme? Should it be abandoned? The answer here is slightly complicated, because I think some aspects of it will have to be rejected, whilst other parts of it can be retained. If we embrace the approach that I recommend, then the *exclusivity* of attention essentialism will have to be rejected, because there will not be one privileged concept of 'attention' which is superior to the others. However, to reject this aspect of attention essentialism is not to say that everything the attention essentialists have said must be totally abandoned. Rather, when we have a proposed analysis of 'attention', we can accept it and deploy it in our theories, *provided that it is theoretically useful*. In this restricted sense we can find a place for some of the theories put forth within the attention essentialist programme, and in this way, we can incorporate them within a pluralistic framework. In a slogan: if it's useful, then I'll take it.

§4.6.1-Specific concepts and their uses.

In §3.5, I gave an argument that was based upon inference to the best explanation for the claim that if we begin assessing proposed 'attention' concepts by whether or not they are theoretically useful, then we will likely end up with *several* concepts that are useful for

different purposes. This was based on the sheer number and variety of roles that 'attention' has been expected to play in philosophy and psychology. However, that argument was quite general, so it will be helpful to bolster it by presenting some concrete cases of certain 'attention' concepts being theoretically useful in particular contexts.

Take Prinz's concept of attention as availability to working memory. We can get a handle on the theoretical use that Prinz's concept has by comparing it with a distinction made by Dehaene and Naccache (2001) and Dehaene et al. (2006) when they discuss the 'global workspace' theory of consciousness. Global workspace theory is (roughly) the view that phenomenal consciousness is identical with information that is 'broadcast' in the 'global workspace'. The global workspace is a functionally defined system in the brain, where information can be stored (typically temporarily). Information in the global workspace then has certain availability relations with other systems in the rest of the brain (which are themselves functionally defined). For example, information that can be used for certain kinds of action control (such as reporting, or the rational manipulation of action and thought) is said to be 'broadcast' in the global workspace. The main claim of global workspace theory is that all and only information in the global workspace system is conscious.[13]

Notice how similar the global workspace is to working memory (at least, on the concept of working memory that thinkers such as Prinz have in mind (see §2.4.2)). Precisely what relationship the global workspace has with working memory has long been one of the core questions for advocates of global workspace theory (see Baars, 1997, 2003 and Shanahan

[13] Dehaene's (2014) book is an extended discussion and defence of his version of global workspace theory. Dennett's (2001b and 2006) 'fame in the brain' theory is a philosophical version of it. His earlier (1991) view has much in common with global workspace theory. Tye (1995), Dretske (1995) and Kirk (2005) have very similar views. I discuss global workspace theory further in Taylor (2013c). One of global workspace theory's strongest original advocates was Bernard Baars (1997, 2002, 2003, 2005 and 2007).

and Baars, 2007). This should alert us to similarities between the conceptual repertoire used by global workspace advocates and by Prinz.

In order to flesh out global workspace theory, Dehaene et al. make a distinction between three stages of processing in the perceptual system, which they label the 'unconscious', the 'preconscious' and the 'conscious' (they sometimes use the labels 'I_1, I_2 and I_3' for these divisions). The first of these is perceptual information that is *inaccessible* to the global workspace (such as repressed Freudian desires).[14] The second category (the 'preconscious') refers to information which *can* be retrieved by the global workspace and used. The final is information that is *in* the workspace. The concept of being *in* the workspace is fleshed out in different ways by different thinkers: Dehaene tends to define it neurally, whilst Baars (2007) tends to rely on metaphor. Importantly, the distinctions between the preconscious and the conscious made by Dehaene et al. are very similar to the distinctions between information *accessible* to working memory and information *accessed* by working memory, which are distinctions that Prinz's concept makes salient. Deheane and Prinz carve up different stages of perceptual processing in much the same way, and their conceptual divisions can be used to perform similar jobs.

Making distinctions between 'available' and 'accessed' information may appear like splitting hairs, but these distinctions really start to matter when they are deployed in certain theoretical contexts. The best example of this is Block's 'phenomenal overflow' argument. The central claim of this argument is that certain experimental results show that phenomenal consciousness has an informational capacity that 'overflows' that of the global

[14] Notice that these distinctions are different from Freud's distinctions of the same names.

workspace (see Block, 2007a and 2007b, 2008, 2009, 2011, 2012 and 2014).[15] Block takes this as evidence that some information in what Dehaene et al. call the 'preconscious' system (which is not in the global workspace) is phenomenally conscious. If Block is correct, then global workspace theory is false.

The point is this: whilst Block's arguments have often been seen as a large problem for global workspace theory, the arguments have often been seen as much less of a problem for views which accept that information that is only 'accessible' to the global workspace can be phenomenally conscious. Block is certainly aware of this fact (2007a, p.492 and 2011). Prinz himself is among the (several) thinkers who deploy the distinction between 'accessed' and 'accessible' information in order to accommodate Block's points (Prinz, 2007a) though he agrees with Block that these results *do* refute global workspace theory. Indeed, Block himself emphasises the importance of such distinctions, describing them as 'certainly useful' (2007a, p.492). This is an instance of the distinctions that Prinz's concept of 'attention' is based upon becoming theoretically important in certain contexts.

This is where Prinz's concept is importantly useful. Prinz's concept of attention may include *too much* when it is measured as a global theory of 'attention', but it can be usefully deployed *in these debates*. Specifically, the distinctions between 'accessed' and 'accessible' information that are at the core of Prinz's concept of 'attention' are important when we are specifically concentrating on the question of what stage in perceptual processing consciousness arises at; when we are thinking about what kind of scope the experimental findings that Block cites have; and how we should change our views about consciousness as

[15] Interestingly, the first version of this argument appeared in Block (1995a) over a decade before its most famous formulation. In his (1995a) Block credits Fodor with suggesting the idea. The results that Block uses for his argument come from Sperling (1960), Kouider et al. (2007a), Landman et al. (2003) and Sligte et al. (2008 and 2009) among other places.

a result. I need not take a stand on any of these controversies, my claim is only that Prinz's concept is *useful* in these debates. So long as we keep the explanatory burden of Prinz's concept narrow, then we can find a use for it, even if it cannot serve to give a privileged and correct reductive analysis of 'attention'. Once we focus on (for example) analysing the phenomenal overflow argument, then the distinction between 'accessed' and 'accessible' information will become vital. These are the distinctions that Prinz's concept makes, and it is in this context that the concept can be deployed, and so long as we keep the scope of application suitably narrow, the concept will retain its use. Once we focus specifically on the case of *perceptual* consciousness, and (even more specifically) on the correct interpretation of the data used in the phenomenal overflow argument, then we find a reason to use Prinz's concept.

I have taken Prinz's concept as my main example, because that concept maps onto a set of conceptual divisions that have already earned their keep in certain debates in cognitive science, and so the claim that it is useful will be easier to see. However, Prinz's concept is not an exception. Recall Kentridge's concept of attention as increased task performance in certain experimental situations, or prioritization of certain pieces of information in the visual field. Such permissive and empirically focussed concepts are clearly useful when we are using the Posner paradigm to probe the abilities that subjects have with relation to invisible stimuli. Such discoveries are useful and interesting, as they give us information about the functionalities of consciousness (by contrasting conscious perception with cases where consciousness is absent, such as that of blindsight or meta-contrast masking). So, when concentrating on these issues particularly, Kentridge's concept will be useful. Again, if we resist the temptation to consider Kentridge's concept as a global

theory of attention, then we can justify its preservation by finding a use for it. I will have more to say about Kentridge's concepts in chapter 7.

This can be taken further, we may say for example that Wu's concept is useful and interesting when assessing the distinction between automaticity and action (this is especially clear in Wu, 2013); Smithies' concept is useful when assessing the connection between consciousness and rational justification. Similarly, the concepts that Watzl delineates can give us useful descriptions with which to think about the phenomenology of attention as organised into foreground and periphery.

Notice that all of these points will equally apply to a thinker who claims that what we should be looking for from a theory of attention is a *mixture* of theoretical fecundity and coherence with the use of 'attention' in empirical and folk psychology. This kind of approach inherits the problems associated with attempting to capture the use of 'attention' in the communities in question *but also* for the same reasons as given above, we would still expect there to be a great many concepts which capture (at least some) of the use of 'attention' in folk psychology and empirical psychology *and* which are theoretically useful. I will return to some of these issues in chapter 7, where I aim to give a plausible explanation of *how* these different concepts of attention arise and why they differ in precisely the ways that they do.

§4.7-Putting the pluralistic pieces together.

I have argued for three things: firstly, that we should abandon assessing proposed 'attention' concepts by whether they offer necessary and sufficient conditions of the folk or empirical use of 'attention'. Secondly, that a better way of assessing the proposed concepts is

in terms of whether they are theoretically useful. Thirdly, that *several* concepts of 'attention' are theoretically useful. My argument for the first claim was the examination of the folk and empirical use of 'attention' in the previous chapter (§§3.3-3.4) as well as the examination of particular proposed accounts of 'attention' in this chapter (§§4.3-4.5 and the appendix to this chapter). My arguments for the second and third claims come from my examination of the number of roles that 'attention' has been expected to fulfil (§3.5) as well as the concrete examples in §4.6.

I am finally in a good position to state my main argument for pluralism. I call it the 'master argument':

Master Argument.

1) 'Attention' is used in importantly different ways by different interlocutors in certain debates.
2) There is no good reason to think that one use of the word is 'better' or 'privileged' above the others.
3) Several of these concepts are worthy of acceptance in philosophy and cognitive science.
4) (Therefore) we should accept a plurality of different 'attention' concepts.

As I mentioned in chapter 3, pluralism about attention is not a position original to me, but where one does find it (e.g. Styles, 1997) one usually sees it simply stated, and rarely argued for. One also does not see the position filled out or made precise. Furthermore, pluralism seems to be a very unpopular position in the philosophical literature, with the majority of thinkers going down the essentialist line. The master argument and the

argument from use in §3.7 are the first fully developed philosophical arguments in favour of pluralism, and the view that I develop is the first philosophically orientated and thoroughly worked out pluralist view in the attention literature.

Since the master argument is important, some notes on it are necessary: the pluralism that the argument is intended to establish is *conceptual* pluralism. That is: that there are many different equally legitimate *concepts* of attention which are worthy of preservation, and that no one of them is privileged or correct above the others. Even though the kind of pluralism advocated is conceptual, it has ontological consequences. This is because several of the concepts of 'attention' on the table have different referents: they refer to different entities in the world.

Premise (1) was the burden of chapter 2. The phrase 'importantly different' in premise (1) is cashed out in terms of different interlocutors in the debates using the term in ways different enough to give different answers to prominent questions concerning attention and its relation to consciousness. Premise (2) was the burden of chapter 3 and the present chapter. Premise (3) is intended to block the suggestion that all of the concepts of 'attention' should be swept aside and that we should start from scratch. The justification for premise (3) comes from §3.5 and §4.6 but these ideas will be returned to later in the thesis.

We find suggestions analogous to premise (3) in other debates in the vicinity. As I mentioned in §3.5.1, the concept 'species' is ambiguous between several distinct concepts. It is incumbent upon those who advocate pluralism about the 'species' concept to provide an argument for why more than one 'species' concept should be accepted in theoretical discourse. Typically, one of the main pieces of justification given for the preservation of the various 'species' concepts is that each one is useful for evolutionary biology (e.g. Kitcher,

1984, Ereshefksy, 1992a and Brigandt, 2003). Premise (3) represents a similar idea applied to the 'attention' debates.

Notice that the master argument is perfectly consistent with the claim that there may be *non-theoretical* reasons for preserving certain concepts of attention as well as theoretical ones. This is consistent with but not implied by the master argument. The 'theoretical use' criterion is intended as a sufficient reason for acceptance of a proposed 'attention' concept, not a necessary one. For example, in §3.7 I put forward a preliminary argument for pluralism based on the fact that pluralism most accurately reflects the use of the word 'attention' in both empirical psychology and folk psychology. The argument of the previous chapter and the master argument are intended to work independently of each other: one can accept one of them but reject the other. However, they both lead to similar conclusions: some variety of pluralism about attention.

I have advocated rejecting the hunt for one concept of 'attention' which captures all and only the uses of the term in folk psychology and empirical psychology. This rejection of an attention concept which mirrors the use of the term in these communities seems *prima facie* to be in tension with the argument I put forward in §3.7 that we should accept pluralism because it most accurately reflects the use of 'attention' in folk psychology and empirical psychology. However, the position advocated in §3.7 is *not* the attempt to develop necessary and sufficient conditions for the folk or empirical use of 'attention', because it accepts that we will not be able to offer such conditions. Rather, my claim of §3.7 was only that because the term is used in many *different* ways in folk and empirical discourse, if we wish to capture a *large* number of these uses, we will need to tailor different concepts for particular uses, and thus we will end up with pluralism.

Notice that the master argument is stated entirely without mention of natural kinds. An opponent of mine may see this as a glaring omission, but (due to the complexities associated with such issues) I will defer discussion of natural kinds until chapter 6.

§4.8-An objection to the master argument.

When faced with this argument, some opponents often insist that there *must* be one concept of 'attention' which is the *right* one. However, at this point I must admit that I no longer understand what my opponents are asking for. I have examined the idea that there is a privileged sense of the word 'attention' in a certain community and found that wanting, and I have examined the philosophical approaches to the question of 'what is attention?' and found them to either be implausible, or to ultimately run straight into pluralism. I have done everything I possibly can to interpret the suggestion that one concept of attention is the best. Having examined these possible avenues, I no longer know what the question 'what is attention *really*?' could possibly mean.

It will be helpful at this point to compare 'attention' to the concept of 'acid' in chemistry, which has been defined in importantly different ways throughout history. The Brønsted-Lowry definition identified an acid as a compound containing an ionisable hydrogen ion, and the Lewis definition defined acids as substances whose constituents have an incompletely filled outer electron shell (Stanford and Kitcher, 2000 cf. Hendry, 2005). It is now widely accepted that there is no single answer to the question of which concept *really is* the acid concept. Correlatively, there often is no satisfactory answer to the question of whether a particular substance is *really* an acid. The real questions are whether it is a Brønsted-Lowry acid, or a Lewis acid, or perhaps some other concept of an acid. Once we

answer these questions the question of whether a substance is *really* an acid need not even enter into it.

Of course, someone could argue that one of the concepts of 'acid' is privileged above the others in the sense of being used by a particular linguistic community, or in the sense of being more useful to chemistry. Let us assume that these possibilities have been exhausted (as they are analogous to the responses to attention pluralism that I have already addressed), but that someone insists that there still is one concept that *really is* the correct acid concept. I suggest that someone insisting that one of the acid concepts was 'best' would be met with puzzlement, and I see no reason to think that the 'attention' concept is any different. The realisation that our concepts have been used in different ways by different thinkers is all part of engaging in discourse at all, and we need not always expect that there *must* be one party who is correct whenever this happens.

It's worth hammering home a point I have made before, which is that the motivations for monism are extremely unclear in the first place. *Why* would we think that there is a privileged concept of 'attention' in the vicinity? Is there any good reason to accept this? In the absence of a clearly developed answer to this question, I do not see why we would work so hard to avoid pluralism.

§4.9-What makes them all 'attention' concepts?

Another question that an opponent might ask is as follows: we may accept that the concept of 'attention' fragments into several importantly different concepts, but what exactly makes all of these concepts *attention* concepts, rather than just a collection of different (and unrelated) concepts?

This question essentially is 'do the concepts that we have delineated deserve the name 'attention concepts' or not?' Personally, I am unconvinced that the answer to this question matters very much. I have already argued that there are a variety of significantly different concepts at play in the literature, no one of which is privileged and (at least some of which) should be preserved in the debates in question. Whether these concepts should be called 'attention concepts' seems to me to be more or less a verbal decision, rather than an important and substantive philosophical question. Whichever way we go on this issue seems to leave the main arguments and views of this thesis basically intact.

So, I remain unconvinced that this question is an especially pressing one. Nonetheless, for those that do consider the question important there are some things that can be said about what unifies the concepts in question, and what makes them deserve the title 'attention concepts'. Firstly, we should notice that what does *not* unify them is a set of certain conditions that all and only attention concepts share. To attempt to give necessary and sufficient conditions for a concept to successfully be an 'attention' concept is perilously close to attempting to give necessary and sufficient conditions for 'attention' itself, and I need not repeat my views on this project.

However, there are some other things that can be said about what makes the concepts in question 'attention' concepts. The main one is that they all share a family resemblance. So, there are various features that we find in some of the concepts, but not others, and there is significant overlap of certain features of the concepts in question, even though there is no core set of features shared by all and only the concepts that we have been looking at. One obvious feature that is shared by many of the concepts in question (though not all of them) is *selectivity.* Another feature that is possessed by many of the concepts in

question is the role of attention in creating *purposive action* (this is especially clear in the accounts of Wu and Mole). Another feature that we find in several of the accounts of attention on the table is that it gives us some kind of epistemological access to certain kinds of information. This is most obvious with Smithies' concept, but we similarly find it in Prinz's concept (given that Prinz thinks that it is essential to attention that it allows information to become available to certain reasoning systems in the brain). We also find epistemological claims built into other concepts of attention in the literature, such as with Campbell's (2002, 2004) claims that attention gives us knowledge of demonstrative reference. Another of the core parts of Campbell's concept of attention is that it involves *feature-binding*, a claim he takes from Treisman and Gelade's (1980) work on attention, and which we saw represented in Rensink's (2013) taxonomy of attention. This is what unifies the concepts that have been delineated, and it is this that gives us good reason to group them together as all 'attention' concepts: their family resemblance.[16]

Some opponents will be unsatisfied at this point, and will likely demand that the concepts in question must have some greater unity in order to warrant the name of 'attention concepts'. Indeed, it may well be said that, in order for my position to deserve the name 'pluralism' at all, I need a more concrete account of what it takes for the concepts in question to be grouped together as 'attention' concepts. Pluralism (the objector claims) implies that the concepts in question are all significantly different, but that they share some very strong unity, which makes them all the same *kind* of concept.

[16] The psychologist John Duncan (2006) believes that 'attention' is a family resemblance term. Insofar as family resemblance unifies the different concepts of 'attention' that have been put forward in the literature, I am in agreement with Duncan.

In response to this, I will say two things. Firstly, I think that to demand such a strong unity runs counter to some of my main claims, and so to demand it is to misunderstand my position. I find it helpful to quote Wittgenstein's claims about 'language'. It is well known that Wittgenstein claimed that the function of our language changes fundamentally depending upon which language game we are engaged in. Wittgenstein claimed that the fact that words function differently based upon which language games they are embedded in is masked by superficial similarities between language games such as syntax and word choice.[17] Wittgenstein then considers the following criticism:

> 'Here we come up against the great question that lies behind all these considerations.-For someone might object against me: "You take the easy way out! You talk about all sorts of language-games, but have nowhere said what the essence of a language-game, and hence of language, is: what is common to all these activities, and what makes them into language or parts of language…" And this is true.-Instead of producing something common to all that we call language, I am saying that these phenomena have no one thing in common which makes us use the same word for all,-but that they are *related* to one another in many different ways. And it is because of this relationship, or these relationships, that we call them all "language".' (1953, §65).

This objection to Wittgenstein about what unifies 'language' is similar to the one to my own view that we are considering presently, and my response can be similar to Wittgenstein's. To demand a strong set of conditions that unifies all of the concepts in

[17] This interpretation of Wittgenstein may be tendentious but nothing I say will hang upon this. I take these ideas from Price (1992).

question is to demand the wrong thing, because I have been insisting that there *is no* common thread that unifies them.

The second point to make is this: perhaps certain thinkers will say that my position does not deserve the name 'pluralism' because the concepts delineated lack the common unity that they demand. If this is my opponent's view, then I won't quibble about whether my position is correctly characterised as 'pluralism'. This is a verbal disagreement of exactly the sort that I am keen to avoid. I will continue to use the label 'pluralism' for my view, but it can be held whether or not you think the title 'pluralism' is apt.

§4.10–Eliminativism?

As with all of the chapters in this thesis, this one can be read independently from the arguments in the other chapters, or as contributing to the overall position that the thesis advocates. Read in the former way it is the first serious and sustained philosophical analysis of the dominant view of attention in the philosophical literature. Read alongside the rest of the thesis, the chapter is intended to bring together elements from the previous chapters in order to assemble the main argument for pluralism. As I said, this chapter has been a little complex, but I think that it has been worth it in order to discuss these important matters. A worry that has been vocalised in the literature is that pluralism is just eliminativism by another name. This issue will be addressed in the next chapter.

§4A-Appendix: Christopher Mole's 'cognitive unison' theory of attention.

Above I promised a detailed analysis of Christopher Mole's (2011a and 2011b) admirably thorough and metaphysically-oriented theory of attention. This appendix will deliver on this promise. Mole's theory is complex, and we will need to understand certain elements of it before we can have enough clarity to assess his view. I shall first outline his view, then present criticisms of it.

§4A.1-Mole's theory.

Mole explains attention in terms of 'acting attentively' which he very clearly offers necessary and sufficient conditions for:

> "Let α be an agent, let τ be some task that the agent is performing, and call the set of cognitive resources that α can, with understanding, bring to bear in the service of τ, τ's 'background set'. α's performance of τ displays cognitive unison if and only if the resources in τ's background set are not occupied with activity that does not serve τ… α performs τ attentively if and only if α's performance of τ displays cognitive unison." (Mole, 2011a, p.51).

Before we can examine Mole's account, we will have to see how he defines a 'task':

> "A subject's 'tasks'… are the things that the subject is in the business of doing and that she is active with. To specify the tasks in which an agent is engaged, we adopt the agent's point of view on her own activities. Normal human tasks are such things as making a cup of tea, following a conversation, or looking for car keys" (2011a, p.52).

Cognitive resources are the available mental processes which can, if required, be dedicated to a certain task. The set of cognitive resources that can be dedicated to τ are what

Mole calls τ's 'background set'. To Mole, attention can only be deployed when it is in the service of a 'task'. As we have seen, Mole also says that attention can only be deployed when no cognitive resources that can be devoted to a task are devoted to another (different) task.

From this, it follows that if there are two tasks, x and y, and the cognitive resources that could be deployed in order to perform x are the same as those that could be deployed to perform y (and x≠y) then one cannot pay attention to x and y simultaneously, on Mole's theory. This is because, if we deploy any cognitive resources to x whilst we are also deploying some to y, then it will follow that some cognitive resources that *could* be dedicated to x are in fact being dedicated to y, and vice versa. Upon Mole's account of attention, neither would be attended to *at all*. I will return to this point later.

§4A.2-Problems over divided attention.

One criticism that has been levelled at Mole's theory before is that it has difficulty accounting for *exogenous* attention and attentional capture. The problem revolves around the claim that (in certain cases of attentional capture) the subject's attention is engaged despite their not being involved in a 'task' at all. Since Mole's view entails that one can only be paying attention when one is engaged in a task, this is a counterexample to Mole's theory. Watzl (2013) presses this criticism, using the example of meandering attention when one is on a long journey as just such a case where attention is engaged, but one is not engaged in a task. I think there are clearer examples of this kind of phenomenon available, such as being stabbed in the arm. Being stabbed in the arm does engage our attention even though it would strain credulity to claim that 'being stabbed in the arm' is a 'task' in Mole's sense (given the relatively high cognitive demands that Mole puts on being a 'task'). I think that

exogenous attention and attentional capture represent an important problem for Mole's account. However, since these issues have been raised before I will not dwell on them here.

Rather, I will concentrate on giving my own criticism of Mole's theory. We can begin examining the problem I have in mind by making a distinction between *partial* and *divided* attention. It is these kinds of attention that will be helpful for my criticisms of Mole's view. Both of these kinds of attention are recognised by both folk psychological usage of the word 'attention', *and* the usage of the word in empirical psychology.

Partial attention is the phenomenon of being able to pay more or less attention to something, it reflects the intuitive fact that attention *comes in degrees*. So, when reading a philosophy paper, we may skim read it, attempting to extract only the core points and arguments to get a general feel for the paper. Alternatively, we may scrutinise it very closely, think of counterarguments and so on. Both activities seem to involve some level of attention to the paper, but in the former case it seems that we are paying *a lot less* attention to it than in the latter case. Divided attention is the phenomenon of attempting to pay attention to *more than one thing at once.* So, whilst driving we may simultaneously pay attention to the road ahead of us, but also the conversation of our passenger.

Mole is eager to allow for the fact that attention can be partial (2011a, p.83ff.). Say that there is some task (call it A), and that we have a set of cognitive resources (call it β) that we *could* potentially dedicate to performing A. So if we are at a state of full attention in performing A, then we will be using 100% of β. Now consider a case of partial attention. Suppose that we are using 60% of the cognitive resources in β in order to perform A, when we could be using 100%, does this count as a case of attention? Mole's answer is that it depends upon what the other 40% is doing. If the other 40% is not being used *for anything at*

all then it will be true that there are no cognitive resources that could be dedicated to the service of A, but which are dedicated to a task different from A. So in this case (on Mole's account) the subject will count as paying attention to A. However, if *some or all* of the other 40% of β is being used in the service of a task other than A (say B) then it will be true that some of the cognitive resources that we *could* bring to bear on A will be in the service of another task, B. So in this case we will not count as performing A attentively, on Mole's account.

So, Mole's account allows for partial attention because it allows for the fact that different levels of attention can be dedicated to one task. The theory allows that one can still count as paying attention to a task, even when one is not dedicating 100% of the cognitive resources that could be dedicated to the task *so long as the excess cognitive resources that could be used to perform that task are not (partially or wholly) taken up by another task.*

Though Mole's response here can be used to accommodate the idea that attention can be *partial*, it runs into problems when it comes to *divided* attention. We can now begin to construct a problem for Mole's theory. Suppose there are two tasks, P and Q (such that P≠Q). Suppose that the set of cognitive resources that one could use in order to perform P is the same as those that could be deployed in order to perform Q (call this set of cognitive resources δ). So if we performed P with full attention, we would be using 100% of δ in the service of P. Alternatively, if we were performing Q with full attention, we would be using 100% of δ in the service of Q. If a subject performs P, and dedicates only 10% of the cognitive resources in δ to the service of P, but *does not do anything at all* with the remaining 90% of δ, then by Mole's theory she will count as paying attention to P. Even though very few cognitive resources are dedicated to P (and she is presumably performing the task extremely

haphazardly), she will still count as paying attention to P, because there are no available cognitive resources that could be used to perform P, but which are dedicated to anything that is not the performance of P (because the remaining 90% of δ are doing nothing at all). Call this case 1.

Now consider another case, case 2. In case 2 the subject is dedicating 95% of the cognitive resources in δ to the service of P, and the remaining 5% of δ is dedicated to the performance of Q. According to Mole's theory, in case 2 the subject will *not* count as paying attention to P at all, even though the amount of cognitive resources that are being dedicated to P in the second case is vastly more than in case 1. This is because in case 2, some of the cognitive resources that *can* be dedicated to P are in fact being used in the service of Q. So the upshot is that in case 1, whilst the subject is dedicating only 10% of δ to P, the subject counts as paying attention to performing P, but in case 2, she is dedicating 95% of δ to P, but Mole's theory counts her as not paying attention to P at all.

A concrete example will help. Imagine that in case 1, a subject is driving her car very haphazardly, only dedicating 10% of her cognitive resources to driving the car, but that she is not doing anything with the remaining 90% of her cognitive resources. Normally we would describe this as a case of inattention (after all, she is not really concentrating on driving the car). Contrary to this, Mole's theory counts her as paying attention to the driving. In case 2, she is dedicating 95% of her available cognitive resources to driving, and is as a result, driving far more carefully and safely than in the first case, but she is also using some of her remaining cognitive resources that could serve the task of driving in the performance of another task (for example, she may be vaguely thinking about her lunch). By the above reasoning, Mole's theory rules that she *is* paying attention to driving in the case 1, even

though she is driving so shoddily, but she is *not* paying attention *at all* to the driving in case 2, even though she is driving far more carefully and with a greater supply of cognitive resources than in case 1.

Now, this clearly seems to be the wrong result. It seems extremely odd that in case 1 we can dedicate only a few cognitive resources to a task, and perform the task extremely shoddily and haphazardly, and yet still count as paying attention to the task, and then in another case dedicate vastly more resources to the task, and perform it much better and with greater accuracy and so on, and yet in the second case we will not count as paying attention to this task at all (by Mole's theory, case 2 will not even count as a case of partial attention). It seems that the correct result is that the subject should count as paying much *less* attention (if any) to the driving in the first case, and paying much *more* in the second case, so it looks as though Mole's theory fails by criterion (1), as it delivers what seems clearly to be the wrong result on this occasion.

§4A.3-Replies.

Perhaps Mole's best response to the problem of divided attention is to claim that in case 2, the subject is not paying attention to P or to Q at all, but she is actually paying attention to some 'wider' task, which is the conjunction of the two tasks: (P&Q). Mole says some things along these lines (2011a, pp.81-2.).

There are various problems with this reply. Firstly, it appears *ad hoc.* Secondly, the reply does not really address the worry that we originally had with Mole's theory. The problem is that Mole's theory must still count the subject as paying attention to P in the first case, and not paying attention to P at all in the second case, even though in the second case she is using far more cognitive resources in the service of P. This worry would not really

have been addressed simply by introducing another possible candidate for what the subject may be paying attention to. The problem began because when asked whether the subject is 'paying attention to P' in the second case, Mole must always answer 'no', and this seems implausible.

A third worry with this reply is that it seems very odd to claim that the subject is paying attention to (P&Q) in case 2, but also to *deny* that she is actually paying attention to P or to Q individually at all. This response would force Mole to say that in case 2, the subject was paying attention to the conjunctive task (driving & thinking about dinner) but that she was not actually paying attention to driving, or paying attention to thinking about dinner *at all*. This is not an outright logical inconsistency, but it is certainly an extremely peculiar result.

§4A.4-Mole's view and Lavie's 'perceptual load' theory

I have put forward my main argument against Mole's view. This argument has been relatively a priori, but we can draw upon empirical psychology to develop another argument against Mole's view which can stand independently of the main objection described in §§4A.2-4A.3. The empirical work is Lavie's 'perceptual load' theory of attention (Lavie, 2005 and Lavie and Yehoshua, 1994I shall argue that if we take Lavie's theory seriously, then Mole will be committed to the claim that (in at least a large number of perceptual tasks) attention cannot be *partial*. I take this to be implausible and thus problematic for Mole.

An extremely crude summary of Lavie's basic claim is that when subjects have to complete a perceptual task which places high demands on processing in the cognitive system, then *all* available attention will be dedicated to that task, and the subjects will fail to

notice task irrelevant stimuli. However, when the perceptual task is simpler, and requires less processing power, then only some attention will be dedicated to it, and the remaining attention will *automatically* be siphoned off to monitor task irrelevant properties of the environment. The main argument in favour of this view is that it neatly accounts for apparently contradictory results which arose out of the 'early' versus 'late' selectionist debates in the mid to late twentieth century (see Lavie, 2007 for more on this).

Lavie's theory is intended to generalise: the view is that *whenever* perceptual attention is not fully dedicated to a task, some of the excess resources will be siphoned off, and the subject will pay a little bit of attention to some things outside of the task. Indeed, not only is this theory empirically supported, but it seems extremely intuitively plausible as well: if the task you are doing does not require full attention, it makes sense to use some of the excess attentional resources to monitor the environment.

If we take Lavie's theory seriously, this is problematic for Mole, because it would entail that (in at least a large number of perceptual tasks) that if attention is not *fully* dedicated to the task, then some cognitive resources that *could* have been dedicated to the task will be siphoned off to perform another task (monitoring task irrelevant properties of the environment). But if some cognitive resources that could be used to perform a certain task are dedicated to another task then Mole's theory must conclude that *neither* task is attended to *at all*. So Mole must conclude that (in these particular perceptual tasks) either attention to a task must be total, or it cannot exist at all. Attention cannot be partial. This is because (on Lavie's theory) if attention is not total then at least *some* excess attentional resources are dedicated to another task. But once this happens, then (on Mole's view) the

subject counts as paying attention to *neither* task. This seems wrong, and it certainly goes against what Mole wishes his theory to entail.

Chapter 5
A Road to Eliminativism?

§5.0-Summary.

In this chapter, pluralism will be defended from the claim that it leads to eliminativism. Firstly, one strategy for avoiding the slide to eliminativism will be rejected. Then it will be argued that pluralism does not imply eliminativism of any threatening kind; but that we do have reason to embrace what I call 'pragmatic eliminativism' about attention. The scope of this chapter will discuss not only 'attention' but also other terms that have been given similar analyses to the one that 'attention' has received in this thesis.

§5.1-The eliminativist challenge.

It may be that someone could raise a question about the overall position that I am advocating, which is that it leads straight to *eliminativism*. If what I am really saying is that 'attention' fragments into several importantly different concepts, and that no one concept is privileged above the others, then really I am just an eliminativist in pluralist clothing. This kind of comment on my view could be levelled in at least two ways. Firstly, someone who was sympathetic to my general picture of attention could say that my arguments, far from being *wrong* do not in fact *go far enough.* This kind of view would be sympathetic to mine in spirit, but draw a more extreme conclusion from what I have been saying.

This picture is similar to what we get from the psychologist Alan Allport's famous review of the psychological debates over whether attentional selection in the auditory system occurs at an 'early' or 'late' stage. Allport claims that 'there is no one uniform computational function, or mental operation in general' that we can identify with attention, and thus that 'qua causal mechanism, *there can be no such thing as attention'* (1993, p.203).[1] Here we see Allport moving from the pluralistic claim that 'attention' is not identical with

[1] Though Allport gives a different view in his (2011).

any one uniform mechanism to an eliminativist claim that it does not exist (or at least, that it does not exist 'qua causal mechanism').

The threat of eliminativism could also be used by an opponent of pluralism as an attack on pluralism itself. This style of argument would proceed by claiming that my arguments aim to establish pluralism about attention, but that pluralism implies eliminativism, and eliminativism is implausible. Therefore, my arguments in favour of pluralism should be rejected.

We find objections along these lines in the literature in various places. Prinz says this:

> '[t]here may be a common denominator [which applies to all and only instances of attention] that can be empirically discovered. If such a common mechanism were found, we might say that "attention" refers to that mechanism. If these phenomena share nothing in common, then we might say that "attention" should be dropped as a term from scientific psychology. We might become eliminativists' (2012, p.91).

Though Prinz uses eliminativism as a reason to avoid pluralism, he provides no argument to show why eliminativism is such a bad thing. It seems to be simply assumed that if we can avoid eliminativism about attention then we should. Over the course of this chapter, I will examine the question of whether eliminativism about attention is something to be avoided, but what I am primarily interested in is the background assumption that is common to Prinz's and Allport's claims, which is that pluralism about attention implies eliminativism (or, at least, that there is a significant threat of eliminativism in the vicinity of pluralism).

§5.2-The argument from analogy.

I addressed the threat of eliminativism in Taylor (2013a) and offered an argument that was intended to show that pluralism about attention does not imply eliminativism. My argument was similar to one employed by De Brigard (2012) which is intended to show something similar. I now think that this argument does contain an important point, but that it is at best suboptimal for resisting the eliminativist threat. I will now outline the argument, then explain what I take to be the main problem with it.

The argument works by analogy; it is argued that pluralism about attention need not imply eliminativism about attention, because there are several other psychological faculties that we can happily accept pluralism about, but which we are nonetheless not eliminativists about. The example that De Brigard and I used was memory. It was argued that in contemporary psychology, the concept 'memory' fragments into several distinct sub-concepts, which refer to different systems in the brain, and these systems are significantly different from each other.

In support of this, I pointed out that different memory subsystems operate in different areas of the brain (for example, pure iconic memory[2] operates in the retinae, whilst long term memory operates primarily in the hippocampus). Importantly, different memory subsystems also operate using different neural mechanisms. Pure iconic memory works by using retinal cells that continue to fire after the eye has been closed, whilst certain kinds of long term memory work by increasing the sensitivity of neurons in the hippocampus to stimulation by other neurons (Kandel, 2002). So, the concept 'memory' has bifurcated into several importantly distinct subconcepts.

[2] I say 'pure' iconic memory because it is now known that 'iconic' memory itself bifurcates into two distinct subsystems. See Sligte et al. (2008 and 2009).

The argument by analogy then concludes with the claim that despite the fact that many psychologists can plausibly be called *pluralists* about memory, we are not thereby eliminativists about memory. The term 'memory' has not been eliminated, and it is not as though we believe that memory does not exist. Therefore, there is no reason to think that if 'attention' similarly fragments, then we must conclude that 'attention' does not exist, any more than it is true that 'memory' does not exist. This was the argument that I employed to resist the eliminativist challenge.

I am now less optimistic about this argument. In the argument, an analogy with memory is used to resist eliminativism, but the problem is that an analogy with memory has also been used to *support* eliminativism about other psychological faculties. For example, in the course of arguing for eliminativism about 'concept' Edouard Machery uses 'memory' as an example of a term that *has* been eliminated from psychology (2009, p.237). The issue is that when De Brigard and I first deployed the argument from analogy, we assumed that eliminativism about 'memory' was false, and then used the analogy with memory to argue that we need not accept eliminativism about 'attention' either. But Machery's argument assumes that eliminativism about 'memory' is *true* and is thus able to use it as an analogy for his case *in favour* of eliminativism about a certain term (in his case, the term 'concept').

I am *not* saying that eliminativism about 'memory' really *is* true, rather I am saying that the status of 'memory' in psychology is too uncertain for either party to use it to support their own view. Each argument will only work by assuming that 'memory' either has or has not been eliminated, but no argument is given by either side to accept their view on what the fate of 'memory' has been. Without such an argument, the argument from analogy is no more powerful than the other argument in precisely the opposite direction.

A plausible conclusion to reach here is that the 'memory' case is more complicated than Machery, De Brigard or myself initially realised. Plausibly, there is a *sense* in which 'memory' has been eliminated, and a *sense* in which it has not. In my view, we cannot adjudicate this matter until we get clearer on what we mean by 'eliminativism'. Then we will be able to judge the attention case and the memory case together. I shall perform the required work over the rest of this chapter.

There is, however something importantly correct about the argument from analogy, which is that it is correct to claim that whether we are pluralists or eliminativists about 'attention', either way it will be in the same boat as 'memory'. 'Memory' fragments into importantly distinct sub-concepts which refer to different things and (if what I have been saying in this thesis is roughly correct) so does 'attention'. This may help to ease some of the initial resistance to my central claims, as I place attention on the same plane as other psychological faculties that have succumbed to the same analysis. This kind of heterogeneity is discovered from time to time, and should not surprise us. Indeed, if much of what I will say in this chapter is correct, the issue is widespread, and occurs with many scientific and philosophical terms, so if this analysis also applies to 'attention' then we should not be surprised or worried.

§5.3-Some similar debates.

Similar argumentative structures as I have been putting forward in this thesis can be found with relation to other debates in philosophy. It is my hope that much that I have to say about 'attention' will apply to these debates as well, and that philosophical progress can be made by looking at these similar debates together. I certainly cannot hope to examine any of these other debates in anything like the detail that they deserve, but I hope to be able to

draw out some similarities that many of the debates have, such that our discussion of attention can inform many of them.

I shall use the label 'heterogeneity view' to refer to any position which takes a pluralist stance toward a certain concept or entity. This is intended to *include* those thinkers who take such a pluralist view further, and draw the conclusion of eliminativism from it. This is a crude categorisation (as there are a great many differences between different sub-varieties of these views) but the term captures what is common to all such views: in response to the question 'what is X?' the heterogeneity theorist responds by saying 'X is many things'.

There are various routes to heterogeneity views. I have already mentioned that throughout the history of biology, the term 'species' has often been defined importantly differently by ecologists, phylogeneticists and those interested in *interbreeding.* (see, e.g. Claridge et al. 1997, Ereshefsky, 1992a, 1998, 2010b and 2010c and Wilkins, 2009). Recently, attention has been paid to the way that 'species' is used in *microbiology* (Ereshefsky, 2010d and Ereshefsky and Reydon, forthcoming). Many of the subconcepts used in this debate divide into further subconcepts (Ereshefsky, 1992b). This has led some to embrace pluralism about the 'species' concept. Unsurprisingly, varieties of species pluralism come in a great many different strengths, from relatively conservative varieties, such as that of Kitcher (1984) and Brigandt (2003) to extremely liberal varieties (Dupré, 1993).

In reaction to this, some thinkers (what can be called 'monists') have argued that one of the 'species' concepts is 'privileged' in some way. Ghiselin (1987 and 1989) and Lee (2003) take this line. At the other end of the spectrum from this monist theory is eliminativism about 'species', which is argued for by Ereshefsky (1992a and 1998). Ereshefsky sees himself

as a pluralist as well as an eliminativist (he doesn't see pluralism and eliminativism as mutually exclusive options). Interestingly, Ereshefsky recently has changed his mind, and argued that the word 'species' should be retained in evolutionary biology (Ereshefsky, 2010c). I will be examining Ereshefsky's argument for this change of mind in §5.6.3.1.

It will not have escaped the reader's awareness that there is another prominent term in the vicinity that has also been subject to a heterogeneity view, which is 'consciousness'. Ned Block describes consciousness as a 'mongrel' concept, and claims that it is ambiguous between several importantly distinct subconcepts (Block, 1995a). Block's main argument proceeds by claiming that Phenomenal-consciousness (P-consciousness), the 'what it is like-ness' is conceptually distinct from Access-consciousness (A-consciousness), which applies to mental events that are poised for direct rational control of action and thought.[3] He then claims that currently prominent theories of consciousness have often claimed to explain P-consciousness, but really have only succeeded in making progress on explaining A-consciousness. Block has argued that P-consciousness and A-consciousness are *conceptually distinct* (1995a, 1995b, 1996, 1997 and 2002) and also that they are empirically dissociable (esp. 2005, 2007a, 2007b, 2011, 2012 and 2014b).

Resistance to Block's arguments comes in a variety of flavours. Some have claimed that the two kinds of consciousness are conceptually distinct but that they do not come apart in reality (Baars, 1995 and 1997; Stazicker, 2011 and Brown, 2012). On the other hand, Burge (1997) has accepted Block's claim that there is a distinction here but claims that P-consciousness is more deserving of the title 'consciousness' than A-consciousness is. Finally,

[3] In later work, Block tends to talk of 'phenomenology' and 'cognitive access' rather than P-consciousness and A-consciousness.

some philosophers (Clark, 2000a, 2000b, Dennett, 1995 and Crane, 2013) argue that the two concepts are conceptually connected in some important sense.

The concept 'consciousness' is further ambiguous. Rosenthal (1986, 1997 and 2005) has long emphasised the difference between transitive consciousness (where one is said to be conscious *of* something) and intransitive consciousness (where one is said simply to be conscious). Another division made by Rosenthal is between state consciousness (as when an experience such as a pain is conscious) and creature consciousness (as when a subject such as a person is conscious). O'Shaughnessey (2002) emphasises the difference between *wakeful* consciousness and *sleeping* consciousness (where the latter includes dreams). This ambiguity can be taken yet further. In his (2009) book on consciousness, Christopher Hill separates 9 different meanings of 'consciousness' (ch.1).

§5.3.1-The general issue.

I have highlighted four terms ('species', 'concept', 'emotion' and 'consciousness') that have been subject to some kind of heterogeneity view, but this is just the tip of the iceberg. Two further heterogeneity views that will be examined in the next chapter are those that revolve around 'concept' (Machery, 2009) and 'emotion' (1997). In addition to these views, heterogeneity views of one form or another have been raised in relation to 'scientific theory' (Vickers, 2008, 2013 and 2014 and Magnus, 2012); 'innateness' (Griffiths, 2002); 'imagination' (Kind, 2013); 'logic' (Beall and Restall, 2000); 'art' (Uidhir and Magnus, 2011); 'value' (Wiggins, 1998); 'person' (Teichman, 1985) and 'life' (Machery, 2012). We could continue this list *ad nauseum.*

One of the most interesting features of these debates is that often, pluralists and eliminativists agree with each other on most of the main points of the argument. Both accept

the heterogeneity claim, and both accept that much of the trouble in the debate is caused because debaters assume that they are talking about the same thing when in fact they are not. It is thus not surprising that eliminativists and pluralists are often lumped together.

This raises a question: is there really a substantial difference between the eliminativists and the pluralists? In some sense, this question is an important one and points out the need to be careful not to descend into a verbal dispute about what exactly does and does not count as 'eliminativism' or 'pluralism'. Nonetheless, we should not simply lump together pluralism and eliminativism as the very same doctrine, as there are a variety of different available views on the table, and which one we embrace will dictate our ontological commitments and our general methodological approach to the debates that are under scrutiny. It is untangling the knot of these views that I will now attempt.

§5.4-Two kinds of eliminativism.

We should start by distinguishing what I call 'ontological eliminativism' and 'pragmatic eliminativism':

> Ontological eliminativism about x is the view that x does not exist.
>
> Pragmatic eliminativism about a given term 'x' is the view that the term 'x' should be eliminated from a particular debate, discipline or set of practices.

Ontological and pragmatic eliminativism are clearly quite different views. We accept ontological eliminativism about certain entities because we doubt whether those entities actually exist. When we examine the motivations for pragmatic eliminativism, we get a very different story. We would be pragmatic eliminativists if we simply thought that there were advantages to avoiding the use of a certain term, and if we thought that the advantages of

avoiding the use of the term in a certain context outweighed the disadvantages. In addition to this difference in motivation, there also seem to be clear cases where one can hold one version of eliminativism, but reject the other. Physicists often speak of frictionless planes and economists sometimes speak of the perfectly rational person, even though the same thinkers would insist that these entities do not really exist. This seems to be an example of ontological eliminativism without pragmatic eliminativism.

Despite these differences between ontological and pragmatic eliminativism, they have often been merged together in the literature. This is especially clear with Machery's argument for eliminativism about 'concept'. Machery says this:

> '"Concept" fails to pick out a natural kind, although it has been assumed by many psychologists to do so. On pragmatic grounds, I conclude that "concept" ought to be eliminated from psychology' (2009, p.246).

From this quotation, it sounds as though Machery is advocating pragmatic eliminativism. It does not seem appropriate to infer that he thinks that concepts *do not exist*, at least not without much more argumentation. However, when discussing this view, Machery is often taken to be a pragmatic *and* ontological eliminativist. Thus, Prinz:

> 'Rather than proposing a new theory to usurp the others, Machery argues that the extant debates are really spurious. On his view, concepts don't really exist' (2010b, p.612).

Equally, when discussing Machery, Margolis and Laurence say this:

> 'These same considerations have led some theorists to advocate *concept eliminativism*—the view that there are no concepts' (2012, §2.5).[4]

I think it is a mistake to lump together ontological and pragmatic eliminativism, and I shall proceed with these distinctions clearly in place. We can now divide the question of whether pluralism implies eliminativism into two questions: whether it implies ontological eliminativism and whether it implies pragmatic eliminativism. I turn to the former of these questions first.

§5.5-Resisting the slide to ontological eliminativism.

In this section I argue that pluralism does not imply ontological eliminativism: pluralists about the 'attention' concept can be realists about attention itself. Much of this section will assume particular metaphysical views concerning truthmaking, as well as some views about language. I take these to be independently plausible, and I will not argue in favour of them directly. If readers reject these views, then this section can be read as a conditional: *if* one accepts these views, *then* one can avoid the slide to ontological eliminativism. My focus is 'attention', but much that I have to say will apply to the other terms I mentioned above.

I said that pluralists can be realists, but what is 'realism' anyway? The question of what realism *is* has generated much complex metametaphysical debate, and assessment of this issue would carry us too far afield. I am inclined to consider the debate over what realism is to be verbal, rather than substantive. For this reason, I proceed by stipulation:

[4] An exception is Weiskopf (2009, p.162) who does not misinterpret Machery.

when I say that we can be *realists* about 'attention' I mean that claims about attention can and often do have *truthmakers.* I now turn to explaining and defending this claim.

§5.5.1-Step one: predicates, properties and truthmakers.

We should start with a distinction between predicates and properties. Predicates such as 'red', 'tall' and 'spherical' are linguistic entities which we use to talk about the world. Predicates *describe* entities. Properties by contrast (at least in the sense that I intend) are features of the world which characterise the entities that exist. For example, a particular electron may have a certain mass, or charge: these are properties of the electron. I will assume realism about properties. Exactly which ontological account of properties one gives will not matter for present purposes (see Armstrong, 2005 for a survey of the available options).

Clearly, there is an important relationship between predicates and properties. When I describe an object using predicates, my description is often *about* the properties of the object. But how best can we elucidate the link between predicates and properties? Here is one particular way, which I take from John Heil:

Φ: 'When a predicate applies truly to an object, it does so in virtue of designating a property possessed by that object and by every object to which the predicate truly applies (or would apply)' (Heil, 2003, p.26).

When we accept principle Φ, the relationship between predicates and properties becomes transparent. On this view, if I truly predicate sphericity of a ball, then the predicate truly applies to the ball in virtue of the property of sphericity that the ball instantiates, and which *every* object that can truthfully be called 'sphericity' instantiates as well.

As simple as this view is, many metaphysicians have held that Φ is implausible, and with good reason.[5] Consider the implications of Φ. Take for example, the predicate 'being non-self-exemplifying', which appears meaningful (example is from Lowe, 2006). If we accept Φ, and also accept that 'being non-self-exemplifying' is a meaningful predicate that can apply truly to objects, then we are led into a Liar-like paradox, because if there is a property of being non-self exemplifying, then it must be non-self-exemplifying. But if it is non-self-exemplifying then it must be self-exemplifying, and so on.

Some readers may be impatient with baroque examples of 'non-self-exemplification' but the point will generalise to more quotidian predicates. Take, for example, the predicate 'red'. Tomatoes, balls and fire can all be 'red', they can all have 'red' truly predicated of them but it seems unlikely that we will find exactly one property shared by all of these entities in virtue of which they are all 'red'. Rather, they all have slightly *different* properties in virtue of which they are red.

Suppose we reject Φ, and reject the claim that for every meaningful predicate there must correspond a property and only one property that is shared by all of the entities that the predicate truly applies to. What positive proposal can we replace Φ with? Here I think we should take seriously Heil's suggestion, which is that we should accept that predications can be true of objects, and what's more, that they can be true *in virtue of the properties of those objects* without the requirement that all and only the objects that the predicate refers to must instantiate one property or set of properties in virtue of which the predicate refers to the object.

It will be helpful to quote Heil's alternative to Φ. He says:

[5] See e.g. Ellis (2001), Heil (2003), Lowe (2006) and Mellor (1993).

> 'It is in virtue of objects' properties that the predicates apply, but significant predicates need not pick out a unique property common to all objects to which they apply' (2003, p.47).

Heil believes that it is in virtue of the *properties* of objects that claims about those objects can be true, even if the objects do not share one property or set of properties that are instantiated by all of the objects that the predicate in question can truly apply to:

> 'while our predicates or concepts often apply-literally and truly- to the world, this need not be taken to imply that these predicates and concepts designate properties shared by everything to which they literally and truly apply' (Heil, 2003, p.58).

To help us get Heil's view more clearly in focus, we can introduce the concept of a *truthmaker*. Truthmakers are the entities in the world in virtue of which certain claims about the world are true. So, if I say 'the cat is on the mat' then the truthmakers for this claim would be the cat, the mat and the particular relation between them.

With this in place, we can couch Heil's suggestion in terms of truthmakers. We can say that the predicate 'red' truly applies to a ball, a tomato and fire, and we can have *truthmakers* for claims such as 'these entities are red' without the need to claim that they must all share one property or set of properties in virtue of which they satisfy the predicate 'red'. To be clear: it is the *properties of the fire, the ball and the tomato*[6] that serve as truthmakers for the predication of 'red' but it does *not* follow that there must be *one* set of properties that

[6] More strictly: it will be properties of the electrons that form the electron shells of the component atoms of the fire, the ball and the tomato,(perhaps together with properties of the radiation they emit) that will serve as truthmakers for predications of 'red'. Whether you think that colour is intrinsic to objects such as tomatoes will affect what you include in the set of truthmakers for colour predications. For example, if you think colour is a Lockean secondary feature, then certain properties of our sensory systems (such as the cells in our retinae) will also be included in the set of truthmakers for colour predications. Nothing of substance I have to say will turn on this issue.

are shared by all three entities that serve as truthmakers for predications of 'red'. Rather, they instantiate different properties which are more or less similar, and these *similar* properties serve as truthmakers for predications of 'red'.

A note on truthmaking: I take truthmakers to be entities in the world that make certain claims true. I remain silent on whether the truthbearers (the things that are made true) are sentences, utterances or propositions. I hope to remain neutral on most of the controversy surrounding truthmaker theory (see Armstrong, 2004).

§5.5.2-Step two: 'Attention' and realism.

With these ideas in place, we can begin to examine how this relates to issues to do with the view of 'attention' that I have been defending. Suppose one accepts that 'attention' is ambiguous between several different concepts, and that no one of them is privileged. If we accept this, then different utterances of the word 'attention' can have different meanings depending upon which subconcept of 'attention' one has in mind. There is nothing odd in this claim: it plausibly applies for all ambiguous terms. When a subject uses the word 'hard' meaning 'difficult to do' then her utterance will take on a different meaning from if the subject used the word to mean 'impenetrable' or 'dense' (example from Wilson, 2006). So, it seems overwhelmingly plausible that token distinct utterances of the same word can take on different meanings depending upon what meaning the interlocutor has in mind.[7] This is the same with 'attention': if the subject has the 'selection for action' subconcept in mind, then their utterance will take on a different meaning from if they have the 'cognitive unison' subconcept in mind.

[7] In cases of semantic externalism, different factors will apply, I address semantic externalism below.

Suppose we have two token distinct claims, each of which makes use of the word 'attention', but where the meaning assigned to the word 'attention' is different in each case. If these different meanings are anything like the concepts of 'attention' that we have been examining in this thesis then they will have different application conditions, and therefore the truthmakers for the claims in question will be different. Again, this is not a surprising phenomenon: it plausibly applies to all claims that involve ambiguous words. For example, the truthmakers for token distinct utterances of 'X is hard' will be different depending upon what meaning one attaches to 'hard'. On one meaning the claim will be true if X has certain properties in virtue of which it is difficult; on the other meaning the claim will be true if X has certain properties in virtue of which it is dense, unscratchable or impenetrable.

The same can apply to the 'attention' case. Take a claim such as 'Frank is paying attention to this exam'. On one concept of 'attention', the truthmakers for the claim will be the properties of Frank in virtue of which he is selecting information about the exam for action; on another concept the truthmakers for the claim will be the properties of Frank in virtue of which the subject is displaying cognitive unison at the task of performing the exam.

Often the meaning assigned to the word 'attention' will be obvious from the context of the utterance, and what it would take for the claim in question to have truthmakers will be correspondingly obvious. For example, if a teacher shouts at a child, saying 'you're not paying attention' then it is clear that she means that the child is not listening to her in a manner appropriate for him to learn what she is trying to teach him. The truthmakers for her claim would be the properties of the child in virtue of which he is failing to listen to her. Likewise, if an empirical psychologist using the Posner paradigm claims that 'the subject is paying attention to the stimulus' then it is clear that she has in mind some concept of

attention that fits in with how the Posner paradigm is supposed to measure attention, and the truthmakers for such a claim would be the properties of the subject in virtue of which they have increased task performance in valid conditions.

These points about context can be taken further to accommodate cases of semantic externalism. In cases of semantic externalism, the meaning of the word will not be (entirely) fixed by what is intrinsic to the subject, but will be (in part) fixed by something external to the subject such as referential relations that the term bears to the external world (Kripke, 1980 and Putnam, 1975) or the particular community in which one makes the utterance (Burge, 1979). If you have sympathy with semantic externalism, then one can substitute my use of the phrase 'whatever subconcept the subject has in mind' with whatever factors one thinks determine meaning, and the spirit of what I have to say will remain the same.

Whether the meaning of the word 'attention' in a particular utterance is fixed by the meaning that the particular thinker has in mind, or by some external relation, it will remain true that different meanings that are assigned to token distinct utterances of 'attention' will have different application conditions. As a result, two claims which contain the word 'attention', (where the meaning of the word is different in each case) will differ in the set of properties that can serve as the truthmakers for the claim (in a nutshell, this is because the truthmakers are determined by the application conditions of the concepts, and the application conditions are different depending upon which meaning the term has).

If we accept this kind of picture, then we get truthmakers for claims about attention, and this forms the basis for realism about attention. This kind of view gives us realism that has some breathing space: we get realism about attention (because we get truthmakers for claims about attention) but we need not put unreasonable demands on what realism

requires. Once we reject principle Φ, and reject the view that there must be one single property or set of properties that can serve as truthmakers for each predicate, then we can embrace the claim that the truthmakers for a certain claim concerning 'attention' can be different from the truthmakers for a different claim containing 'attention' where the meaning of the term is different. The fact that different sets of properties can serve as truthmakers for claims containing the same predicate should not worry us: it is plausibly the case even with quotidian predicates such as 'red'.

Obviously, an opponent could question whether what I have offered really is deserving of the title 'realism'. Such an opponent could choose to reserve the title 'realism' for something more demanding and austere than my view. If readers are unhappy labelling my position 'realism' then my view can be stated without mention of it: my view is that claims containing the predicate 'attention' can be and often are *true* and this is because they can and often do have truthmakers. That is good enough for me.

I am willing to grant the use of the word 'realism' to an opponent, but I will say that if we do have an austere view of what counts as 'realism' then we will likely end up being anti-realists about a great many things. As I argued above, it is plausible that many of our predications are made true by different sets of properties. If we insist that such a view does not deserve to be labelled 'realism' in the case of attention, we must presumably insist the same when it comes to other more everyday predications, and we will end up being anti-realists about the entities that these predications are about also. Of course, such views have been held (e.g. Ladyman and Ross, 2009) but I am not concerned with rebutting global views like this here. Rather, I am only concerned with resisting the claim that there is something particular about my view on 'attention' that leads us to ontological eliminativism.

§5.6-Pragmatic eliminativism.

I have argued that pluralism need not imply ontological eliminativism. We can now consider the claim that pluralism implies pragmatic eliminativism. Recall that pragmatic eliminativism is the view that a certain term should be dropped from a certain set of practices or debates. An opponent may claim that even if a pluralist can avoid ontological eliminativism, pluralism still entails pragmatic eliminativism which (it may be said) is bad enough.

We should firstly note that the kind of pluralism that I have put forward is *consistent* with the denial of pragmatic eliminativism. One can perfectly consistently argue that the word 'attention' is ambiguous between several importantly distinct concepts, but insist that we should still use the word 'attention' in all of our practices. So, a pluralist does not *have* to embrace pragmatic eliminativism. It is important to bear this point in mind.

However, even if we do not *have* to commit to pragmatic eliminativism about 'attention', are there any good positive reasons to do so? Indeed, some thinkers *have* held the view that the term 'attention' should be avoided. Mole (2009) mentions Donald Broadbent as an advocate of this view. In a 1982 paper on the topic, Broadbent says this:

> 'The topic of this paper is one that is often termed 'attention' and it may seem unduly artificial to have given it a more cumbrous title. 'Attention' is a word in ordinary language, and can reasonably be used as a label for experiments in a particular area. Yet it has also been used on occasion as a theoretical concept, a mysterious asset or energy which is sometimes attached to human functions and sometimes not. This use of attention as a theoretical concept… is not very helpful and avoiding the word in the title is a step toward clarity.' (p.253).

Throughout this article, Broadbent tends to only use the term 'attention' either when discussing the work of others, or by placing the term in scare quotes to indicate that he does not advocate the use of the term (Broadbent avoids use of the term in his (1958) work as well). For example:

> 'Controlled processing… interferes with other tasks even when no action is required; and is therefore seen as requiring some further resource ('attention'?). These concepts have been very widely influential; they are beginning to appear in quite elementary texts… Some doubts will be expressed about them later' (1982, p.278).

So, Broadbent advocates pragmatic eliminativism about 'attention'. However, in these passages, beyond vague comments that attention is a 'mysterious asset or energy' Broadbent does not really give any clear reasons in favour of pragmatic eliminativism. Without further argument, Broadbent's assertions are far from convincing. Nonetheless, it is my view that Broadbent's basic intuition was correct: the term 'attention' should indeed be avoided, at least in a great many theoretical contexts. In the remainder of this section, I will develop and defend an argument for this claim. In the process, I will refine the version of pragmatic eliminativism that I advocate, because pragmatic eliminativism is itself a position that can come in a variety of strengths, and there are important differences between the different versions that one could opt for.[8]

[8]The discussion in this section is indebted to Chalmers (2011) and Vickers (2014), though neither of them discuss 'attention', and their views differ from mine in various ways. Also see Sidelle (2007) and Balcerak Jackson (2014) for more on these issues.

§5.6.1-The argument for pragmatic eliminativism.

I call the argument for pragmatic eliminativism the 'pragmatic argument':

Pragmatic Argument.

1) The word 'attention' is used in different ways by different thinkers in various debates concerning 'attention', and this is causing confusion and difficulty within these debates.

2) Pragmatic eliminativism can help resolve these problems, and it has additional benefits as well.

3) The advantages of pragmatic eliminativism significantly outweigh the disadvantages.

4) (Therefore) we should adopt pragmatic eliminativism.

I will assume that premise (1) is plausible. Chapter 2 was intended as a detailed collection of case studies to demonstrate this. Notice that one need not accept pluralism itself in order to accept premise (1). For these reasons, I will dedicate most effort to defending (2-3).

§5.6.2-Premise (2).

The point of pragmatic eliminativism can often be missed. In my view, it is misunderstanding the position that has led to a great deal of resistance to it. It is *not* meant as a blanket ban on all disputes in which the concept 'attention' has featured. Rather, the point is that *the main claims of such disputes would be better stated without mention of the term 'attention'*. So, we can still have many of the debates that we wanted to have but we should do so without mention of 'attention'. When we encounter a debate that involves a claim that

makes mention of 'attention', pragmatic eliminativism would involve substituting the word 'attention' with what that particular thinker means by the word 'attention' and then restating the claim in those terms.[9] We then preserve the substance of the thinker's claim, but it is now impossible to become engaged in debates which arise as a result of different interlocutors having different views on the meaning of the word 'attention'.

We can get an idea of this advantage by returning to one of our main case studies, which is the dispute over whether GY is attending to the stimuli in question or not. Suppose we take the main claims of the thinkers involved in this debate, and rephrase them in terms which do not contain the word 'attention'.

Kentridge claims the following:

i) GY was attending to the stimuli in question, but was not conscious of them.

We can now apply the pragmatic eliminativist approach to Kentridge's claim, and substitute the word 'attention' for what Kentridge himself takes the word to mean:[10]

i*) GY was (voluntarily or involuntarily) prioritizing information in a selected part of the visual scene. But GY was not conscious of the stimuli in question.

De Brigard and Prinz's claim can be written as follows:

ii) GY was not attending to the stimuli in question, and was not conscious of them.

When we apply the pragmatic eliminativism method, this becomes:

[9] This draws on the 'method of substitution' that Chalmers (2011) discusses.
[10] As I said in chapter 2, Kentridge puts forth two definitions of attention. The pragmatic eliminativist method can be applied using his other definition as well.

ii*) Information about the stimuli was not available to GY's working memory, and he was not conscious of the stimuli.

Smithies' claim is the same as (ii) above, but when we substitute his account of attention into his claim it becomes:

iii*) Information about the stimuli were not rationally available to GY as a justifying reason for action and thought, and GY was not conscious of the stimuli.

Now, when we compare (i*), (ii*) and (iii*) we can see that they do not contradict each other. Once we state the claims without use of the word 'attention', the friction between the different positions disappears. The method of pragmatic eliminativism makes this fact salient, and it means that the interlocutors are no longer able to apparently disagree as a result of differing views over the meaning of the word 'attention' (because 'attention' does not feature in their claims). Notice that the substance of each thinker's claim has been retained. Now the interlocutors can continue their debate, without worrying about different meanings of 'attention'. If what I have been saying in this thesis is correct, the term 'attention' is ambiguous in certain crucial ways, and this fact invites methodological confusion and cross-talking when we employ the term. Pragmatic eliminativism halts these problems in their tracks. In addition to putting a stop to such problems, there is an additional advantage in embracing pragmatic eliminativism in that the revised claims are also much clearer and more precise than the original ones.

These are two core advantages of pragmatic eliminativism: it has prophylactic power in preventing certain troublesome kinds of debates arising and it introduces more precision

and clarity into the debates. These advantages have general application, beyond the debates over GY. Suppose two thinkers are arguing over whether 'attention' has property y. Let us suppose that we ask them to state their claims in ways that do not contain the word 'attention'. Suppose then one of them says that 'selection for action includes property y' and the other says 'cognitive unison does not include property y'. If both interlocutors agree on the truth of both of these statements, then it seems very likely that the original debate was due to simply understanding the word 'attention' in different ways (because both interlocutors can agree on all of the facts without use of the word 'attention'). In this case, what appeared to be a straightforward (presumably empirical) disagreement about attention and property y is in fact simply thinkers talking past each other. Pragmatic eliminativism brings this starkly into light, and halts the possibility of such cross-talking over 'attention' from arising again.

Importantly, the advantages of pragmatic eliminativism extend beyond its prophylactic power: it does not simply serve to avoid debates which turn on different views on the meaning of the word 'attention'. Pragmatic eliminativism also helps in debates that do *not* turn on such a difference about the meaning of 'attention', as it reveals and clarifies the main point of difference between the two positions in the debate. Suppose for example that on banning use of the word 'attention' one interlocutor says 'accessibility to working memory involves property y' and the other says something like 'accessibility to working memory does not involve property y' then the interlocutors here are disagreeing on the truth value of a certain claim that does not involve the term 'attention', and so this is good evidence that the original dispute was *not* due to understanding the word 'attention' in

different ways. In such a case, we will have clarified the original dispute, pinpointed the differences and can focus our efforts on assessing the claims that remain.

Of course, if we eliminate the word 'attention' and we find that the two interlocutors are still disagreeing, even when they state their claims without mention of 'attention', it does not automatically follow that the dispute *does not* turn on differences in opinion about the meaning of certain terms. As I argued in chapter 2, the term 'working memory' has also been understood in ways different enough to give different answers to certain questions, such as the question of whether the contents of working memory are necessarily phenomenally conscious. Avoiding use of the word 'attention' does not cure all debates of terminological difficulties, but it can at least serve to purge the debates of problems surrounding 'attention'. A reader may be wondering whether 'working memory' might also be a candidate for elimination. A discussion of this would take us too far afield, though it is an interesting question.

Some debates may involve verbal *and* substantive elements. Pragmatic eliminativism is useful here too, helping us distil the aspects of a debate which turn on different meanings of the word 'attention' and the aspects of a debate which may turn upon empirical facts, a priori disagreement or something like that. Take the question of whether attention to x is necessary for perceptual justification of beliefs about x (e.g. Siegel and Silins, forthcoming; Silins and Siegel, forthcoming and Campbell, 2011). Upon banning the use of the word 'attention' in this dispute, we may come to accept that there are various different faculties in the human cognitive system, and that some of them are necessary for perceptual justification, and some of them unnecessary. However, setting aside the use of 'attention' will not resolve *all* of the issues at play here, as there will still be residual questions to ask to do with *which*

faculties are necessary for justification, and *why* they are necessary. In this case, we will have clarified the dispute and filtered out the claims that are due to differing views on the meaning of 'attention' and we will be left with a collection of residual questions.

When we discuss the theoretical advantages of pragmatic eliminativism, it is often tempting to think of them in terms of debates *between* different interlocutors, but of course it will also serve to prevent equivocations that *particular* thinkers make between different meanings of the word 'attention'. It is not uncommon to find individual thinkers sliding between different uses of the word 'attention'. I gave an example of this with Bommarito's work on attention in chapter 3.

Another advantage of pragmatic eliminativism is that it forces interlocutors to state what is interesting about their claims, independently of reference to 'attention'. When a thinker makes use of the term 'attention' in a certain claim, it is easy to assume that what one is saying is interesting or important simply because it is about 'attention'. Suppose we embrace pragmatic eliminativism and replace a claim about 'attention' with a claim about a certain mechanism in the brain, or a certain way of performing a task. This will force thinkers to explain why talk of *this particular* mechanism or way of performing a task is theoretically important for the purposes in question. Pragmatic eliminativism thus removes the temptation to sanction claims simply on the basis of the fact that they are about 'attention'. [11] This will force thinkers to more seriously engage with the substance and importance of their claims.

[11] Vickers (2013, pp.244-249) makes a similar point in the course of his argument for eliminativism about 'theory'.

This concludes my discussion of the prime advantages of pragmatic eliminativism: firstly it makes it impossible for debates to arise as a result of different views on the meaning of the word 'attention'; secondly it helps to clarify disputes which (partially or wholly) *do not* turn on different meanings of 'attention' and finally it makes it impossible to assume that a certain claim is independently interesting simply because it concerns 'attention'.

An important qualification is that I intend pragmatic eliminativism to apply to current debates in the cognitive sciences and philosophy that concern what attention is, what functions attention serves and what relations attention has to other faculties of the mind. I do not intend pragmatic eliminativism to apply to claims which focus on how the *concept* of 'attention' has been used in certain contexts. For example, consider claims such as: 'William James thought 'attention' was (…)' or 'the concept 'attention' is used by Iris Murdoch to accommodate for (…)'. Such claims are doubtless interesting from a meta-philosophical or exegetical perspective. I think that claims like this can safely be excluded from the remit of pragmatic eliminativism about 'attention', because they are not really statements about what attention is or what attention does, they are rather statements about how the *term* 'attention' has been (or is) used in certain contexts. The scope of pragmatic eliminativism is not intended to apply to second-order claims such as these because debates of this kind are not the ones that are causing the problems that we have been looking at (cf. Vickers, 2014, p.120).

Perhaps it will strike a reader as *ad hoc* to allow the word 'attention' to be preserved in these contexts but not in others. To see that this is not so, recall that the argument for pragmatic eliminativism takes its impetus from theoretical pragmatics: among the reasons for pragmatic eliminativism is the claim that we should eliminate 'attention' *because it is*

causing trouble in certain debates and contexts. Whether a certain concept is troublesome is not an intrinsic property of the concept itself, but it is a result of the relational properties that a concept has with a certain debate or context. For this reason, it is perfectly consistent to advocate the elimination of the term from one context (where it is troublesome) and not from another (where it is not troublesome).

As an example, consider the concept 'sport'. It is not hard to think of debates where use of this concept is unproblematic. An example of a context where the concept is unlikely to cause trouble would be a debate over the historical question of what caused rugby-football to split into two separate sports. In such a context, the two interlocutors could agree enough on the meaning of 'sport' for the debate to be fruitful. However, in another context the same concept could cause great difficulty. Suppose, for example, if the interlocutors turned to the question of whether chess counts as a sport. In *this* context we would (rightfully) suspect that the term 'sport' could usefully be eliminated. This would force the interlocutors to examine what work the term 'sport' is doing for them. Whether a certain concept is troublesome is a relational feature that it holds with a certain debate, and this fact demonstrates why it is consistent to advocate eliminativism about the 'attention' concept in some contexts but not others. For this reason, the position I advocate can be called *selective* pragmatic eliminativism.

This discussion of 'selective' pragmatic eliminativism raises a worry. An opponent may at this point object that the problems I have been discussing are highly specific to certain particular debates. An opponent may accept that what I have said may hold true of the debates discussed in chapter 2, but is there any good reason to think that these problems are *widespread* throughout attention research? Unless there is such good reason, then we

need not assume that pragmatic eliminativism is really required, except in just a few specific cases, or so claims my opponent.

There are various reasons that I do not think that there is much hope for insisting that the problems here are not very widespread. Firstly, I should say that even if we *do* restrict the scope of pragmatic eliminativism to the debates over the sufficiency claim and the necessity claim that I have focussed on specifically, this would still be a very substantial position. Secondly, notice that there is nothing *specific* to the debates about the sufficiency and necessity theses that leads to the generation of these problems: similar problems could be generated in a great many debates about 'attention'. This suggests that a great many such debates may be suffering for similar reasons. At the very least, it is worth employing pragmatic eliminativism as a useful diagnostic tool throughout these debates.

Finally, note that the terminological confusions with 'attention' have been noted and bemoaned before, though for different reasons from those put forward by myself. Those that have complained about this include empirical psychologists (e.g. Allport, 1993 and 2011, Uttal, 2011, ch.6 and Bachmann, 2011) and philosophers (Gennaro, 2014). Indeed, even monists have noted these problems (Mole, 2011b and 2012). If we take these claims seriously, this should give us reason to think that the problems are widespread. Taking all of these considerations together, it is not too tendentious to shift the burden of proof at this stage: if someone does claim that the problems here are not widespread or worrying, then the onus is on them to show this. I hold out no hope for such a position.

§5.6.3-Premise (3) and monism.

So much for the main advantages of pragmatic eliminativism. However, the argument in favour of pragmatic eliminativism only goes through if the advantages of the

approach outweigh the disadvantages. This is premise (3). To assess this premise, we will have to ask *what* the disadvantages of doing away with 'attention' could be. When we do away with the term 'attention', what claims will we lack the power to express? What claims can we make in terms of attention, which we cannot make in terms of its subconcepts?

Examples of theoretically important and substantive claims that cannot be stated without mention of 'attention' are difficult to find. For every claim containing the term, it seems to be possible to put forward a claim containing a subconcept of 'attention' (or a collection of claims containing various subconcepts of 'attention') which preserves what is theoretically important about the original claim. Of course, there is no limit to *how many* different claims we may put forward, or how many subconcepts can be employed within these claims, so if a certain thinker feels that something important has been missed by translating the original claim into a claim without mention of 'attention', she can simply form a new claim (possibly with a new subconcept) which captures what she feels was lost.

One dispute which cannot be carried out in the absence of the term 'attention' is the debate over *what attention is* which has been examined at length in the previous chapter. For example, Wu claims that 'attention is selection for action' whilst Mole claims that 'attention is cognitive unison'. If we replace the term 'attention' in these claims with the individual accounts of 'attention' that the particular thinkers embrace, we will be left with the following two claims: 'selection for action is selection for action' and 'cognitive unison is cognitive unison'. Obviously these claims are true. However (my opponent may insist) something important has been lost from the debate, which is the question of *which one of them is really correct about attention!* By avoiding use of the word 'attention', we become unable to state the

main contention of this debate, which is *what attention is.* So (the objection concludes) in the interest of preserving this debate, we should resist pragmatic eliminativism.

This objection has no bite if one is attracted to the brand of pluralism that I recommend. If one accepts my view, one will likely think that whether we select the cognitive unison concept or the selection for action one will largely be a pragmatic decision based on what we want to *do* with the concept, and that neither thinker in this debate is latching onto any deep truth about what attention *is*. For this reason, the pluralist is unlikely to think that this debate is worth preserving, and as such is unlikely to think that preserving the debate is a sufficient reason to resist pragmatic eliminativism.

Obviously, a monist may well completely reject that view, and will likely think that there *is* an important and substantive dispute to be had over what attention *really is*, which we cannot have if we embrace pragmatic eliminativism. Let us set aside my earlier arguments and assume that we do accept monism. Must monists reject pragmatic eliminativism out of hand? I do not think so.

Firstly, notice that *even if* we accept that there does exist one privileged concept of attention which is 'correct', it is worth asking the following question: what would the interest be of engaging in a dispute over which concept this was? What would turn on such a dispute? Suppose we have three concepts, X, Y and Z and that we can use these three concepts to fulfil a great many theoretical roles that 'attention' has been put to: they shed light on consciousness, demonstrative reference, conceptual development, etc. In this situation, what exactly do we *gain* from making a claim such as 'X is attention'? What *progress* would we have made by making such a claim? We would already have shed light on demonstrative reference, consciousness and so on using X, Y and Z; to link *one* of the

concepts in question with 'attention' does not serve to increase our understanding of any of these faculties, nor does it increase our understanding of the faculty or mechanism *that we identify attention with.* Suppose that the concept 'X' refers to some mechanism in the brain, that we can investigate using empirical science. We would not learn anything new of interest about this mechanism by making the claim that the mechanism is identical with attention.[12]

So even if one is a monist, I am unconvinced that one should be interested in having a debate over which concept of 'attention' is the 'correct' one. However, suppose that we *are* convinced that this debate is a worthwhile one, which we are interested in having and which is important. The selectivity of the pragmatic eliminativism argued for in §5.6.2 gives us the resources to accommodate for this. We could *restrict* the remit of pragmatic eliminativism to exclude this debate specifically. We could say that we should still make use of the term 'attention' in order to make claims such as 'X is attention', but we can still embrace pragmatic eliminativism in *other* areas.

This suggests a general claim, which flows from the selectivity of the eliminativism I advocate: if there are contexts where we are sure that a certain debate is worthwhile *and* cannot be adequately carried out *without* mention of attention, we can restrict our pragmatic eliminativism to allow for that. So even if there are debates like this, the pragmatic eliminativism that I recommend has the resources to allow for them.

To restrict pragmatic eliminativism in this way may seem like something of a concession, but there will still be many contexts where we should expect pragmatic eliminativism to apply *even if one is a monist.* As I said in §5.6.2, many of the debates in

[12] Cf. Chalmers (2011, pp.535-536) and Vickers (2014, pp.119-120)

question will contain substantive claims, independently of their use of 'attention'. It may well be a very interesting claim from an ethical point of view that a certain way of assessing one's life is what is responsible for what we normally think of as 'modesty', whether or not this manner of assessing one's life is 'attention'. Here we can recover much of the substantiality of Bommarito's (2013) view without mentioning 'attention'. It may also be interesting that certain processes need to be in place which couple perceptual inputs to particular outputs in a certain way, if a subject is to be capable of action. Here again, much of interest can be saved from Wu's (2008) view, without invoking 'attention'. It is also interesting from both an empirical and philosophical point of view that GY can perform better in Posner tasks that contain valid cues than ones that contain invalid ones. This doubtless tells us something interesting about blindsight and the functionalities of consciousness. Again, these claims can be retained without mention of 'attention'.

I suspect that in practice, most thinkers who are currently engaged in debates that make use of 'attention' will be at best only secondarily concerned with the definitional question of how we should understand the concept 'attention' itself. Many who are engaged in *applying* attention to issues in philosophy of mind, epistemology, empirical science and so on will be willing to set aside definitional issues over 'attention', and recast their claims without it. For thinkers like this, pragmatic eliminativism makes a lot of sense, whether they are monists or not.

§5.6.3.1-Second thoughts: Ereshefsky.

Marc Ereshefsky has until relatively recently been one of the most dedicated defenders of the view that 'species' ought to be eliminated from biology. However, he has

now changed his mind (2010c), arguing instead that the term should be retained because it is 'entrenched'. Ereshefsky says:

> 'Students are taught the term from their earliest encounters in biology. Field guides and taxonomic monographs use the word 'species'. And the term is even found in our governments' laws. From a practical standpoint, it would be hard to eliminate the word 'species'… Removing the term 'species' from biology would be an arduous task, and that speaks in favour of keeping it' (2010c, pp.420-421).

One could of course say that this applies to 'attention' as well: we should not eliminate 'attention' because it would just not be worth the work. In response to this view, I accept that 'attention' is 'entrenched' in a similar way to 'species': the term is widely used throughout cognitive science, in undergraduate textbooks and so on. Indeed, like the word 'species' the word 'attention' is found in the laws of both the UK and the USA and doubtless many other countries. However, I think that Ereshefsky is overstating the difficulties here. Firstly, I think it is quite recalcitrant to insist that we should not make changes that will be pragmatically useful just because it would involve effort to do so. Secondly, it is not news to be told that the conceptual repertoire of the empirical sciences changes in many important ways over time. We would expect conceptual change in biology and psychology over the coming years anyway; there is no reason to think that 'species' or 'attention' should be immune from such revision.

The third point against Ereshefsky is deeper, and grows out of my discussion of selective pragmatic eliminativism. Ereshefsky mentions the use of the word 'species' in undergraduate textbooks and in our governments' laws. If we were indeed convinced that the use of the terms in textbooks and law-making made the task of removing them too

arduous, then the selectivity of the pragmatic eliminativism that I advocate here becomes relevant again. We can include the use of the term in lawmaking, and undergraduate textbooks as places where pragmatic eliminativism need *not* apply. If what I have said above is correct then this makes perfect sense: if we are worried about the use of a certain term in one context, we should not be too worried if we preserve it for use in another, more innocuous context.

I think that one of Ereshefsky's mistakes here is to assume that 'eliminativism' must be an all or nothing affair: we either do without the term 'species' *tout court* or we retain it for all occasions. However, if what I have been arguing is along the right lines then this is an unnecessary overgeneralisation, and selective pragmatic eliminativism is the more sensitive position. So, I think that these worries can be met: Ereshefsky changed his mind from eliminativism to pluralism for bad reasons.

§5.6.4-Pragmatic eliminativism and realism.

My opponent could here raise the following objection: surely (says my opponent) I have tried to marry heaven and hell. I want the ontology of the realist, but the theoretical advantages of the eliminativist. Are these really consistent? This reaction is reasonable, because there are a great many versions of realism which *are* inconsistent with a great many versions of eliminativism. My response to this worry is that when we make the adequate distinctions between different *versions* of realism and eliminativism, then we open up a space for holding a particular version of one doctrine consistently with a particular version of the other one. It is only by making the appropriate distinctions that the consistency emerges.

To see this, it will be helpful to state the core commitments of each doctrine:

(Realism): Claims containing the predicate 'attention' can be (and often are) true, because they can (and often do) have truthmakers.

(Pragmatic Eliminativism): We have good reason (based upon theoretical pragmatics) to avoid use of the term 'attention' in (at least) a large number of debates that currently make use of the term 'attention'.

These are the two claims that are at the heart of this chapter, and they are consistent with each other. At the very least, *if* they are inconsistent, then this inconsistency is not obvious and we would require a detailed argument to show *why* they are inconsistent. I stressed above that the substantive debates in this vicinity are over the plausibility of each of these positions; the issue of whether they deserve the titles 'realism' and 'eliminativism' is at best a secondary consideration.

We can hammer home the point about their consistency by turning to a particular version of the 'inconsistency' worry under scrutiny. It may be said that in embracing pragmatic eliminativism, we lose the ability to make claims about 'attention' because we have banned talk of 'attention'. For this reason, we cannot have claims about attention that have truthmakers, and so we cannot have the realism that I have promised. This objection misunderstands the nature of the pragmatic eliminativism that I have been arguing for. I do not claim that it is *impossible* to make claims about 'attention' (many of which can be *true*). My claim is only that in certain debates (probably a high number of debates) it is methodologically prudent for us to refrain from doing so.

Before I move on, it will further be helpful to introduce a statement of the pluralism that I have argued extensively for, in order to see that it is consistent with the two claims given above:

> (Pluralism): The term 'attention' is ambiguous between several importantly different concepts, no one of which is privileged. Several of these concepts are worthy of preservation for theoretical reasons.

The overall position that emerges from these three views is this: we have the term 'attention' which is ambiguous between several different concepts, where several of the concepts are theoretically useful, and should be preserved. However, when we deploy these concepts in our theorising, we should avoid use of the overarching predicate 'attention' for theoretical reasons; even though we could *if we wished* make true claims using the predicate 'attention'.

§5.7-Another reason to be an eliminativist about 'attention'?

In this section, I will briefly discuss the argument of psychologist Britt Anderson, who seems to argue for *ontological* eliminativism about attention (2011). Anderson's argument is toward the claim that 'there is no such thing as attention' (p.2). The argument is centred around the idea that attention research has been focussed on dichotomies that have been taken as axiomatic, rather than experimentally shown, and the claim that attention has been 'reified'.

It is important to mention the presence of an argument for ontological eliminativism about attention in the psychological literature but I will not dwell on Anderson's argument for too long, as it does not really seem to reach the conclusion of full ontological

eliminativism. For example, at one point he says that '[w]e need to recognise attention is an effect and not a cause' (p.3) and that 'I assert attention should be treated as an effect' (p.4). Anderson still seems to think that attention exists, if only as an 'effect' so either he is not really an ontological eliminativist, or his view is internally inconsistent. We can leave Anderson's view here.

§5.8-A last objection.

It has been suggested to me that if my general picture is correct, then I may end up proving *too much.* The complaint is that if I accept this kind of picture about 'attention' then I may be forced to accept it about a great many different concepts. It may transpire that we become pragmatic eliminativists about almost everything![13]

This objection is difficult to assess because how much of a problem one thinks that it is will depend upon how keen one is to avoid this kind of analysis being applied to other philosophically prominent concepts. Chalmers, for example, seems happy to allow his own version of the heterogeneity strategy to apply throughout philosophy (2011, esp. pp.531-532). For someone like Chalmers, this objection has no bite.

My own view on this matter is that we should avoid sweeping statements about what concepts will and will not be susceptible to the arguments that I have put forward. To what extent different concepts have been interpreted differently by different philosophers will be a matter of degree, and it is a matter of degree how much damage this kind of issue does in a debate. Whether what I have been saying about 'attention' can be applied to a certain concept is something that would require us to establish how that concept has been

[13] This objection was put to me by Tony Cheng.

employed in the debate in question. This would require an in depth study of the debate. Clearly, we are in no position to do this from the armchair.

If it is true that something like my account can apply to a significant number of other concepts then I will have no problem with that. I would expect that many of the concepts that I have compared 'attention' to (such as 'species') will be susceptible to roughly the kind of account I have given, but I do not know how far the general methodological approach can be appropriately applied throughout philosophy. I certainly do not think it is obvious that a great many philosophically interesting concepts will succumb to such an analysis (I don't even know if this is *plausible*), even if my account is on the right lines for 'attention'.

§5.9-Conclusion.

I take pragmatic eliminativism to be a sensible development of pluralism about 'attention'. Nonetheless, these two views are independent: neither one implies the other. For this reason, I intend the argument in favour of pragmatic eliminativism to function as a 'backup' argument: if readers are unconvinced by my arguments for pluralism, then perhaps the argument for pragmatic eliminativism will be more attractive. Equally, if you are convinced by my argument for pluralism but unconvinced by pragmatic eliminativism then that is good enough for me: I will count you as an ally. However, I think we have good reason to embrace both views.

Throughout this thesis, I have constantly deferred discussion of natural kinds until later. Some thinkers will certainly object to this, as many consider discussion of natural kinds to be crucial to any pluralist or eliminativist position. We will finally turn to this issue in the next chapter.

Chapter 6
Psychological Eliminativism and Natural Kinds.

§6.0-Summary.

This chapter will consider the nature of eliminativism in psychology, and how these issues relate to natural kinds. It will be shown that issues over eliminativism in philosophy of psychology have usually been linked very closely to natural kinds. It will be argued that this whole approach should be rejected. Two eliminativist arguments in the literature will be targeted specifically in making this case. It will be argued that the eliminativist strategy pursued in the previous chapter of this thesis is more plausible than any approach that invokes natural kinds. In addition, several objections to the views defended in this thesis will be considered. These objections are also based around claims about natural kinds, and it will be argued that these objections fail for similar reasons. A general scepticism about the usefulness of the idea of 'natural kinds' for philosophy of psychology will emerge.

§6.1-The natural kinds question and three eliminativist arguments.

In the previous chapter, I gave an argument for pragmatic eliminativism about 'attention'. I suggested that the term 'attention' be dropped from a certain collection of debates within philosophy and psychology, and my reasons for thinking this were that the theoretical advantages of dropping the term heavily outweigh the disadvantages. It may be thought that I have ignored an important issue in the vicinity, which is the question of whether 'attention' refers to a natural kind. Indeed, we find natural kinds being linked to issues over eliminativism very clearly in the 'attention' literature. Here is Declan Smithies again:

> 'if the various different mechanisms of selection [that have been called 'attention'] have nothing in common besides the function of selecting information for some purpose or other, then it is not clear that attention is a natural kind…In which case, the concept of attention should simply be eliminated from a mature cognitive science and replaced by a more fine-grained taxonomy' (2011a, pp.251-252).

Here Smithies expresses a commonly held view: if a term does not refer to a natural kind then it should be eliminated. The corresponding view is of course that a term should be *preserved* if it *does* refer to a natural kind. This tight link between natural kinds and eliminativism is commonplace throughout philosophy of psychology. Particularly prominent examples of this kind of thought come from Paul Griffiths, Edouard Machery and Elizabeth Irvine. All three of these thinkers argue for eliminativism about a certain term ('emotion', 'concept' and 'consciousness' respectively), and all three do so by arguing that the term in question does not refer to a natural kind. Here is Griffiths:

> 'It is unlikely that all the psychological states and processes that fall under the vernacular category of emotion are sufficiently similar to one another to allow a unified scientific psychology of the emotions… In a slogan, emotions are not a natural kind… I have described my position as a form of eliminativism about emotion, because it implies that the term 'emotion' and some specific emotion terms like 'anger' are examples of 'partial reference'… In the same sense that there is really no such thing as jade, only jadeite and nephrite, there is no such thing as emotion' (2004a, pp.901-902).

Similarly, Machery argues that 'concept' designates three different mechanisms, which do not form a natural kind. Machery takes this to support eliminativism about 'concept'. He summarises the link between natural kinds and eliminativism succinctly:

> 'when it is found that a scientific term fails to pick out a natural kind, there is a presumption that it should be eliminated' (2010a, p.206).

Machery puts more flesh on this claim elsewhere:

> 'the scientific eliminativist makes a case that the class of concepts does not possess the properties that characterize the classes that matter for the empirical sciences. Or, to use a slogan, that this class is not a natural kind. If "concept" does not pick out a natural kind, then it is unlikely to be a useful notion in scientific psychology. It is even likely to stand in the way of progress in psychology. If this is the case, the term "concept" ought to be eliminated from the theoretical vocabulary of psychology' (2009, p.230).

Similarly, one of the main claims that Irvine makes in support of eliminativism about 'consciousness' is that the term does not designate a natural kind, but many different subkinds (2013, esp. chs. 7 and 10). As we shall see, many have responded to these arguments by claiming that in fact the term in question *does* designate a natural kind, and this is usually taken as reason to reject eliminativism about it. Typically, the debate then revolves around the question of whether the term refers to a natural kind or not.

There are obvious similarities between the arguments of Griffiths, Machery, Irvine and my own. They all involve the claim that a particular term that is now widely used in philosophy, psychology and neuroscience should be eliminated. These similarities raise several questions with relation to *my* arguments. Firstly, is attention a natural kind? Secondly, if it is *not* one, then is this a good reason to be an eliminativist about it? Thirdly, if it *is* one, then does this show that eliminativism about 'attention' is false? In the course of this chapter I shall argue that the first question is a poor one to ask, and that the other questions should be rejected as a result. I will argue that the whole 'natural kinds' approach to questions of eliminativism in psychology should be rejected. I will be examining many of the issues by focussing on the arguments from Machery and Griffiths. Both of these thinkers

offer rich and varied arguments for their core claims, and it is not my intention to discuss their views in exhaustive detail. I will restrict myself to examining the general structure of their arguments.

§6.2-What is a natural kind?

We must begin with the question: what is a natural kind? I certainly cannot do full justice to this question here but there are some (relatively) uncontroversial things that we can say about what natural kinds are supposed to be (taken from Quine, 1969). Firstly, they are groups of entities that are *of interest to science* and secondly, they permit scientifically interesting generalisations over a range of cases. One intuition driving this second claim is that it seems that when we make certain discoveries about *some* members of a certain class of entities, we can be reasonably confident that these discoveries will tend to hold true of all members of that class and that they will continue to do so in the future.

A more controversial claim is that natural kinds are *objective divisions in nature*, which exist independently of human thought. This claim is what many thinkers have in mind when they say that natural kinds *carve nature at the joints.* We will touch on the idea of 'joint carving' in some of what follows. Another controversial claim is that natural kinds can be found in many different sciences. Electrons, carbon and dogs are each thought to be natural kinds that belong to physics, chemistry and biology respectively. Natural kinds must be distinguished from natural kind *terms*. The former of these is a set of entities in nature (these form the kind itself) and the latter is the term that refers to these entities. Electrons form a natural kind, whilst the term 'electron' is the natural kind term that refers to electrons.

§6.2.1-Two red herrings.

Firstly, we should set aside two accounts which will *not* be relevant to the present discussion: the essentialist account of natural kinds and the conventionalist one. The simplest way to summarise the essentialist view is that for something to be a member of a natural kind, it must instantiate a set of intrinsic properties that are necessary and sufficient for it to be a member of that kind. Kripke (1980) and Putnam (1975) use anti-descriptivism about reference to motivate an essentialist view of at least some natural kinds. Kripke gives an essentialist treatment of 'gold' as substances with atomic number 79, and Putnam famously puts most weight on the controversial claim that water is H_2O. One particularly well worked out essentialist theory of natural kinds comes from Brian Ellis (2001 and 2005).[1]

The essentialist view is interesting in its own right for many reasons, but I would like to set it aside here because it is not best suited to the question of whether a certain term in psychology refers to a natural kind. This is for two reasons. Firstly, an essentialist treatment of natural kinds has usually been seen as inappropriate for 'higher order' kinds such as the kinds of biology and psychology. Though the essentialist account of natural kinds may work well for the kinds of physics, its appropriateness for the kinds of biology or the higher level entities studied by psychology is doubtful. It is generally held that essentialism (at least of the 'traditional' kind that I have outlined here) is a mistaken view of biological kinds, and many philosophers consider it to be a product of pre-Darwinian thinking (e.g. Hull, 1965 and Sober, 1980).

This is evident in the work of Ellis, who is himself an essentialist. Most of Ellis' examples of natural kinds that can be analysed along essentialist lines are taken from

[1] Ellis does not use anti-descriptivism about reference to motivate his view.

fundamental physics and chemistry, and it is highly controversial whether the essentialist view is even an appropriate treatment for the kinds of chemistry (Hendry, forthcoming). When discussing biological kinds, Ellis says:

> 'Some classes of things that are plausibly natural kinds are really clusters of similar natural kinds, which are conceptualised as being things of the same kind, although there is no set of standing intrinsic properties or structures that would distinguish the members of these kinds from those of all other kinds. These are what we might call 'cluster kinds'. The members of the various animal and plant species, for example, are members of cluster kinds' (2001, p.32).

There are some philosophers (e.g. Devitt, 2008) that have defended essentialism about biological kinds. However, his essentialism departs in certain important ways from the kind of natural kind essentialism that thinkers such as Ellis have in mind. Most notably, Devitt emphasises that the properties that form a species' essence are often only *partly* intrinsic to the organism (2008, p.346). This is because Devitt notes the importance of ancestry in setting the boundaries of biological kinds. In any case, Devitt's view is particularly extreme, and I shall discuss it no more here.[2]

The second reason that I would like to set aside essentialism bears more directly on attention. In chapter 4, I have already examined the view that attention has an essence, and found it to be lacking. Much of what I said in that chapter can apply *mutatis mutandis* to the claim that attention is an essentialist natural kind. I leave the essentialist account here.

[2] For criticism, see Ereshefsky (2010a).

The second account of natural kinds we will need to set aside is the conventionalist one. Conventionalism comes in a variety of flavours (e.g. Goodman, 1978 and Heil, 2005) but it can roughly be summarised as the view that natural kinds are simply the result of a choice of how to classify the world. Science chooses to classify the world in a certain way, but this does not latch on to any privileged 'joints' in nature: how science classifies the world is a result of convention. On this account, claiming that a certain term refers to a natural kind will not be a very substantive claim. I set this account of natural kinds aside as well.

§6.2.2-A more promising approach?

Interestingly, when we consult the literature on eliminativism in psychology, we find that there is broad agreement over which account of 'natural kinds' is the most pertinent. This is the homeostatic property cluster (HPC) account which is mainly due to the work of Richard Boyd (1989, 1991, 1999, 2003 and 2013). Boyd puts forth his HPC view in slightly different ways in different places (the most clear summary of Boyd's view is his 1989, pp.16-17). There are, however, certain core features of the account that can be extracted from Boyd's writings.[3]

The first is that members of a natural kind instantiate a *property cluster* (e.g. Boyd, 1989, p.16 and 1991, p.129). That is, there must be a cluster of properties that reliably repeats itself in nature in 'an important number of cases' (1989, p.16) in order for the set of entities that instantiate that cluster of properties to qualify as a natural kind. Importantly, it is *not* the case that *all* of the entities within the kind must instantiate *all* of the properties within the

[3] A note on terminology: the HPC view is sometimes called 'cluster kind realism'. Confusingly, some refer to Boyd's account as 'essentialism' or 'weakened essentialism' (Ereshefsky, 2010a) though the account is significantly different from the essentialist account I have already put aside, so I shall not refer to it as essentialism.

property cluster: instantiation of the property cluster can admit of exceptions (1989, p.16). We find exceptions of this kind in nature: it may be the case that dogs tend to have four legs. 'Four legged-ness' may be one of the properties in the property cluster that is definitive of the kind of dogs. However, a three legged dog is still a dog: failure to fulfil all of the properties in the cluster *does not* automatically exclude something as counting as an instance of the kind in question. This is what marks Boyd's view out from traditional essentialism.

Such repeated co-occurrence of a cluster of properties is not *sufficient* for natural kindhood. The continued re-occurrence of the properties must be underwritten by a *homeostatic mechanism* (1989, p.16). A homeostatic mechanism is some common (typically causal) mechanism that tends to generate the properties in the property cluster. This causal mechanism is a feature of the world which produces and *explains* why the properties tend to cluster together in nature. On the HPC account, this mechanism also explains why it is possible to infer from the fact that some members of the kind instantiate certain properties to the claim that any member of the kind will tend to instantiate those properties: it is because these properties are caused by *the same mechanism* (Boyd, 1991, pp.130-131 and 1999, p.68).

It is open-ended what this mechanism could be. Typical examples of such mechanisms include genetic inheritance, pressures in the natural environment and common descent, but there is no reason why things such as the social class into which one is born, the school one went to or parental influence on one's cognitive development could not count as appropriate mechanisms. I will be focussing on mechanisms in §6.5.2.

Given a certain mechanism, it will of course be an empirical question exactly *which* properties the mechanism causes. This is why Boyd insists that the properties in the cluster must be discovered a posteriori rather than a priori (1989, p.16). Indeed, the insistence on the

claim that members of a kind must be underwritten by a common causal mechanism is largely motivated by the fact that the extension of a natural kind term should be determined by the causal structure of the world, rather than convention or a priori stipulation. As Boyd says:

> 'the unity of the property-cluster which defines [the kind] is *causal rather than conceptual*… a natural kind is associated *causally* with a large family of methodologically important properties' (1991, p.141. Italics in original).

This quotation raises two further points that will be important for my argument. Firstly, notice that Boyd insists that there must be a *large* number of properties in the cluster. As we shall see, this point is emphasised by many of Boyd's followers. Secondly, notice that Boyd insists that the properties in the cluster must be 'methodologically important'. By this, Boyd means that the properties definitive of a kind are in part dependent on which properties the scientific discipline in question is interested in. This leads Boyd to claim that natural kinds are *not* objective divisions in nature that exist entirely independently of human thought, but exist partially due to the disciplinary context in which they are discussed. This is Boyd's 'accommodation thesis'. He is explicit on this issue:

> 'The naturalness of a natural kind depends on the disciplinary context within which it is employed… what makes natural kinds *natural* - what makes reference to them contribute to projectibility judgments– is that reference to them allows us to achieve

> in [sic.] *accommodation* between our classificatory and inductive and explanatory practices… and the causal structure of relevant phenomena' (1999, p.69).[4]

And:

> 'Natural kinds are features, not of the world outside our practice, but of the ways in which that practice engages with the rest of the world' (2003, p.538 See also Boyd, 2013, p.53).

Boyd's HPC view is probably the most popular view of natural kinds in philosophy of science today, at least as an account of the kinds of the life sciences such as biology and psychology. Indeed, Samuels and Ferreira describe it as a 'consensus' view (2010, p.222). The popularity of the HPC view extends beyond philosophy, and has even filtered into biology itself (Rieppel, 2005).

The influence of the HPC view to the debates over eliminativism that I mentioned above is very clear. Griffiths spends much time developing and expanding on Boyd's view (1997, chs. 7-9). Like Boyd, Griffiths emphasises the following criteria:

> 'A category is (minimally) natural if it is possible to make better than chance predictions about the properties of its instances… Ideally, a natural kind should allow *very reliable* predictions in a *large domain* of properties' (2004b, §1.2)

At other times, Griffiths places a great deal of weight on the idea of a homeostatic mechanism (1997, p.242), a notion he lifts directly from Boyd. Machery also borrows Boyd's account of kinds, insisting that members of a kind must share:

[4] In the first part of this quotation, Boyd is actually summarising Ruth Millikan's view rather than describing his own, though he explicitly says that he agrees with it.

'a large set of scientifically relevant properties… because of some common causal mechanism' (2005, pp.447-448 and 2009, p.232).

Machery also emphasises that:

'[a] natural kind is a class about which many generalizations can be formulated: its members tend to have many properties in common' (2005, p.448).

Notice that Griffiths and Machery place strict constraints on what can count as a natural kind. The former remarks that natural kinds should support 'very reliable' predictions in a 'large domain' of properties; the latter demands that members of a kind must support 'many' generalizations, and that members of a kind must share a 'large' number of scientifically relevant properties. This will be important for my argument later. Like Griffiths and Machery, Irvine defers to the HPC view in her eliminativist argument (2013, pp.93-95). From now on I will follow the trend of the thinkers in this field, and assume that the HPC view is the most relevant for the question of eliminativism in psychology and biology.

§6.3-A space of kinds.

As we have seen, debates over whether a certain term should be eliminated from psychology often revolve around the question of whether the term refers to a natural kind, where the relevant notion of 'natural kind' is the HPC view. In this section and the next I shall reject this approach, by arguing that whether a certain term refers to a natural kind (in the HPC sense) is often a matter of personal interpretation rather than empirical fact. In the present section I shall give some general arguments for these claims and in the next section I will present some concrete examples of this occurring in the debates in question.

We can begin with the issue of vagueness in the HPC view. At several times, Boyd explicitly says that the extension of some natural kind terms is *vague.* Here is what Boyd says on this matter (here 'F' is the set of properties that members of the kind tend to share, and 't' is the natural kind term that refers to the kind):

> 'there will be many cases of extensional "vagueness" which are such that they are not resolvable even given all the relevant facts and all the true theories. There will be things which display some but not all of the properties in F (and/or in which some but not all of the relevant homeostatic mechanisms operate) such that no rational considerations dictate whether or not they are to be classed under *t*' (1989, p.17).

Here we see two possible ways that a natural kind can be vague: when a certain entity possesses some but not all of the properties definitive of the kind, and when the homeostatic mechanism is only partially operative. Indeed, Boyd argues (esp. 1991, p.141-142) that the account *has* to be vague in order to be an accurate account of biological kinds, because dominant evolutionary theory suggests that there must be organisms that are indeterminate in classification between a parent species and an emergent one.

As I noted above, Boyd claims that members of a kind must tend to share a 'large' number of properties in order to qualify as a natural kind, and advocates such as Machery and Griffiths follow him in this. Obviously, the word 'large' is vague.[5] The issue is this: how we interpret Boyd's account will depend upon how strictly we interpret this criterion. Consider someone who interprets 'large' as meaning *a very great many* properties. Someone who interpreted 'large' in this way would be more likely to *exclude* a certain collection of properties shared by a group of entities from being sufficient for making those entities a

[5] Machery (2005) notes this problem, but does not really address it.

kind. Conversely, someone who interpreted 'large' in a much more liberal way would be much more permissive, and would be willing to count far *fewer* common properties as sufficient for a group of entities to count as a kind. These two interpretations of Boyd would deliver very different constraints on what it takes for something to count as a HPC kind. Two thinkers could then examine the same collection of entities, and one would claim that they *do not* form a HPC kind, whilst the other would insist that they *do.*

A related problem of interpretation with Boyd's account is this: in examining a collection of entities, and attempting to decide whether they form a natural kind, we must obviously ask whether the entities are 'similar' to each other in certain ways, and *how.* This immediately raises a question of *fineness of grain*. One thinker could use a coarse grained analysis and claim that some entities count as a natural kind because they tend to share some very abstract properties, specified at an extremely vague and coarse grained level of detail. Someone else could insist that the class of entities must share properties at a *much more fine grained* level of analysis in order to count as a natural kind. We will be encountering problems like this later.

There are many crucial areas at which the HPC view is vague. Recall that the properties in the cluster that is definitive of a HPC kind should be of interest to the science that uses the terms that refer to the kinds. This opens up the possibility of different thinkers having different views on exactly what counts as a 'scientifically relevant' property. Someone with an austere view on what counts as a 'scientifically relevant' property would be more likely to exclude something from counting as a kind, whilst someone more liberal would be willing to count the very same entities as forming a HPC kind.

As I pointed out above, one criterion on a collection of entities counting as a HPC kind is that there must be a large number of scientific generalizations that can be made about the entities in question. Once again, this opens up room for different interpretations of the HPC view, as two thinkers could have different opinions about *how many* generalizations would have to hold true, or how successful such inductive practices would have to be.

We can imagine a space of different interpretations of the HPC account. Different interpretations of the terms that I have just mentioned will deliver a different point in the space of possible interpretations of Boyd's view, and possible interpretations of Boyd within the space can vary enormously in strength. These differences in strength will deliver different versions of the HPC view, and application of these different versions will deliver different verdicts on the question of whether a certain collection of entities is a natural kind. It is difficult to see how such disagreements could ever be resolved, because Boyd's account is itself too vague to settle the matter one way or the other. Importantly, such interpretive disagreements would not be disagreements about which account of natural kinds is correct *per se*. Such disagreements would arise between two thinkers both of whom agree that the HPC account is the correct one. The result of this is that in many debates over whether a certain term refers to a natural kind, all that the different sides are doing is interpreting Boyd's account in ways different enough to give different answers to the central questions.

§6.4-Two case studies.

It will help if we see how these issues play out in practice by examining two debates where these problems emerge.

§6.4.1-Concepts.

Machery's argument for the claim that concepts are not a natural kind revolves around his claim that there are few scientifically relevant properties that concepts share, and there are few scientifically relevant generalisations that can be made about concepts (esp. 2005, p.450 and 2009, p.239).[6] Dan Weiskopf disagrees: he thinks that 'concept' refers to a natural kind (2009 and 2010).[7]

Weiskopf attacks Machery by claiming that there are several important properties that are shared by the set of entities referred to by the term 'concept', and several interesting generalisations that can be made about them. One of these is the phenomenon of 'conceptual combination' (2009, 164-165). Weiskopf claims that all concepts combine to produce new meaning (as when we combine WOODEN and SPOON to form the concept WOODEN SPOON).[8] This fact is one part of Weiskopf's argument that 'concept' designates a natural kind.

This is Machery's response:

> '[the claim that all concepts are involved in conceptual combination] is correct. It is, however, unclear why this is taken to justify conserving the notion of concept.

[6] For some of the controversy surrounding Machery's arguments, see Piccinini and Scott (2006), Margolis and Laurence (2010) and Strohminger and Moore (2010).

[7] Weiskopf believes that there are sub-kinds of concepts. When discussing kinds, Weiskopf says: 'Kinds are understood here as groupings of entities that participate in a body of empirically discovered reliable generalizations, and which participate in those generalizations due to some set of properties they have in common' (2009, p.147).

[8] In this chapter, I follow convention by using capitals to refer to concepts.

> Evidence suggests that in conceptual combination, prototypes, exemplars, and theories fulfil different functions… Thus, prototypes, exemplars and theories are likely to be used by different subprocesses of the process underwriting concept combination' (2009, 245).

Weiskopf is pointing out that all concepts are involved in conceptual combination, and he is happy to accept this as sufficient to count as a scientifically relevant property which can help to count concepts as forming a natural kind. Machery responds by insisting that at a finer level of analysis, the processes involved in 'conceptual combination' are quite different (they fulfil different functions, and work in different ways).

Also in support of his claim that concepts form a natural kind, Weiskopf (2009, pp.166-167) points out that concepts are all stored in long-term memory. Machery's response to this claim is as follows:

> '[concepts] are stored in long-term memory all right, but the rules that govern storage, permanence, and retrieval are likely to be different. Thus, there is so far no serious evidence that the notion of concept underwrites non-trivial scientific generalisations' (2009, 245).

Whereas to Weiskopf, being stored in long term memory is sufficient to count as a scientifically relevant property, Machery does not think this is enough. Machery demands there must be *more* properties that the entities have in common, and they must all have similarity at a finer grain of analysis: they must all be stored in long term memory *in the same way*, using similar rules for storage and retrieval.

Weiskopf and Machery are placing different constraints upon *how much* various entities must have in common in order to count as a natural kind. Weiskopf is liberal: he only requires concepts to share a few properties, specified at a coarse grain of analysis in order to count concepts as a natural kind. Machery is stricter: for him, concepts must share *more* properties, and they must exist at a much finer grain of analysis. This is why Machery claims that concepts are not a natural kind, and Weiskopf claims they are.

We can find similar issues elsewhere in this debate. Samuels and Ferreira (2010) think that concepts are a natural kind. They point out that they are all used to reason inductively, are involved in linguistic comprehension, are used to store knowledge, are non-proprietary, are used by default in higher cognition, are subject to temporal and inter-subjective variation, are internally connected and are internally coherent. The conclusion that Samuels and Ferreira reach is that concepts *are* a natural kind.

Machery's response to Samuels and Ferreira is to insist that these properties only support 'few' as opposed to 'many' scientific inductions and generalizations, and that this disqualifies them from counting as a natural kind (2010b, 237-238).[9] Again, it is easy to see what is at issue here: a strict interpretation of the claim that natural kinds must support 'many' interesting generalisations will deliver a verdict in line with Machery's view; Samuels and Ferreira seem to have something more liberal in mind, which vindicates the converse verdict.

[9] Machery admits that they may count as a homeostatic cluster kind (2010b, 238). This is because Machery reserves the term 'natural kind' for homeostatic cluster kinds that support many interesting generalisations, so on this terminology it is possible for a set of entities to be a homeostatic cluster kind and *not* a natural kind. This terminology departs from Boyd's choice of terms, but it will not matter for what I have to say in this chapter.

At one point, Machery responds to a similar objection by claiming that the properties in question are not *of interest to psychologists* and therefore, the entities that instantiate these properties do not form a natural kind (2009, p.244). Here again, it is clear that this debate is an interpretive one: Machery has a strict interpretation of 'scientifically relevant' which delivers the verdict that the properties in the cluster are not 'scientifically relevant' enough to count qualify them as forming a natural kind.

Interestingly, both Machery and Samuels and Ferreira mention the fact that whether a certain term designates a natural kind may well be open to interpretation (Machery 2010b, 238 and Samuels and Ferreira 2010, 223) but both of them still insist that their interpretation is the correct one. If what I have been urging is along the right lines, we should not expect there to be any one 'correct' answer to these questions, as the criteria that Boyd lays down for us to judge them on are open to interpretations that are different enough to deliver different answers to the question of whether concepts are a HPC kind.

§6.4.2-Emotions.

It is fair to say that there is *even more* heated debate over whether emotions are a natural kind than over whether concepts are (e.g. de Sousa (1987), Nussbaum (2001), Prinz (2004)). As already mentioned, at the centre of this debate is Paul Griffiths, whose eliminativism about 'emotion' is based on the claim that the vernacular term 'emotion' refers to two or three separate natural kinds. I say 'two or three' because he argues that 'basic' emotions are 'affect programs', and that these form a kind distinct from the more higher-level cognitive emotions (1997, chs. 1-6). This gives Griffiths the conclusion that the term 'emotion' refers to at least two different natural kinds. However, at times Griffiths seems unsure of whether to go further and argue that the term 'emotion' refers to *three* natural

kinds, not merely these two. This is because sometimes he suggests that there might be a third category of emotions (including love), which are what he calls 'internalized model[s] of appropriate behaviour'. He gives different reasons at different times for doubt about whether this third category is strictly speaking a third kind of emotion. At some points (e.g. 2004b) he seems doubtful of whether these 'internalized models of appropriate behaviour' are a good account of mental processes such as love. At other times (e.g. 1997, p.246) he seems to think that even if the account were correct, such internalized models of appropriate behaviour would not strictly speaking count as emotions, but only pretend emotions (because real emotions must be sincere). Nonetheless, whether Griffiths thinks that 'emotion' refers to two or three natural kinds, the main point is the same: the term 'emotion' refers to more than one natural kind, and thus should be eliminated from psychology and neuroscience.

Many of the problems found with the 'concept' debate show up in an examination of Griffiths' arguments as well. For example, Griffiths claims that the folk psychological term 'emotion' does not refer to a natural kind, but he also argues that all of the entities that the folk term refers to are 'irruptive motivational states' (e.g. 1997, pp.245-247). By this, he means that all of the processes that the folk term 'emotion' refers to are 'states which interfere with the smooth unfolding of plans designed to secure our long-term goals' (1997, p.246). This is the unifying element that all instances of emotion have in common, or so claims Griffiths. Let us assume this is true. This immediately raises a question: why not say that the fact that all emotions interfere with the smooth unfolding of plans designed to secure long-term goals is sufficient to qualify them as a natural kind?

Griffiths' reply is that, though it is true of all emotions that they have this common unity, they fail to be a natural kind because the psychological mechanisms that implement them are different (1997, p.246). However, at this point, an opponent could resist the claim that the natural kinds of psychology need to share a common unifying psychological mechanism *specified at this level of detail.* An opponent could claim simply that (at an abstract level of functional specification) all emotions have a similar purpose (as specified by Griffiths himself, as the interference of secure long-term goals) and equally that the mechanisms responsible for this are *similar enough* (again, at a high level of functional abstraction) to warrant counting them as a natural kind. Because Boyd's and Griffiths' accounts of natural kinds are open to interpretation, this move cannot be blocked by insisting that such collections of entities would not 'really' be natural kinds.

Above, we found that if we set our analysis at a high enough level of abstraction, then concepts will count as a natural kind. The same is true of the emotions: at a sufficiently coarse grained level of analysis, they can count as a natural kind. Again, the question of whether the category is a natural kind comes down more to interpretation than to the structure of the world.

A further example of these problems arises from the debate between Griffiths and Louis Charland, who argues that emotions form a natural kind (1995 and 2002). Like Griffiths, Charland explicitly commits to the HPC view of psychological kinds (2002, p.512). Charland's argument is based around two claims. Firstly, Charland argues that all emotions 'involve their own distinct mode of representation' (p.522). The kind of representation that Charland has in mind is that 'emotions are normative or evaluative judgements' (p.522). The second part of Charland's argument is that 'there are reliable generalizations and principles

of inference that govern emotional behaviour (p.522), though Charland admits that there are only a small number of such laws (p.524). All emotions are 'felt affective' representational states, and it is these unifying features are enough to qualify emotions as natural kinds, or so claims Charland. Griffiths replies to Charland by pointing out that there are a great many different affective states and that 'the category of felt affective states is so broad, it is natural within this framework to seek distinctive *kinds* of processes involving affective feelings' (2004b, §2.2).

Charland is happy to accept the category of 'felt affective states' as being sufficient to count members of the category as natural kinds. Griffiths' view is more divisive: he claims that the differences between different *kinds* of felt affective states entails that they are not a natural kind. This is precisely the problem we saw with the 'concept' debate: one side uses the commonalities shared by some entities, specified at an abstract and vague level of detail, and concludes that the entities are a kind; the other side demands *more* commonalities at a *finer* level of detail and concludes that they are *not* a kind.

I have argued that it is a mistake to think that there must be one answer to the question of whether a certain term *really* refers to a HPC kind. On an Ellisian account of natural kinds, it is plausible that there would be one objective answer to the question of whether a certain collection of entities form a natural kind, because to Ellis what is and is not a natural kind is an entirely mind-independent fact (2001, pp.19-23). Boyd's account is not like that: Boyd himself purposely allows vagueness into the account, and (if what I have said is correct) this vagueness is so great as to render many debates over whether a certain set of entities is 'really' a natural kind a matter of interpretation rather than empirical discovery.

§6.5-Natural kinds and attention.

I have argued that there are important problems for debates that revolve around the question of whether a certain term refers to a natural kind. We can see these problems emerging again with regard to two objections to the arguments that have been developed in this thesis.

§6.5.1-Two natural kinds-based objections to my view.

There are at least two ways that an opponent of mine could attempt to use natural kinds to resist the central arguments of this thesis. The first would be accept my arguments for the claim that there are various different concepts of 'attention', but to claim that *all* of the entities that the various different concepts refer to share enough in common to form a natural kind. It could then be argued that this justifies preservation of the term 'attention': we should preserve the term attention because the different concepts that it is ambiguous between all collectively refer to a natural kind. This general approach would be a way of accommodating my main argument for pluralism (because it accepts that there are several distinct 'attention' concepts) but resisting the slide to pragmatic eliminativism (because it uses natural kinds to sanction preservation of the term 'attention').

In order to make this suggestion vivid, we can return to the quotation from Smithies that I gave at the beginning of this chapter. In that quotation, it was suggested that all of the entities that are referred to as 'attention' involve 'selecting information for some purpose or another'. I am unsure of whether this is true but let us assume for the sake of argument that it is. At this point, my opponent could say that all of the entities referred to by the term 'attention' all collectively form a natural kind, because they all share the common property of 'selecting information for a purpose'.

Such a response to my argument would be (at best) extremely unclear, for reasons that should now be obvious. On one interpretation of Boyd's analysis, the (vague and abstract) characterisation of attention as 'selection of information for a purpose' will be sufficient to count the entities referred to as 'attention' as forming an HPC kind. On a stricter (and equally legitimate) interpretation of Boyd's criteria, they will *not* count as a natural kind. As we saw above, there is nothing in Boyd's account to help us decide between these two options. What would result is a debate of the sort that we saw with Machery, Weiskopf, Samuels and Ferriera, Griffiths and Charland. We have one side using one interpretation of Boyd to oppose eliminativism and the other side using another interpretation to *support* eliminativism.

There is also a second way that my opponent could use natural kinds to resist my arguments, which stumbles for similar reasons. Rather than taking all of the entities that all of the concepts of 'attention' collectively refer to, and trying to find similarities between them, an opponent may admit that we have a range of different possible candidate meanings for the term 'attention' and that they refer to entities that are significantly different from each other. However, the opponent could say that we should select the concept that refers to a natural kind, and prefer that concept to the others. The idea is that we should reject the concepts which do not carve nature at the joints, and prefer the one that *does*.[10] This would be a way of opposing my argument for pluralism in the first place.

The first problem with this suggestion will be obvious now: if my opponent wishes to claim that a certain concept refers to a natural kind, then she must face the problem that a

[10] This suggestion has been made to me independently by Ned Block and Robert Foyle. Ted Sider (2011, pp.62-63) makes similar suggestions, though his notion of 'carving nature at the joints' is significantly different from Boyd's, and he does not mention 'attention'.

different enough interpretation of the core terms in Boyd's analysis will deliver the result that it is *not* a natural kind. A related problem is that if we employed a loose interpretation of Boyd's view, we will likely reach the conclusion that a *great many* of the concepts that have been put forward as candidate meanings of 'attention' will refer to natural kinds. Thus, my opponent's move may end up establishing *pluralism* about attention, on the basis that (on a loose understanding of Boyd) several of the 'attention' concepts refer to natural kinds.

To compound difficulties here, we can marry the claim that different theoretical interests will deliver different views on what counts as a 'scientifically relevant' property with an argument I made in chapter 3, that 'attention' has been put to a *great many* uses and that there is no reason to think that the theoretical interests of one group of thinkers must match up with those of another. For these reasons we would expect our judgement of whether a certain concept of attention refers to a natural kind to change, depending upon which theoretical roles we wish to use the concept for and thus how we interpret what it takes for a property to be 'scientifically relevant' when assessing whether the concept refers to a natural kind.

My opponent could accept that (on a loose interpretation of some of Boyd's criteria) several concepts of attention will refer to natural kinds. My opponent could however insist that it could still be that one concept refers to a kind that is *the most* natural one, and this one is to be preferred to the ones that refer to *less natural* kinds. On this view, we should not select the concept that refers to *any* natural kind, but we should only accept the concept that refers to the *most natural kind.*

However, this suggestion just drives many of the problems back a step, and invites new ones as well. The first problem is that which concept of 'attention' refers to a kind that

is the *most* natural will itself be open to interpretation. This is (among other things) because it is open to interpretation how we should weight the different criteria in Boyd's analysis. Suppose someone placed a great deal of weight on the claim that a certain set of entities should support inductive generalisations, and they weighted this criterion much more strongly than the criterion that a set of entities should share a large number of scientifically relevant properties. We would expect this person to have a different view on whether a certain kind was the 'most natural' from someone who weighted these criteria the other way around. Again, I see nothing in Boyd's analysis that can help us resolve this issue.

Furthermore, the claim that we should prefer the term that refers to the *most natural* kind is independently dubious. Science often makes use of different terms, which vary in the levels of description involved, and correlatively in how 'natural' the sets of entities that they refer to are (at least by Boyd's criteria). This kind of point is familiar from the long-running debates over how the kinds of the special sciences interact with the kinds of fundamental physics (e.g. Block and Fodor, 1972; Fodor, 1974 and 1997). Of course, it is also true that *individual* disciplines make use of different levels of descriptions, some of which describe their target phenomena in great detail, others of which discuss them in extremely abstract terms. A good example of this in psychology is David Marr's famous 'three level' theory of vision (1982).[11] It is a point that has been made before (but which bears repeating) that different levels of analysis have often been used simultaneously to aid explanation and prediction in science. This gives us good reason to be sceptical of the claim that we should select which concept refers to the 'most natural' kind and reject the other concepts.

[11] See also Carl Craver's (2009, pp.585-589) discussion of the different levels of abstraction at which the hippocampus has been studied. See also Piccinini and Craver (2011) and Bechtel (2008).

§6.5.2-'Mechanisms'.

At this point, it is worth raising a related difficulty for using the HPC view to decide whether a certain term refers to a natural kind, which stems from the idea of a 'mechanism'. Recall that (to Boyd) there must be a homeostatic/causal mechanism that tends to produce the cluster of properties that are definitive of the kind in question. Therefore, whether we take two entities to be a member of the same kind is contingent on whether we think that they are supported by the same mechanism. Remember that there is no a priori restriction on what a 'mechanism' could be in Boyd's theory. The issue is that a large element of *decision* is involved in deciding whether two entities are supported by the same mechanism. If this is the case, there will be a great deal of decision involved in whether we claim that a certain set of entities form a natural kind or not.

We can draw out some of these problems here with a concrete example. Take one of Rensink's concepts of attention as the selection of information by the retina for transmission along the optic nerve. We can compare this with Prinz's concept of attention as availability to working memory. Suppose we want to decide whether these concepts refer to processes that are caused by the same mechanism. One candidate mechanism here is genetic inheritance. Genetic inheritance is one of the archetypal mechanisms that support Boydian natural kinds such as biological species. Certainly the systems underpinning both Rensink's and Prinz's concept can be seen as the result of genetic inheritance (along with some environmental factors). We are genetically predisposed to have functioning retinae, as well as to have functioning working memory systems which allow information to be encoded and used for deliberation, report etc. So, when we take the candidate mechanism of 'genetic inheritance', we reach the conclusion that they are caused by the same mechanism. Thus, they could fall within the same kind.

However, the retinae and working memory have different locations in the nervous system, and there are (of course) some differences in function of these different areas. So, we could say: 'the mechanism that underpins Rensink's concept of attention is the retina, which has these properties […] whilst the mechanism that underpins Prinz's concept is working memory, which has these ones […]. So, the entities that Prinz and Rensink are talking about are subserved by different mechanisms. Therefore, they do not fall within the same kind'.[12]

The problem is this: when we select a particular mechanism as the cause of a certain set of properties, we must group together that set of properties with other properties that the mechanism also causes (because it is essential to the entities in a Boydian natural kind that they be caused by the same mechanism). This means that *which* mechanism we select will impact on what *other* entities can fall within the same kind as the set of properties we are investigating. But it is often very unclear what could lead us to selecting *one* candidate mechanism over another.[13]

§6.5.3-Endogenous and exogenous attention again.

I shall close the critical element of my discussion with a brief case study. Consider the distinction between exogenous and endogenous attention, which we encountered in chapter 2. There has been debate over whether the endogenous/exogenous distinction holds up (Awh et al., 2012); and there is a great deal of debate over the neural correlates of these two kinds of attention (see Corbetta and Shulman, 2002 and Peelen et al. 2004 for two

[12] The problem can be put in more abstract terms: suppose we have two clusters of properties, A and B. Sometimes there will be one candidate mechanism (say θ) that generates the properties in A, but *does not* generate the properties in B; and another candidate mechanism (say ϕ) that generates the properties in *both* A and B. If we count θ as the mechanism that generates A, then we will conclude that A is caused by a mechanism that does not cause B, and thus they do not fall within the same kind. If we take ϕ to be the mechanism that generates A, then we conclude that A and B are caused by *the same* mechanism, and thus they could fall within the same kind.

[13] For more on 'mechanisms' in psychology, see Machamer et al. (2000) and Craver (2009).

radically opposing views). However, let us set these controversies aside and ask the following question: given the available empirical evidence, do endogenous and exogenous attention form a natural kind?

So, we must look at the two kinds of attention, and see if they have enough in common to qualify them for natural kindhood. When we do this, we find that endogenous and exogenous attention do indeed have some commonalities. For example, directing either kind of attention to a stimulus makes subjects faster (and often more accurate) to respond to those stimuli. Both kinds of attention can increase a subject's sensitivity to the *contrast* of a stimulus (Ling and Carrasco, 2006). Both kinds of attention impact on perceptual processing in the occipital, parietal and frontal cortex (Carrasco, 2011, p.1488). When either one is directed toward a square that has a gap in it, it is easier for subjects to correctly identify which side of the square has the gap. It is also *more difficult* to discern the location of a gap on a square when attention is directed *away* from it, for both the exogenous and endogenous kinds of attention (Montagna et al. 2009).

However, endogenous and exogenous attention also have differences. The former is voluntarily deployed whilst the latter is not. They also take different amounts of time to engage. Exogenous attention takes about 100ms to engage, and the performance of exogenous attention peaks at about 100-125ms, and disappears by about 250ms. By contrast, endogenous attention takes about 300-500ms to engage, and can be maintained for as long as the subject wills it (Carrasco, 2011, pp.1489-490 and Montagna et al. 2009, p.735). Strikingly, it was found that directing exogenous attention to a particular location *impairs* subjects' ability to discriminate which of two dots appeared first at that location. However, endogenous attention *improved* subjects' abilities to perform this task (Hein et al. 2006).

Finally, it has been claimed that exogenous attention is highly developed in babies, whilst endogenous attention does not develop until much later (Gopnik, 2007).

I could continue the iteration of similarities and differences between these two kinds of attention (see Carrasco, 2011, pp.1486-1500), but let us return to the natural kinds question with which we began. It is clear that similar problems as have already been encountered could easily occur here: one side could emphasise the similarities that the two kinds of attention have, and then reach the conclusion that the two kinds of attention fall within the same Boydian natural kind. An opponent could then say that these similarities are not enough and demand many more similarities before she allows that the two kinds of attention could qualify for natural kindhood. Such an opponent could then emphasise the differences that the two kinds of attention have, and use these to argue that the exogenous and endogenous attention are very different, and should be studied differently. Of course, the discovery of *more* similarities or differences will not help the debate here, because Boyd's criteria for natural kindhood can be stretched one way or the other, so that it will always be possible to reach either conclusion. As before, the problem is *not* with lack of empirical knowledge, but with the HPC view itself.

§6.6-Getting to eliminativism without natural kinds.

Nothing I have said about natural kinds in this chapter supports pluralism or eliminativism about 'attention' or about any term within psychology. Rather, I consider the natural kinds issue to be too murky to help us decide one way or the other, so I recommend an embargo on arguments of this type. The rest of this chapter will be more optimistic; I shall argue that the views of Machery, Griffiths and Irvine should be reconstructed so as to avoid the problems that I have pointed out.

We can uncover some of the important ideas here by asking the following question: suppose that we can all agree that a certain term *does not* refer to a natural kind. How do we get from *this* claim to the claim that the term in question should be eliminated from a particular set of practices? What provides this extra step? Machery is aware of this question, and indeed highlights the need for extra argument to take us from the claim that a term does not designate a natural kind to the claim that the term ought to be eliminated. Machery says this:

> 'The main considerations that bear on [this issue] are pragmatic. A theoretical term that has been found to fail to pick out a natural kind should be kept if it plays a useful role. Not all terms in science are assumed to pick out natural kinds… However, by the same token, a theoretical term that has been found to fail to pick out a natural kind should be eliminated if it fails to play a useful role or if it plays a harmful role. I believe that the latter is likely to be the most common case' (2009, p.239. cf. his 2005, p.465).

Indeed, Machery *does* think that the term 'concept' is doing harm in psychology, he says :

> 'The notion of concept ought to be eliminated from the theoretical vocabulary of psychology because it might prevent psychologists from correctly characterizing the nature of knowledge in long-term memory and its use in cognitive processes' (2009, p.220).

Machery concludes:

> 'Thus, elimination of "concept" would probably help reframing the research on concepts and eliminate the unproductive controversies between proponents of different paradigms' (2009, p.243).

So, Machery thinks that a term that does not refer to a natural kind should be eliminated *provided there are good pragmatic reasons to do so*. It is important to stress this: for Machery the fact that a certain term does not refer to a natural kind is *not* what takes us to eliminativism; what takes us there is theoretical pragmatics. Whilst Machery places by far the most weight on the claim that 'concept' does not refer to a natural kind, the main impetus for eliminativism is still pragmatic: we should eliminate 'concept' because it is theoretically useful to do so.

My suggestion is this: given that we have encountered so many problems when it comes to natural kinds, why not simply stay quiet on whether the terms refer to natural kinds, and state the eliminativist argument *entirely* in terms of theoretical pragmatics? Consider Machery's argument for eliminativism about 'concept'. I suggest that we remove the claim that 'concepts are not a natural kind' and simply state the argument like this: 'it is theoretically useful to do away with the term 'concept' for the following reasons…' Concerns of theoretical pragmatics are going to be needed to take the argument to eliminativism anyway, so why not allow the argument to rely *exclusively* on these facts, and do without troublesome claims like 'concepts are not a natural kind'? Here we preserve the impetus for eliminating 'concept' without becoming enmeshed in the difficulties with trying to argue that 'concept' is not a natural kind.

This kind of proposal can apply *mutatis mutandis* to Irvine's arguments for eliminativism about 'consciousness'. Irvine says:

> 'Arguing that 'consciousness' fails to pick out a scientific kind provides further support for an eliminativist claim about 'consciousness' (2013, p.157).

However, as I pointed out above, the link between denying that a certain term designates a natural kind and embracing *eliminativism* about the term is somewhat obscure. Here is one way that Irvine takes us to eliminativism about 'consciousness':

> ''consciousness' is not a viable target for scientific investigation and instead promotes methodologically flawed research programs' (2013, p.158).

Here we can clearly see concerns of theoretical pragmatics creeping into the motivation for eliminativism. We see this surfacing once again in Irvine's most explicit statement of the link between denying that 'consciousness' is a natural kind and embracing eliminativism about it:

> 'If a concept does not itself refer to a scientific kind, but to a group of scientific kinds, then its continued use may lead to confusion. For example, if X refers to a group of scientific kinds (a,b,c) then asking questions about X can be interpreted as questions about any of (a,b,c), all of which have different answers. Debates may then ensue in which all sides talk past each other' (2013, pp.162-163).

As with Machery, Irvine's arguments ultimately take their force from theoretical pragmatics and again, Irvine's arguments can be reformulated to avoid mention of natural kinds. Irvine can say: 'the term 'consciousness' stands in the way of scientific research for the following reasons: it is ambiguous, it gives rise to verbal disputes(…) and should thus be eliminated.' This retains the power of Irvine's arguments without commitment to troublesome claims about natural kinds.

In summary, my view is this: there is often no satisfactory answer to the question of whether a certain term refers to a natural kind. It is also extremely unclear how the claim that ''x' does not refer to a natural kind' takes us to eliminativism about 'x' *unless* we bring in concerns related to theoretical pragmatics. But once these concerns are brought in, claims about natural kinds can safely drop away from the argument, and the pragmatic arguments can stand alone.

An important qualification is that reconstructing the arguments in question to avoid mention of natural kinds may be a substantial job. This applies especially to Griffiths, who merges together claims about theoretical pragmatics with claims about natural kinds in a way that means it is not obvious how to separate them (e.g. Griffiths, 1997, pp.228-235). Above, I gave some examples of how such reconstruction could be carried out on some of Irvine's and Machery's core claims, so I think we have good reason to be optimistic about the prospects of reconstruction.

This is not to say that I do think that 'concept', 'emotion' and 'consciousness' should be eliminated. Machery, Griffiths and Irvine all put forward many rich and varied arguments which would need to hold up before we could reach that conclusion. What I am saying is that their route to eliminativism faces issues that can be resolved by employing a strategy more similar to my own. *If* we want to be eliminativists, then *this* is the way to do it.

§6.7-An objection and a reply.

An opponent of mine could insist that by setting the natural kinds issue aside, I have missed something of major importance. They may say that when a term *does* refer to a natural kind, then it is a term that *carves nature at the joints* and this gives us good reason to

preserve it. The worry driving this suggestion is that if we do shed a term that refers to a natural kind, we will have given up on a particularly privileged way of describing the world. For this reason, it is crucial that we consider the natural kinds issue when deciding whether a term should be eliminated or not.

In many ways this objection misses the point of what I have been saying because the objection takes as its starting point the claim that we should not do away with terms that refer to natural kinds. For this reason, the objection assumes that we *can decide* that a certain term refers to a natural kind. I have been arguing that this is usually not the case in the debates in question, and that there are principled reasons why this should be so. Another point to make against this objection is that it gains its force from the claim that natural kind terms carve nature at the joints in some strong and important way. Much of this chapter has been aimed at undermining this intuition, because as we have seen whether a certain class of entities is a natural kind is often more down to interpretation and decision than the structure of the world.

Furthermore, it is worth questioning the inference from ''x' refers to a natural kind' to ''x' should not be eliminated' in the first place. Suppose we were convinced that a certain term 'carved nature at the joints' better than another set of terms in the vicinity. What reason do we have to think that *this* was sufficient to safeguard the term from elimination? Why should the fact that the term in question carves nature at the joints override pragmatic considerations? The answers to all of these questions are (at best) extremely unclear, and until we have good answers to them, this objection will lack force.

§6.8-Conclusion.

I really did not want this to happen. I tried for a very long time to decide whether attention was a natural kind. Eventually I realised that I couldn't decide, and that there were good philosophical reasons why I couldn't. It is these reasons that I have tried to explain in this chapter. What makes me especially keen on emphasising these problems is that I keep seeing them in more or less *every* debate over whether something is a natural kind. At the very least, I hope I have shown that these issues deserve considerably more discussion than they are currently receiving.

Chapter 7
Theoretical Use, Classification and the Future of Attention Research.

§7.0-Summary.

This concluding chapter has three aims. Firstly, the arguments of the thesis so far will be summarised. Secondly, some points that have received insufficient attention in previous chapters will be outlined and expanded upon. This will include a more thorough explanation of precisely what has fuelled the differences in concepts of 'attention' that have occupied this thesis extensively. Thirdly, some comments about how I think attention research should develop in coming years will be made. In general, the chapter aims to draw on much of what has been said in previous chapters of the thesis, connecting together issues and placing them in the context of a coherent and complete view.

§7.1-The story so far.

The current piece of work is the first dedicated and sustained philosophical argument in favour of pluralism about attention. I have presented arguments in favour of my view, defended it from objections and presented criticisms of other views. I have also offered arguments for what I think is a plausible extension of the view, which leads us to a variety of eliminativism about 'attention'. In what follows I shall summarise the original contributions of each chapter.

In chapter 2 I argued that the debate over whether attention is necessary and sufficient for consciousness really comes down to different interlocutors having different views on the meaning of the word 'attention'. The locus of the discussion around the sufficiency claim included the results surrounding the GY experiments, the meta-contrast masking experiments and Smithies' argument in favour of the sufficiency claim. The 'dual-task' results surrounding the necessity claim were also examined. I claimed that the best way to resolve these difficulties is to focus on certain conceptual issues to do with what we mean by 'attention'. The rest of this thesis has been performing the tasks which I think are

crucial if these debates are to move forward fruitfully. This chapter and the paper that it is based on (Taylor, 2013a) represent the first thorough investigation and analysis of these problems.

Chapter 3 examined the more optimistic reaction to these problems, which is the claim that we should decide which account of attention is 'better' than the others. In that chapter, we examined the use of 'attention' in folk psychology, empirical psychology and in terms of 'theoretical use'. Again, it is my view that these issues have generally been under-analysed in the literature. Often philosophers and psychologists have claimed that 'the folk' have one settled view on the matter of what attention *is*. I argue that such a view should be rejected. As a result, I argued that attempting to ally our account of attention to folk psychology is a grim route to go down. Furthermore, I argued that things are little better in the empirical literature. I have also focussed on the sheer number of different theoretical roles that attention has been put to. Finally, the chapter offered a preliminary argument for pluralism about 'attention' which is based on the claim that pluralism most accurately represents the way 'attention' is used in empirical and folk psychology. This chapter adds substantially to several debates regarding the folk psychological usage of the term 'attention', the theoretical use of 'attention' and how 'attention' has been understood in empirical psychology.

Chapter 4 turned towards the thinkers who have attempted to give reductive analyses of attention, where this is understood as giving necessary and sufficient conditions for 'attention' in non-circular terms. I labelled this view 'attention essentialism'. I examined several particularly prominent versions of attention essentialism. After arguing that all of them have difficulty, I gave arguments for the conclusion that attention essentialism (as a

theory of the folk or empirical concept of attention) should itself be rejected. Also in this chapter, the main argument for pluralism (which I call the 'master argument') was presented and defended from some initial criticisms. This chapter and the paper that it is based on (Taylor, 2014) represent the first analysis of the attention essentialist approach as a whole. The chapter also puts forth a unique and original argument for pluralism about 'attention' which draws on chapters 2 and 3.

Chapter 5 was a detailed examination of one particularly prominent criticism of my view. This was the claim that pluralism leads to *eliminativism.* Indeed I would go so far as to say that this criticism is the *only* attack on pluralism that has received any real emphasis in the literature on attention. Whilst maintaining its primary focus on attention, chapter 5 also attempted to widen the purview of the argument, extending it beyond 'attention', and to other concepts that have received pluralist or eliminativist treatments. This was in the interests of making the claims more relevant to a wide variety of debates, and making the overall position I advocate more panoptic. I argued that pluralism need not imply what I called 'ontological eliminativism'. I employed some plausible theories from metaphysics to make this case, and ultimately defended a variety of realism about 'attention'. I also identified another variety of eliminativism which I call 'pragmatic eliminativism', which I argued we have good reason to accept. This chapter is the most thorough analysis of the different kinds of eliminativism in this field, it draws debates about attention together with debates in contemporary metaphysics over truthmaking and realism and the chapter also contains an original and detailed argument for pragmatic eliminativism about 'attention'.

Chapter 6 was an examination of how eliminativism in psychology is related to issues to do with natural kinds. This issue is relevant to the arguments of this thesis, as well

as several other prominent arguments for eliminativism in philosophy of psychology. The views I particularly focussed on were Machery's and Griffiths' arguments for eliminativism about 'concept' and 'emotion' respectively, but Irvine's argument for eliminativism about 'consciousness' also received discussion. It was argued that these arguments stumble because the account of natural kinds that they assume makes the question of whether a certain term refers to a natural kind more one of interpretation than fact. I argued that we should abandon the entire natural kinds based approach to issues surrounding eliminativism. I offered what I think is a way to reconstruct the eliminativist arguments in question. This chapter also considered various natural kinds-based objections to my own view, and argued that these objections stumble for similar reasons. This chapter provides original contributions to various debates over the status of 'attention', 'emotion', 'concept' and 'consciousness', it draws general conclusions about how eliminativism in psychology relates to natural kinds and it also raises questions over whether Boyd's view of natural kinds is particularly useful to psychology at all.

In the master argument for pluralism that I put forward in chapter 4, I claimed that certain attention concepts should be preserved because they are 'theoretically useful' and gave some examples. I also promised to put more flesh on these ideas. In this chapter, I turn to this task. In §7.2 I will offer a plausible story of how differences in theoretical use drive differences in classification; and thus how different theoretical uses can be expected to drive differences in concepts of 'attention'. I also examine in more detail the process of how we go about deciding if a concept is 'theoretically useful' and thus worthy of preservation. These arguments will lead me on to considering a final criticism of my view (§§7.2.3). In §7.3 I will

sketch out how we can expect these debates to develop, if my view is taken seriously. Finally in §7.4 I will close with some remarks about the project I have been engaged in.

§7.2-Theoretical use and pluralism.

Sometimes the interests of one group of people who use a term do not match up with the interests of another group of people who use the same term. What often results is different parties classifying the world differently, but using the same term to do so. I believe that this is what is happening in the 'attention' debates. Explaining how this comes about will allow us to get in mind what factors I think have driven differences in different 'attention' concepts, and it will also provide an explanatory framework in which to understand how the 'attention' debates have developed in the ways that they have.

§7.2.1-Tomatoes and fruit.

A point which was recognised at least as far back as Locke, but which bears repeating, is that classification is *interest relative.* How we choose to classify entities in the world depends (among other things) on what our concerns and interests are when we carve up the world using language.[1] To see this, take the question of whether a tomato is a fruit. Biologists have certain interests when taxonomising the world: they are interested in explaining why there are certain similarities in nature, in making predictions about how a group of organisms will behave, and extrapolating information about a subset of organisms to the group of organisms as a whole. Given the interests of biology it makes sense to group together all entities that are the ovaries of flowering plants, because such a taxonomy allows us to explain why all the ovaries of flowering plants tend to share properties that biology is interested in. The use of the word 'fruit' in biology is thus applied to those things that are

[1] John Dupré is particularly well known for examining these kinds of issues (1981, 1993 and 1999) and much that I say below will be indebted to his work. See also Sainsbury (2014) and Philips (2014).

the ovaries of flowering plants, as this is a biologically useful set of individuals to group together. That is why it makes sense to count a tomato as a fruit, if one is a biologist.

To a chef, things are different. The properties of tomatoes in virtue of which chefs are interested in tomatoes are not that they are the ovaries of flowering plants. Chefs are not interested in genetic inheritance, evolutionary adaptation and so on. Rather, chefs are interested in gustatory and aesthetic properties of foods. So to a chef, the fact that a tomato has a very low sugar content compared to apples, pears, raspberries and so on is much more important than the fact that they are the ovaries of flowering plants, because sugar content affects taste so much. For this reason it makes sense to class tomatoes with things that have similar sugar content like carrots, courgettes, aubergines etc. This is why chefs count tomatoes as a vegetable. It would make less sense to classify them as a fruit, because tomatoes do not share many properties that chefs are primarily interested in with other fruits like apples and pears.

Here biologists are employing one 'fruit' concept (which groups organisms in one way), and chefs are employing another (which groups organisms differently). Notice that the two 'fruit' concepts are each sensitive to differences that the other one is not. The chef's concept can be used to carve up organisms based upon the gastronomic and aesthetic properties of them, whilst the biologist's applicability is completely insensitive to such divisions. Similarly, the biologist's concept is sensitive in a way that the chef's is not. It is important to hammer this point home: each one can be used to make divisions that the other cannot make, so which concept we employ will depend upon whether what we are doing depends crucially on making those divisions.

A core point to bear in mind is that the classificatory systems of both the chef and the biologist are *non-arbitrary* and *useful*, it is just that they are useful for different purposes. This can easily be obscured by the fact that on the surface, chefs and biologists can seem to disagree about some empirical fact. When one side asserts an utterance like: 'a tomato is a fruit' and the other side asserts: 'a tomato is not fruit', it is natural to assume that they disagree about some property of tomatoes which is presumably discoverable by empirical means. The truth of course is that the two parties do not really disagree about some empirical fact, they are just using the word 'fruit' differently because one party finds it useful to use the word 'fruit' in one way, and the other side finds it useful to use it another way.

§7.2.2-Theoretical uses and the different 'attention' concepts.

What the 'fruit' example shows is how interests drive classificatory practices, and how different interests can drive differences in the meanings of concepts. There are striking similarities in the 'attention' case: different theoretical uses drive the difference between different definitions of the word 'attention'.

As we have seen, one of Kentridge's definitions of attention is in terms of *task performance*. Kentridge is engaging in a classificatory practice (defining attention in terms of task performance) that is being driven by a certain theoretical interest (using the means of contemporary psychology to empirically test certain abilities of subjects in relation to invisible stimuli in valid and invalid trial conditions). Elsewhere, of course, we have Prinz's concept. Recall from §4.6 that there are certain empirical results that have been used in order to attack the global workspace theory of consciousness. As we saw in that section, assessing how much these results are a problem for GWT requires certain conceptual distinctions. It is

this role that *Prinz's* concept can fulfil, and these distinctions that the concept is useful for. This is part of what makes Prinz's concept the way that it is: Prinz wanted a concept of 'attention' that could do this particular work.

As was the case with the two 'fruit' concepts, each concept makes distinctions that the other is insensitive to. Kentridge's concept is insensitive to the difference between information broadcast in working memory, information accessible to working memory, and inaccessible information. On Kentridge's concept, so long as subjects have increased task performance in certain situations then they will count as attending to the items in question, no matter how the information they have about the items stands in relation to working memory and the global workspace. Kentridge's concept also does things that Prinz's cannot. Kentridge's concept can be used to track increased task success with relation to verbally unreportable stimuli, whereas (as I argued in §2.6) Prinz's concept is insensitive to these issues. Kentridge's concept is useful if we are interested in doing what Kentridge is doing: investigating increased or decreased task performance with relation to invisible (and thus unreportable) stimuli.

It is striking how similar the 'fruit' and the 'attention' cases are: in both cases, each subconcept can perform functions the other cannot, and each is non-arbitrary and useful. For all of these reasons, it is entirely unsurprising that the two concepts should differ and it is also unsurprising that they should be useful for different purposes. The reason that they differ in the first place is *precisely because they are meant for different purposes.*

This general picture has wide application. Wayne Wu is primarily interested in explaining *action,* and this explains why his concept of 'attention' places such an emphasis on its role in action. Wu's concept is insensitive to distinctions the other concepts *are*

sensitive to, but it is sensitive to some distinctions that others are *not* sensitive to (it is used to explain the difference between mental action and automaticity). This is because these are the distinctions that a concept *has* to be sensitive to *if we want to do what Wu is doing*. Smithies is interested in the connection between rational justification and phenomenal consciousness, which explains why his concept places an emphasis on this. Again, if we are interested in what Smithies is interested in, we will need a concept like Smithies', and this is why his concept is the way it is.

It is not just that different thinkers are picking different definitions of attention more or less at random, resulting in the differences in question. Something more sophisticated than that is going on: the difference in how different thinkers define attention is driven (in large part) by differences in how they think it *should* be defined, and this is in turn driven by the uses for 'attention' that they have in mind. As was the case with fruit, the difference arises from a different classificatory scheme, and *that* arises because of different interests.

This is another strong reason to think that we should not expect agreement about how 'attention' should be defined. If it were the case that the different sides of the debate were selecting different definitions arbitrarily, then perhaps we might expect them to be willing to shed their own definitions in favour of another, and we could expect eventual agreement. However, we cannot expect this in the current context because accepting an opponent's 'attention' concept would involve giving up on the distinctions that one's own concept makes salient, which will be unacceptable if one is wedded to theoretical purposes for which such distinctions are essential.

§7.2.3-Objection: redundancy.

At this point, a worry might be raised regarding my reason for preserving certain concepts of 'attention'. Recall from §4.7 that I justified preservation of various 'attention' concepts by claiming that they are useful in certain theoretical contexts. One of my examples was Prinz's concept, which is useful for assessing certain empirical results pertaining to global workspace theory. Part of my case for thinking that the concept would be useful involved showing that it maps on to a set of conceptual divisions which have already gained currency and importance in cognitive science. Specifically, I likened his distinction to one that is made by Dehaene and Naccache (2001) and Dehaene et al. (2006). This may raise a question about redundancy. An opponent may say that we really *do not need* Prinz's concept at all, because Prinz is just echoing what Dehaene has said. The idea here is that anything that Prinz's concept can do Dehaene's concept can also do, therefore Prinz's concept is not required.

My first response to this objection is to say that I am sceptical of the claim that Prinz's concept is *exactly* the same as Dehaene's concept, and as such I do not think that *everything* that Prinz's concept can be useful for would also be fulfilled by Dehaene's. The differences I have in mind are based around the fact that Prinz's concept makes mention of availability *to working memory*, thus linking the issues to do with global workspace theory to issues to do with working memory. I think that this is important, as versions of global workspace theory have often been extremely vague over the issue of precisely how the 'global workspace' is supposed to interact with working memory (or whether they are simply supposed to be the same thing). Recall from §2.4.2 that Prinz's account of working memory is very specific, and must be carefully distinguished from other accounts if we are not to fall victim to exactly the kind of problems we have been encountering in this thesis.

So I am not convinced that Prinz's concept is *exactly the same* as Dehaene's. However, even if we assume that Prinz's concept *is* the same, I do not think that this objection is especially worrying. Recall that we are interested in whether a certain concept is theoretically useful. If it really were true that Prinz's concept perfectly matches up with concepts that have already been deployed in the literature, that would not show that Prinz's concept was useless but rather that the concept *is* useful because it is the same as one that has been used all along. Prinz's concept would be the same as one already at play in the literature and it is *this concept* that we would want to preserve, whether it is original to Prinz, or Dehaene, or whoever. If we do have a concept which matches up with another one in the literature then I do not think it is important who we attribute it to, in fact I suspect that it may be more useful simply to call the concept the 'availability to working memory' concept rather than attributing it to one particular thinker. What is important is that we recognise the potential uses of a concept and separate it out from those concepts that have different uses, whether or not it is original to Prinz. I claim that the concept in question does have such uses and that it is thus worthy of preservation.

Recall that I justified preservation of certain 'attention' concepts by saying that they are *theoretically useful.* Importantly, I do not think that thinkers necessarily need *to be aware* that their concept has a certain theoretical use in order for the concept to be so useful, and thus worthy of preservation. Concepts such as 'prime number' are perfectly useful when we are doing cryptography, whether or not this use was recognised by those that first introduced the concept of a prime number. Similarly, concepts of complex and imaginary numbers are useful in electrical engineering, even though we wouldn't want to claim that the mathematicians who originally introduced those concepts were aware of this usefulness.

§7.3-Where do we go from here?

In a nutshell, what I have been arguing over the course of this thesis is that many of the debates surrounding 'attention' only have traction if we assume that there is some entity which is particularly deserving of the name 'attention' and which has a substantial nature which can be discovered in some way, presumably by science or philosophy. This (I have argued) should be rejected. Differences in views over what attention is and what attention does are often mere differences in how the thinkers in question understand the concept 'attention' and (crucially) no one of them is correct above the others, because they are driven by different theoretical interests, and different purposes.

I have given some examples of concepts that I think should be accepted as useful. However, I did not offer a list of all and only those attention concepts that are useful and thus worthy of acceptance. In fact I think we should resist the urge to give a complete list of attention concepts, and to claim that no other attention concepts will be useful (and thus worthy of acceptance). This is because we do not know how attention research will continue to develop in the coming years. Subsequently, we do not know which theoretical roles will become important in the future and whether new concepts will have to be introduced in order to take the theoretical weight of our future investigations.

This comment about 'future research' raises an interesting question. Suppose you accept much of what I have been saying. This raises a question of how I think that attention research *should* develop if my view is taken seriously. What would remain the same? What would change? In this section I will summarise how I think research should continue and how our general approach may change, if the pluralistic view is embraced. The take-home

message is that I think that the debates over attention can benefit from a much greater emphasis on theoretical pragmatics.[2]

In general, I think that we should shift away from questions of the form 'what is attention?' or 'is theory x the *correct* theory of attention?' and 'does 'attention' have property x?' and towards questions such as 'what are the different concepts of 'attention' that have been delineated, and what work can they do for us in our theories?' The questions we should concentrate most of our efforts on should be more those of *what good* a certain concept can do, rather than whether that concept offers us a good analysis of 'attention', or whether it is 'better' than other proposed concepts of attention. This kind of shift of focus naturally follows from two things which I have argued for extensively over the course of this thesis. The first is the rejection of the hunt for a 'correct' theory of attention. The second is an acceptance of the claim that theoretical use is a reasonable criterion to judge 'attention' concepts by. These two commitments lead us to the normative view that I advocate.

This kind of shift is sometimes associated with pragmatism. I would not call myself a pragmatist, as that term has been understood in many different ways. However, it is worth pointing out that in some versions of contemporary pragmatism something like the following claim is made about how to approach conceptual analysis generally: 'instead of asking "What is *X*?," one should focus on the roles one wants *X* to play and see what can play that role' (Chalmers, 2011, p.538. We find something like this shift in other contemporary pragmatists as well, e.g. Price (2003 and 2011)). I wholeheartedly agree with this view.

[2] Of course, it will not come as a surprise that I hold this view, given the emphasis on theoretical pragmatics in my arguments in §5.6 and §6.6.

It is worth noticing just how far we have slid in our examination of the issue of how attention relates to consciousness. We began by asking whether attention is necessary and sufficient for consciousness, and we were forced to address the question 'what is attention?' Now we have slid further away from this question, towards the question of 'what concepts have been called 'attention' and what work can these concepts do for us?'

The rejection of the question 'what is attention?' may strike some as extravagant, given its prominence in many circles within attention research. However, notice that the more theoretically useful concepts of 'attention' we delineate, the less interesting the residual question 'what is attention?' will become. One of the main reasons we became interested in attention in the first place is because we hoped that it could shed light upon other interesting faculties of the mind (such as consciousness, modesty, knowledge of demonstrative reference and so on). By introducing the concepts which can aid us in charting these relationships, we start to obtain answers to these questions, independently of whether we ally one particular concept to 'attention' above the others. Of course, to some thinkers (e.g. Mole, 2011a) the question 'what is attention?' will have independent interest, regardless of whether or not the roles that 'attention' has been put to can be fulfilled by a range of concepts that have at one time or another been called 'attention'. However, if what I have been saying is correct, then pursuing the question 'what is attention?' is something of a waste of time; and thus giving up on the question is no loss.

We will have to give up on a lot of other questions that are currently prominent in attention research as well. Most obviously, if what I have been arguing is correct, then the question of whether attention is *sufficient* for consciousness will not be a substantive and independently interesting question. Rather, the answer will flow from which concept of

'attention' we choose to use, which will itself be derivative on the theoretical context in which we discuss the issues. If what I have said is correct, there will be no one privileged answer to the question of whether attention is sufficient for consciousness, *because* there is no privileged answer to the question of what attention *is*. Indeed, on the approach I recommend, many of the questions which revolve around whether attention has a certain property will have to be given up on, as a great many of them will simply reflect differences in the concepts of 'attention' at play.

My view strikes some people as very radical, it can also appear as though I advocate a simple rejection of all of the work on attention that has gone before. However, the approach I recommend is not intended to simply sweep aside all the work on attention that is currently out there. Rather, the shift towards pragmatism is intended to *aid* this work, by rejecting certain questions which are obfuscating important issues, and focussing our efforts on the more substantial and interesting questions that the debates are circling around.

§7.4-Concluding thoughts.

In my view, one of the key jobs that philosophy of psychology and philosophy of mind should be in the business of doing is philosophical reflection on the conceptual underpinnings of work on the mind. This includes both reflection on the science of the mind, and metaphilosophical reflection on philosophy of psychology and philosophy of mind themselves. I believe that philosophers are particularly well set to carry out such a job, given their training in the analysis of argument and debate, and their understanding of subtle issues that arise related to language and how it can trick us. I also believe that the successful completion of such conceptual philosophical work is a clear form of progress in the overall enterprise of understanding the mind. Indeed, I think that it is just as important a kind of

progress as *any* philosophical attempt to get to grips with human beings' mental architecture and the place of such architecture in the natural world. I am also convinced that empirical work in psychology and neuroscience can benefit greatly from such philosophical analysis and that if empirical science ignores what philosophers have to say then it will be at best impoverished and at worst completely crippled.

Most of the core ideas in this thesis have been presented at one time or another over several years at more conferences than I am interested in remembering. I have received extensive criticism and feedback from philosophers and psychologists alike. I have subsequently refined my views, and responded to all of the serious criticisms I have received. A philosopher can do no more than this.

Nonetheless, the analysis that I have offered in this piece of work is far from perfect. I am not so naïve as to expect a reader to agree with everything I have said. I am sure that every single one of my arguments could be challenged in serious ways by someone sufficiently enamoured with the opposing view. Indeed I would be delighted if this did occur, since the common reaction to many ideas in philosophy is deafening silence. Having said all of that; I do think that my view is reasonable, as are the arguments that I have given in favour of it. Certainly I think that my arguments *at least* show that pluralism deserves to stand as a prominent and plausible alternative to the other options available in the literature.

As a final thought, I will quote *Naming and Necessity.* As readers will doubtless be aware, in this book Kripke launches a famous attack on the Fregean view of reference, and attempts to replace it with his own. When reflecting on the Fregean view, Kripke says this:

> 'What I think the examples I've given show is not simply that there is some technical error here or some mistake there, but that the whole picture given by this theory of how reference is determined seems to be wrong from the fundamentals…What I am trying to present is a better picture-a picture which, if more details were filled in, might be refined so as to give more exact conditions for reference to take place.' (1980, pp.93-94).

In this thesis, I have tried to offer what I think is a *better picture* of what attention is like, which begins from a different starting point from my opponents' view. I have attempted to offer another way of thinking about the issues that are currently being studied in attention research. I have tried to identify some common assumptions that are taken for granted in much work on attention, and give arguments to lead us to reject them. These assumptions deserve to be explicitly stated and rigorously assessed, not merely taken for granted. In addition to assessing these assumptions, I have also tried to construct an alternative view of attention and an alternative methodology for approaching the debates that surround the troublesome concept. I do not think it is too arrogant to say that this is a case of philosophical progress.

Appendix.
Is the grain of vision finer than the grain of attention? Response to Block.

Note: The material in this chapter originally appeared as Taylor (2013b). Unlike the main chapters of this thesis, its focus is very much first-order rather than metaphilosophical. I include it here as an appendix simply because it represents an important contribution to recent debates that centre around attention and its relationship to consciousness. Uninterested readers can skip this appendix. Since the publication of the paper on which this appendix is based, much more work within this lively debate has appeared (Block, 2013b and 2014a, Richards, 2013 and Tye, 2014). For reasons of space, I will not be able to give a thorough analysis of all of the complex issues that have been raised since the paper first appeared. Therefore, I reproduce the paper here with only minimal changes.

§A.0-Summary.

In many theories in contemporary philosophy of mind, attention is constitutively linked to phenomenal consciousness (e.g. Prinz, 2012). Ned Block (2013a) has recently argued that 'identity crowding' provides an example of subjects consciously seeing something to which they are unable to attend. Here I examine the reasons that Block gives for thinking that this is a case of a consciously perceived item that we are unable to attend to, and I offer a different interpretation.

§A.1-Attention and Crowding.

Block argues that in (at least some) cases of 'perceptual crowding', we have good reason to think that we can consciously perceive an item, even though we are unable to direct our attention toward the item in question. Block marshals several considerations to make this case (which will be examined below) but his main argument comes from the following example of a kind of perceptual crowding that he calls 'identity crowding' (2013a, pp.173-176):

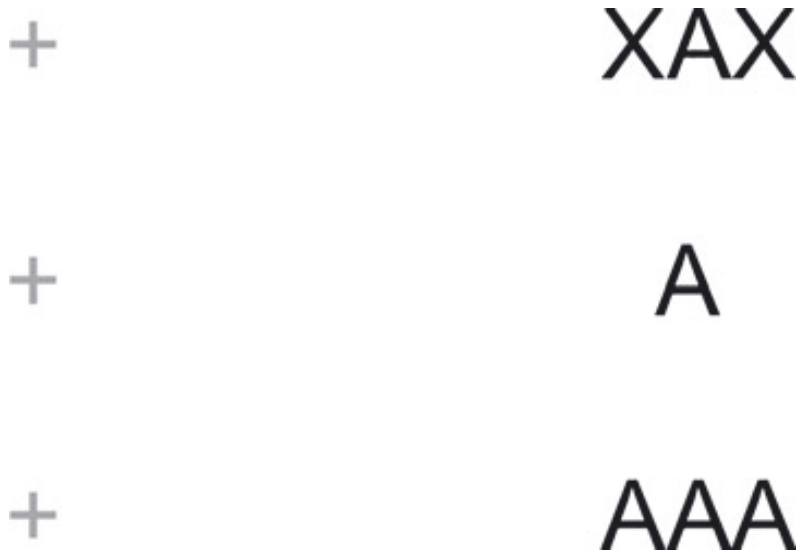

Figure 6: An example of 'identity crowding'. Fixate your gaze on each cross in turn and attempt to identify the letter(s) to the right of it. From Block (2013a, p.174): John Wiley and Sons, Inc. Reproduced with permission.

Fixate your gaze on the top cross in figure 6. You will find that you are unable to direct your attention toward the middle letter in the set of letters (the 'XAX') on the right of the top cross. It does, however, seem that when we focus our gaze upon the cross in the *middle* row, we can focus our attention upon the 'A' to the right of that cross. The crucial case for Block's argument comes from the bottom row. When we focus our gaze upon the cross at the bottom, we are unable to direct our attention towards the *middle* 'A' in the collection of A's on the right of that cross. It does seem plausible, however, that we can direct our attention toward at least one of the other two 'A's on the bottom row (the 'flankers'). Nonetheless, we are capable of realising *that* there is a third item in addition to the flankers on the bottom row, and we are also capable of identifying the middle item in the bottom row as an 'A'. Block claims that the middle 'A' of the bottom row is consciously seen, but we are unable to direct our attention to it. Block summarises this conclusion by saying that the 'grain' of vision is finer than the 'grain' of attention (2013a, p.176).

Importantly, Block claims that 'there can be conscious perception of an object without attention to that object' (2013a, p.170), though he does accept that 'there may still be diffuse spatial attention to the area' (2013a, p.173). Thus, Block's claim is not that consciousness is possible in the *entire absence* of attention, but rather that there can be representations of some object in phenomenal consciousness in the absence of attention *to that particular object.*

§A.2-An analysis of Block's argument.

I accept that the middle 'A' on the bottom row of figure 6 does not have object attention directed upon it. I also accept that it is seen. What I dispute is Block's main claim, that it is seen *consciously.*

I interpret Block's claim in the following way:

> Core Claim (CC): In figure 6, when we are focussing our gaze upon the plus at the bottom of the diagram, we can have a phenomenally conscious representation of the middle 'A' in the bottom row, even though we cannot direct our attention towards that 'A'.

CC contrasts with the claim that really we do not have a phenomenally conscious representation of the middle 'A' in the bottom row, but instead we have a representation of a 'texture' or 'indeterminate shape'. If this alternative interpretation were true, then the fact that subjects cannot attend to the middle 'A' of the bottom row should not concern us, because the middle 'A' is not represented in consciousness. We can call this alternative explanation the 'texture interpretation' and it is this that Block is keen to deny (2013a, pp.171-172).

The central argument that Block gives for accepting that subjects had a phenomenal representation of the middle 'A' (and thus that CC is correct) is that subjects had access to various pieces of information about the middle 'A'. As Block says:

> 'Since identity-crowding allows detection (i.e. distinguishing between presence and absence), differentiation from the background, discrimination from other items and visual identification of the items-all consciously-it is difficult to see a rationale for denying that one can consciously see them' (2013a, p.175).

In order to deny Block's claim that the middle 'A' is consciously seen, we will need an explanation of how the subjects may have access to this information about the middle 'A' without committing us to the claim that this middle 'A' is represented consciously. I shall give such an explanation, then I shall defend it from some of Block's other claims in support of CC.

Let us start with the claim that though we may not be able to direct our attention to the middle 'A' as such, we are still able to direct our attention toward the *group of items as a whole*. When we direct attention towards the group itself, we will come to have some kind of phenomenal representation of the group of items to the right of the cross. It may be that this phenomenal representation is determinate and detailed enough to represent (at least one of) the *flankers* specifically as an 'A'. However, the texture may be too indeterminate to represent the middle letter specifically as an 'A'. This would fit neatly with the texture interpretation and be an alternative to CC.

But if the representation of the group lacks the detail to represent the middle 'A' specifically, then how can we explain Block's main point, which is subjects' abilities to

discriminate and identify the middle item as an 'A'? I claim that when we focus our gaze at the bottom cross, and move our attention toward the bottom group of letters, we may have a phenomenal representation, which is too indeterminate to represent the middle 'A' but which may represent at least one of the *flanking* letters specifically as an 'A' (this will not worry those who disagree with CC, because it seems plausible that at least one of the flankers can be subject to attention). Then the experience of the group may make the subjects able to judge that the group is cluttered (that is, that it contains more than just the flankers). This would allow us to conclude that there was something there in addition to the two flanking letters. We could then notice that the overall representation of the items has a certain congruity, or uniformity. From these pieces of information, we will be able to infer that there is another item in addition to the flankers, and that the middle item is the same as the flankers, i.e. that it is an 'A'. What this response keys into is that the subjects may represent abstract properties of the group of items in their phenomenal consciousness, and that subjects may be using this information to make the judgements that they do, they need not represent the middle 'A' individually.

So the claim is that subjects can infer from:

(1) (At least one of) the flanking letters is an 'A' (the representation of the group of letters may be specific enough to allow for this identification).

And:

(2) The experience represents the group as 'cluttered' (indicating that there are more items there than just the flankers).

And:

(3) The overall representation of the group of items shows congruity and uniformity (indicating that the middle item is probably the same as the flankers).

To the claim that:

(4) There is a middle item which is probably also an 'A'.

What is crucial for this interpretation is that nowhere do we need to claim that there is a specific phenomenal representation of the middle 'A', and yet we are able to explain subjects' capacities to report that there is a middle letter, and to identify it. So Block's claim, that subjects' abilities when confronted with the task shows that subjects had a phenomenal representation of the middle 'A', will not hold up, and CC will not have been established. All we need to claim is present in consciousness is a representation of the group which is detailed enough to identify at least one of the flankers, and also appears cluttered and congruous. This will explain subjects' abilities.

I do not think that this appeal to things such as 'congruity' should strike us as odd. We have good independent reason to think that the visual system is capable of detecting whether a collection of items is 'congruous' with each other or not, and for spotting anomalous items. We know this from the phenomenon of 'visual pop-out' (see e.g. Wolfe and Horowitz, 2004) where certain items which are significantly different from other items that they are placed among (such as a slanted line in a collection of straight lines) will 'pop-out' at us, and they will attract our attention. The reason that they pop-out is because the visual system can discriminate that they do not belong with the other items, they are anomalous or incongruous. All of this would seem to indicate *at the very least* that the visual system is capable of judging whether a collection of items is 'congruous' or 'uniform' and thus this alternative interpretation of the data should be considered.

The plausibility of my claim should be tested using introspection. Focus upon the bottom cross and attend to the collection of three 'A's. Now do the same but with the top set (the 'XAX') group. It strikes me as plausible that when we reflect upon the phenomenology of these contrasting experiences, the top one will strike us as 'more messy' and more incongruous than the bottom one, even though it is difficult to say exactly why. Equally, the bottom group of 'A's is likely to strike us as 'more cluttered' than the middle row, where there is only one 'A'.

Notice that this account is compatible with some other abilities that subjects have in relation to the middle 'A' of the bottom row. For example, Block notes that subjects can have *de re* thoughts about it (they can ask of it 'what is that?'). This Block takes as further evidence that the middle 'A' was represented in phenomenal consciousness (Block, 2013a, p.177. See also Dretske, 2007, Siegel, 2006 and Tye, 2009a, p.59). The interpretation just given can accommodate this. It could be that the group of letters is represented as cluttered, and that we can use this information to reach the conclusion that there is another item there in addition to the flankers, and *this* would allow us to wonder 'what is that?' about this extra item, and thus it could be the basis of *de re* thought. Nowhere need we claim that the middle 'A' is actually represented in phenomenal consciousness.

§A.3-Objections and replies.

In this section I shall discuss some other considerations that Block brings forth in favour of CC.

§A.3.1-Objection 1.

Block discusses and rejects the idea that subjects may be using inference to identify the middle 'A'. He references an experiment where subjects had triplets of Gabor patches

(grids) each of which were slanted in a specific orientation presented in the periphery of their visual field, and then were asked to identify which triplet they had just been exposed to. Block notes that performance was different when they had been exposed to the triplet of patches slanted to the orientation '///' as opposed to the triplet '/\/' (see Petrov and Popple, 2007, pp.3-5). He concludes that:

> '[t]he… experiment suggests that subjects' success in identity crowding is genuinely perceptual and not just a cognitive inference from the look of uniformity and identification of the flankers. For there are considerable asymmetries between left and right tilts that only can be explained perceptually' (2013a, p.175).

In response to this, I should stress that my claim is not that subjects *just* used information about the identity and uniformity of the *flankers* to make their judgement, but that they used this information about the flankers *together with the overall look of clutter, congruity and uniformity of the experience of the group itself.* This experience would afford them information that would not have been given simply by identifying the flankers and then drawing an inference from that. So Block's claim that subjects were not *only* using the identity of the flankers to draw their conclusions does not serve to rebut this interpretation.

A point that is made by Tye (2010, p.416) is worth reemphasising now, which is that something can *contribute* to the phenomenology of an experience, without actually being *represented* in the experience. This is plausibly the case with middle '/' in the Petrov and Popple experiment and the middle 'A' in the bottom row of figure 6. These items may contribute to the phenomenology of the experience of the group (by making the group appear more congruous/incongruous or more cluttered/uncluttered). Indeed, it is possible that the contribution made by these items is different from the contribution that other items

would have made if they had been present. Nonetheless, it does not follow that these items are themselves represented in consciousness.

In the above quotation, Block claims that the 'asymmetries' in discriminatory abilities between identifying a '///' and identifying a '/\/' in the Petrov and Popple experiments 'can only be explained perceptually' (Block, 2013a, p.175). In response to this, I accept that in this experiment, *all* of the Gabor patches were perceived and processed to a high degree of detail in the visual system, but this need not imply that a representation of each one of them was phenomenally conscious. This point is important with relation to the controversial middle 'A' in figure 6. It may be said that in order for subjects to be able to make judgements about how congruous or cluttered the group of 'A's is, information about the middle 'A' must be processed to a high level of detail in the visual system. Again, this is true, but the issue here is not whether the visual system processes information about the middle 'A' but whether the 'A' is actually represented in phenomenal consciousness. We know that a great deal of visual information is processed unconsciously (e.g. Milner and Goodale, 1995 and Milner, 2012), so the fact that the crowded items in figure 6 and in the Petrov and Popple experiments are *perceived* does not by itself tell us anything about whether they are phenomenally conscious. I shall have more to say about unconscious perception below.

§A.3.2-Objection 2.

Block claims that if we are to accept that the textures that are phenomenally conscious to the subjects in the experiments have a high level of detail, then in the end there will be no difference between seeing the texture and seeing the objects that comprise it. Block says this:

> '[m]aybe a texture can be letterish but can it be A-ish or Times-Roman-A-ish? If one allows such detailed textures, it is not clear that there is any incompatibility between seeing textures and seeing objects that compose them. One can see an object and at the same time see it as fitting into a pattern that includes other objects' (2013a, p.176).

I say two things in response to this. Firstly, the texture that I have suggested need not represent the group of letters in this much detail in order to explain subjects' abilities. The texture in question need not represent the group as a collection of Times-Roman-A's. Rather, all that the texture required would have to represent is that the group contains *at least one* 'Times-Roman-A' (which will be one of the flankers, to which we *can* direct attention) and that the group is also cluttered and congruous, the middle letter does not itself need to be represented as a 'Times-Roman-A'. From this level of detail, subjects' inferences can do the rest. So there is an important distinction between the level of detail of the texture that I have suggested and the level of detail of the texture that Block is here criticising.

Secondly, there is good reason to think that Block's claim here begs the question. It is one of the main contentions of the texture interpretation that we can see a group of items collectively along with some of their features without seeing all of the individual items that make up the group. This is familiar from discussion of the speckled hen, where it is claimed that we can see a speckled hen without seeing each of the speckles that makes it up (e.g. Tye, 2009a). This kind of experience seems ubiquitous, often we will see a brick wall from a distance, and judge that it is a brick wall, and indeed our experience makes us able to tell a great deal about each individual brick that makes up the wall (that it is red, of a certain shape etc.) but it is at the very least not clear that we have a phenomenally conscious representation of each brick individually. So to insist that we cannot represent something as

a collection of 'Times-Roman-A's' without representing each individual A seems to simply assume one of the core claims that Block's opponents would deny.

§A.3.3-Objection 3.

Block refers to another example of crowding to support his case (figure 7).

Figure 7: From Intriligator and Cavangh (2001): Wiley and Sons, Inc. Reproduced with permission.

When one fixates upon the cross in figure 7, one will be able to attend to some of the bars, though not each one of them individually. Block claims (contra Tye, 2010) that we see each individual bar, even those that we are unable to attend to, and notes as evidence for this the fact that the bars are differentiated in experience from the background, and that '[o]ne can see the white space in between the items' (2013a, p.177). This Block takes as evidence that we consciously see each individual bar, he thinks this is 'obvious' (2013a, p.177).

An alternative interpretation is that we see a collection of 'black bars and white spaces' without seeing each individual bar and each individual white space. This would explain how subjects know that there are white spaces there. It may well also be that subjects can differentiate the left-most bar from the background and the right-most bar from the background (because they can attend to these bars) and then infer, again based on properties of the group such as congruity and uniformity, that *all* of the bars are differentiated from the background. This would explain how subjects are capable of differentiation of the bars from

the background, and are aware of the white spaces, without committing us to the claim that each bar and each space is individually consciously seen.

§A.4-Unconscious Seeing.

Block discusses more empirical data concerned with object seeing and attention. He references an experiment by Freeman and Pelli (2007) where subjects were exposed to a collection of letters (sometimes crowded, sometimes not), and then after the letters disappeared from the screen, the subjects were cued to recall one letter. It was found that subjects' abilities for recall of crowded letters and uncrowded letters were the same, but that subjects' abilities significantly decreased when exposed to Armenian letters rather than Roman ones (see Freeman and Pelli, 2007, p.8). This Block takes as evidence that seeing the crowded items depends upon letter-recognition (2013a, p.180).

Block claims that before the cue, the crowded letters were each individually seen, even though attention was not directed upon them. Part of Block's case here is that each letter was not perceived as a 'bag' of unorganised features but specifically as 'letter-representation' (2013a, p.180), and also that there is no positive reason to think that the crowded letters were each subject to attention (2013a, p.182). Block does, however, accept that the crowded letters were likely seen *unconsciously.* In order to maintain this claim, Block argues that 'seeing' is a natural kind that has conscious and unconscious subkinds (2013a, pp.180-181). Block concludes thus:

> '*[s]ince unconscious seeing is still seeing, seeing an individual item is compatible with failure of object-based attention to it.* So the Freeman and Pelli experiment shows that there *can*

be seeing an object-even if unconscious seeing-without attention to that object' (2013a, p.180).

Block says: 'I have not argued against the possibility that what is in consciousness in Freeman and Pelli is merely textural until the cue' but that 'object seeing (if only unconscious object-seeing) is compatible with the lack of object-attention engendered by crowding' (2013a, pp.181-182). Block also notes that 'the Freeman and Pelli experiment suggests that crowded object seeing does not have to be merely textural' (2013a, p.182).

This argument would be a problem for someone who claimed that all (conscious or unconscious) perception of crowded objects *had* to be merely textural, or that all (conscious or unconscious) object perception was impossible in the absence of attention, but I see no reason for an opponent of CC to hold either of these claims. Block's opponents can accept the main claim that *unconscious* crowded object seeing may not be textural, and may occur in the absence of attention but still claim that *conscious* object seeing cannot occur in the absence of attention, and that *conscious* perception of crowded items is textural (or, at least, that it is textural when the crowded objects escape attention).

Perhaps we might extend the remit of Block's argument and claim that since unconscious seeing of crowded objects is compatible with an absence of attention we have good reason to think that conscious seeing of crowded objects is as well. However, this would be an illicit leap, as many of Block's opponents (e.g. Prinz, 2012) would claim that attention is precisely what makes the difference between something's being unconscious and its being conscious, so to infer from the fact that something unattended to can be perceived unconsciously to the claim that it can be unattended to and perceived consciously would be to assume that attention is not what makes the difference between something's being

unconscious and its being conscious, which is precisely what Block's opponents would deny. It is certainly true that much unconscious perception occurs in the absence of attention, but if what we are interested in is the question of whether there is a link between attention and consciousness, I do not think this should worry us.

§A.5-Conclusion.

The contrary interpretation of the data that I have suggested allows us to keep the relationship between object attention and conscious object seeing very tight, and thus in the absence of reason against it, we need not commit to CC. For this reason, Block's argument is unsuccessful.

Bibliography

Allport, A. 1993. "Attention and Control. Have we been asking the wrong questions? A critical review of twenty-five years." In Meyer, D. E. and Kornblum, S. (eds.) *Attention and performance, XIV.* (Cambridge, MA: MIT Press). pp.183-218.

Allport, A. 2011. "Attention and integration." In Mole, C., Wu, W. and Smithies, D. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch.2, pp.24-59.

Allport, D. A., Antonis, B. and Reynolds, P. 1972. "On the division of attention: a disproof of the single channel hypothesis." *Quarterly Journal of Experimental Psychology*, 24 (1), pp.225-35.

Anderson, B. 2011. "There is no such thing as attention." *Frontiers in Psychology.* 2, pp.1-8.

Armstrong, D. 2004. *Truth and Truthmakers.* (Cambridge: Cambridge University Press).

Armstrong, D. 2005. "Four disputes about properties." *Synthese.* 144, pp.309-320.

Armstrong, K. 2011. "Covert spatial attention and saccade planning." In Mole, C., Wu, W. and Smithies, D. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch.4, pp.78-96.

Aru, J. and Bachmann, T. 2013. "Phenomenal awareness can emerge without attention." *Frontiers in Human Neuroscience.* 7 (Article 891), pp.1-2.

Awh, E., Vogel, E. and Oh, S. 2006. "Interactions between attention and working memory." *Neuroscience* . 139, 201-208.

Awh, E., Belopolsky, A.V., & Theeuwes, J. 2012. "Top-down versus bottom-up attentional control: A failed theoretical dichotomy." *Trends in Cognitive Sciences,* 16(8), pp.437-443

Azzapardi, P. and Cowey, A. 1997. "Is blindsight like normal, near-threshold vision?" *PNAS* 94, pp.14190-14194.

Baars, B. 1995. "Evidence that phenomenal consciousness is the same as access consciousness." *Behavioral and Brain Sciences.* 18 (2), p.249.

Baars, B. 1997. *In The Theater of Consciousness: the workspace of the mind.* (New York: Oxford University Press).

Baars, B. 2002. "The conscious access hypothesis: origins and recent evidence." *Trends in Cognitive Sciences.* 6, pp.47-52.

Baars, B. 2003. "How conscious experience and working memory interact." *Trends in Cognitive Sciences.* 7, pp.166-172.

Baars, B. 2005. "Global workspace theory of consciousness: toward a cognitive neuroscience of human experience?" *Progress in Brain Research.* 150, pp.45-53.

Baars, B. 2007. "The global workspace theory of consciousness." In Velmans, M. and Schneider, S. (eds.) *The Blackwell Companion to Consciousness.* (Singapore: Blackwell). Ch. 18, pp.236-246.

Bachmann, T. 2011. "Attention as a process of selection, perception as a process of representation, and phenomenal experience as the resulting process of perception being modulated by a dedicated consciousness mechanism." *Frontiers in Psychology.* 2, p.287.

Balcerak Jackson, B. 2014. "Verbal disputes and substantiveness." *Erkenntnis.* 79, pp.31-54.

Beall, J. C. and Restall, G. 2000. "Logical Pluralism." *Australasian Journal of Philosophy.* 78 (4), pp.475-493.

Bechtel, W. 2008. *Mental Mechanisms.* (Sussex: Psychology Press).

Bechtel, W. and Mundale, J. 1999. "Multiple realizability revisited: linking cognitive and neural states." *Philosophy of Science,* 66 (2), pp.175-207.

Bird, Alexander and Tobin, Emma, "Natural Kinds", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/natural-kinds/>.

Block, N. 1995a. "On a confusion about a function of consciousness." Reprinted in Block, N., Gulzeldere, G. and Flanagan, O. (eds.) *The Nature of Consciousness: Philosophical Debates.* 1997. Reprinted 2002. (USA: MIT Press). Ch.20. pp.375-415. Citations from this version.

Block, N. 1995b. "How many concepts of consciousness?" Reprinted in *Consciousness, Function and Representation: Collected Papers, Volume 1.* 2007. (USA: MIT Press). Ch.10, pp.215-247.

Block, N. 1996. "How not to find the neural correlates of consciousness." Reprinted in *Consciousness, Function and Representation: Collected Papers, Volume 1.*2007. (USA: MIT Press). Ch.14, pp.299-309.

Block, N. 1997. "Biology versus computation in the study of consciousness." Reprinted in *Consciousness, Function and Representation: Collected Papers, Volume 1.*2007. (USA: MIT Press). Ch.11, pp.249-268.

Block, N. 2001. "Paradox and cross purposes in recent work on consciousness." Reprinted in *Consciousness, Function and Representation: Collected Papers, Volume 1.*2007. (USA: MIT Press). Ch.15, pp.311-338.

Block, N. 2002. "Concepts of consciousness." Reprinted in *Consciousness, Function and Representation: Collected Papers, Volume 1.*2007. (USA: MIT Press). Ch.13, pp.275-296.

Block, N. 2003. "Mental paint." Reprinted in *Consciousness, Function and Representation: Collected Papers, Volume 1.*2007. (USA: MIT Press). Ch.24, pp.533-570.

Block, N. 2005. "Two neural correlates of consciousness." Reprinted in *Consciousness, Function and Representation: Collected Papers, Volume 1.*2007. (USA: MIT Press). Ch.17, pp.343-361.

Block, N. 2007a. "Consciousness, accessibility and the mesh between psychology and neuroscience." *Behavioural and Brain Sciences.* 30 (5/6), pp.481-499.

Block, N. 2007b. "Overflow, access, and attention." *Behavioural and Brain Sciences.* 30 (5/6), pp.530-548.

Block, N. 2008. "Consciousness and cognitive access." *Proceedings of the Aristotelian Society.* 108 (3), pp.289-317.

Block, N. 2009. "Comparing the major theories of consciousness." In Gazzaniga, M. (ed.) *The Cognitive Neurosciences: Fourth Edition.* (Hong Kong: MIT Press). Ch.77, pp.1111-1122

Block, N. 2010. "Attention and mental paint." *Philosophical Issues.* 20 (1), pp.23-63.

Block, N. 2011. "Perceptual consciousness overflows cognitive access." *Trends In Cognitive Sciences.* 15 (12), pp.567-575.

Block, N. 2012. "Response to Kouider et al.: which view is better supported by the evidence?" *Trends in Cognitive Sciences.* 16 (3), pp.141-142.

Block, N. 2013a. "The grain of vision and the grain of attention." *Thought: A Journal of Philosophy.* 1 (3), pp.170-184.

Block, N. 2013b. "Seeing and windows of integration." *Thought: A Journal of Philosophy.* 2 (1), pp.29-39.

Block, N. 2014a. "The defective armchair: a reply to Tye" *Thought: A Journal of Philosophy.* 3 (2), pp.159-165.

Block, N. 2014b. "Rich conscious perception outside focal attention" *Trends in Cognitive Sciences.* In press. Available at: http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/2014.corrected.proof.pdf. Accessed 04/08/2014.

Block, N. and Fodor, J. 1972. "What psychological states are not." *The Philosophical Review.* 81 (2), pp.159-181.

Block, N. and Stalnaker, R. 1999. "Conceptual analysis, dualism and the explanatory gap.' *The Philosophical Review.* 108, pp.1-46.

Bommarito, N. 2013. "Modesty as a virtue of attention." *The Philosophical Review.* 122, (1), pp.93-117.

Boyd, R. 1989. "What realism implies and what it does not." *Dialectica.* 43 (1-2), pp.5-29.

Boyd, R. 1991. "Realism, anti-foundationalism and the enthusiasm for natural kinds" *Philosophical Studies.* 61 (1/2). pp. 127-148.

Boyd, R. 1999. "Kinds, complexity and multiple realization: comments of Millikan's "Historical Kinds and the Special Sciences"". *Philosophical Studies.* 95 (1/2), pp.67-98.

Boyd, R. 2003. "Finite beings, finite goods: the semantics, metaphysics and ethics of naturalist consequentialism, part I" *Philosophy and Phenomenological Research,* 66, pp.505-553.

Boyd, R. 2013. "What of pragmatism with the world here?" In Baghramian, M. (ed.) *Reading Putnam.* (USA: Routledge).

Breitmeyer, B. G., Ro, T. and Singhal, N. S. 2004. "Unconscious color priming occurs at stimulus-not percept-dependent levels of processing." *Psychological Science.* 15, pp.198-202.

Brigandt, I. 2003. "Species Pluralism does not imply species eliminativism." *Philosophy of Science.* 70 1305-1316.

Broadbent, D. E. 1958. *Perception and Communication*. (Oxford: Pergamon Press).

Broadbent, D. E. 1982. "Task combination and selective intake of information." *Acta Psychologica.* 50, pp.253-290.

Brown, R. 2012. "The myth of phenomenological overflow." *Consciousness and Cognition.* 21, pp.599-604.

Burge, T. 1979. "Individualism and the mental." *Midwest Studies in Philosophy.* 4, pp.73-121.

Burge, T. 1997. "Two kinds of consciousness." In Block, N., Gulzeldere, G. and Flanagan, O. (eds.) *The Nature of Consciousness: Philosophical Debates.* Reprinted 2002. (USA: MIT Press).Ch.24, pp.427-434.

Campbell, J. 2002. *Reference and Consciousness.* (Oxford: Oxford University Press).

Campbell, J. 2004. "Reference as attention." *Philosophical Studies.* 120 (1/3), pp.265-276.

Campbell, J. 2011. "Visual attention and the epistemic role of consciousness." In Mole, C., Smithies, D. and Wu, W. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch.14, pp.323-341.

Carey, S. 2009. *The Origin of Concepts*. (Oxford: Oxford University Press).

Carrasco, M. 2011. Visual attention: the past 25 years." *Vision Research.* 51, pp.1484-1525.

Carrasco, M., Ling, S. and Read, S. (2004). "Attention alters appearance." *Nature.* 7 (3), pp.308-313.

Carruthers, P. 2000. *Phenomenal Consciousness: a naturalistic theory.* (Cambridge: Cambridge University Press).

Chalmers, D. J. 1995. "Facing up to the problem of consciousness." *Journal of Consciousness Studies.* 2 (3), pp.200-219.

Chalmers, D. J. 1996. *The Conscious Mind: in search of a fundamental theory.* (USA: Oxford University Press).

Chalmers, D. J. 1997. "Availability: The Cognitive Basis of Experience?" In Block, N., Gulzeldere, G. and Flanagan, O. (eds.) *The Nature of Consciousness: Philosophical Debates.* Reprinted 2002. (USA: MIT Press). Ch. 22. pp.421-424.

Chalmers, D. J. 2010. *The Character of Consciousness.* (New York: Oxford University Press).

Chalmers, D. J. 2011. "Verbal disputes." *The Philosophical Review.* 120 (4), 515-566.

Charland, L. 1995. "Emotion as a natural kind: Towards a computational foundation for emotion theory." *Philosophical Psychology.* 8 (1): 59-84.

Charland, L. 2002. "The Natural Kind Status of Emotion." *British Journal of Philosophy of Science.* 53: 511-537.

Cherry, C. 1953. "Some experiments on the reception of speech with one and with two ears." *Journal of the Acoustical Society of America.* 25, pp.975-9

Chun, M., Golomb, J. and Turk-Browne, N.B. 2011. "A taxonomy of internal and external attention." *Annual Review of Psychology.* 62, pp.73-101.

Churchland, P. 1981. "Eliminative materialism and the propositional attitudes." *Journal of Philosophy.* 78 (2), pp.67-90.

Clark, Andy. 2000a. "A case where access implies qualia?" *Analysis.* 60(1), pp.30-37.

Clark, Andy. 2000b. "Phenomenal immediacy and the doors of sensation." *Journal of Consciousness Studies.* 7 (4), pp.21-24.

Clark, Austen. 2006. "Attention and inscrutability: a commentary on John Campbell's *Reference and Consciousness.*" *Philosophical Studies.* 127, pp.167-193.

Claridge, M., Dawah, H., and Wilson, R., (eds.), 1997.*Species: The Units of Biodiversity*, (London: Chapman and Hall).

Cohen, M. A. and Dennett, D. C. 2011. "Consciousness cannot be separated from function." *Trends in Cognitive Sciences.* 15 (8). pp.358-364.

Cohen, M.A., Cavanagh, P., Chun, M. and Nakayama, K. 2012a. "The attentional requirements of consciousness." *Trends in Cognitive Sciences.* 16 (8), pp. 411-417.

Cohen, M.A., Cavanagh, P., Chun, M. and Nakayama, K. 2012b. "Response to Tsuchiya *et al.*: considering endogenous and exogenous attention." *Trends in Cognitive Sciences.* 16 (11), p.528.

Corbetta, M. and Shulman, G. 2002. "Control of goal-directed and stimulus driven attention in the brain." *Nature Reviews: Neuroscience.* 3, pp.201-216.

Crane, T. 2001. *Elements of Mind: An Introduction to the Philosophy of Mind.* (New York: Oxford University Press). Reprinted 2009.

Crane, T. 2013. "Unconscious Belief and Conscious Thought." In Kriegel, U. (ed.) *Phenomenal Intentionality: New Essays.* (New York: Oxford University Press).

Craver, C. 2009. "Mechanisms and natural kinds." *Philosophical Psychology.* 22 (5), pp.575-594

Davidson, D. 1969. "The individuation of events." Reprinted in *Essays on Actions and Events.* (New York: Oxford university Press). 1980. Ch.8. pp.163-180.

Davidson, D. 1970. "Events as particulars." Reprinted in *Essays on Actions and Events.* (New York: Oxford university Press). 1980. Ch.9. pp.181-187.

De Brigard, F. 2010. "Consciousness, attention and commonsense." *Journal of Consciousness Studies.* 17 (9-10), pp.189-201.

De Brigard, F. 2012. "The role of attention in conscious recollection." *Frontiers in Psychology.* 3, pp.1-10.

De Brigard, F. 2014. "A commentary on Philip Koralus' 'The erotetic theory of attention: questions, focus and distraction." Symposium for *Mind and Language* on Philip Koralus' theory of attention. Available at: https://dl.dropboxusercontent.com/u/56406837/debrigaard%20comments.pdf. Accessed 11/Sept/2014.

De Brigard, F. and Prinz, J. 2010. "Attention and consciousness." *Wiley Interdiscerplinary Reviews: Cognitive science.* 1 (1), pp.51-59.

De Gardelle, V., Sachur, J. and Kouider, S. 2009. "Perceptual Illusions in brief visual representations." *Consciousness and Cognition.* 18, pp.569-577.

De Sousa, R. 1987. *The Rationality of Emotion.* Cambridge, MA: MIT Press.

Dehaene, S. 2014. *Consciousness and the Brain: decoding how the brain codes our thoughts.* (Warnock: Penguin)

Dehaene, S. and Naccache, L. 2001. "Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework." *Cognition.* 79, pp.1-37.

Dehaene, S., Changeux, J., Naccache, L., Sackur, J. and Sergent, C. 2006. "Conscious, preconscious, and subliminal processing: a testable taxonomy." *Trends in Cognitive Sciences.* Vol.10. pp.204-211.

Dennett, D. C. 1991. *Consciousness Explained.* Reprinted 1993. (St. Ives: Penguin).

Dennett, D. C. 1995. "The Path not Taken." Reprinted in Block, N., Gulzeldere, G. and Flanagan, O. (eds.). *The Nature of Consciousness: Philosophical Debates.* 1997. Reprinted 2002. (USA: MIT Press). Ch.21. pp.417-419. Citations from this version.

Dennett, D. C. 1996. "Facing backwards on the problem of consciousness." *Journal of Consciousness Studies.* 3 (1), pp.4-6.

Dennett, D. C. 2001a. "Surprise, surprise." *Behavioral and Brain Sciences.* 24 (5), p.982.

Dennett, D. C. 2001b. "Are we explaining consciousness yet?" *Cognition.* 79, pp.221-237.

Dennett, D. C. 2002. "How could I be wrong? How wrong could I be?" *Journal of Consciousness Studies.* 9, pp.13-16.

Dennett, D. C. 2006. *Sweet Dreams: Scientific Obstacles to a Science of Consciousness.* (USA: MIT Press).

Descartes. R. 1988. "Replies to Objections", in *The Philosophical Writings Of Descartes* (3 volumes), translated by J. Cottingham, R. Stoothoff, D. Murdoch, and A. Kenny, Cambridge: Cambridge University Press.

Devitt, M. 2008. "Resurrecting Biological Essentialism." *Philosophy of Science.* 75 (3), pp.344-382.

Dretske, F. 1969. *Seeing and Knowing.* (London: Routledge).

Dretske, F. 1995. *Naturalizing the Mind.* (USA: MIT Press).

Dretske, F. 2006. "Perception without awareness." In Gendler, T. S. and Hawthorne, J. (eds.)*Perceptual Experience.* (New York: Oxford University Press). Ch. 4. pp.147-180.

Dretske, F. 2007. "What change blindness teaches about consciousness." *Philosophical Topics.* 21, pp.215-230.

Duncan, J. 2006. "Brain mechanisms of attention." *The Quarterly Journal of Experimental Psychology.* 59, pp.2-27.

Dupré, J. 1981. "Natural kinds and biological taxa." *The Philosophical Review.* 90, pp.66-90.

Dupré, J. 1993. *The Disorder of Things.* (Cambridge, MA: Harvard University Press).

Dupré, J. 1999. "Are whales fish?" In D. L. Medin and S. Atran (eds.) *Folkbiology.* (Cambridge, MA: MIT Press).

Egly, R., Driver, J. and Rafal, R. D. 1994. "Sifting visual attention between objects and locations: evidence from normal and parietal lesion subjects." *Journal of Experimental Psychology.* 123 (2), pp.161-177.

Eilan, N., Hoerl, C., McCormack, T. & Roessler, J., 2005, *Joint Attention: Communication and Other Minds*, Oxford: Oxford University Press

Ellis, B. 2001. *Scientific Essentialism.* (USA: Cambridge University Press).

Ellis, B. 2005. "Scientific Realism." *Ratio.* 18 (4), pp.371-384.

Ereshefsky, M. 1992a. "Eliminative Pluralism." *Philosophy of Science.* 59 (4), pp.671-690.

Ereshefsky, M. 1992b. *The units of evolution: essays on the nature of species.* (Cambridge, MA.: MIT Press).

Ereshefsky, M. 1998. "Species pluralism and anti-realism." *Philosophy of Science.* 65 (1), pp.103-120.

Ereshefsky, M. 2010a. "What's wrong with the new biological essentialism." *Philosophy of Science.* 77 (5), pp.674-685.

Ereshefsky, Marc, 2010b "Species", *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2010/entries/species/>.

Ereshefsky, M. 2010c. "Darwin's solution to the species problem." *Synthese.* 175, pp.405-425.

Ereshefsky, M. 2010d. "Microbiology and the species problem." *Biology and Philosophy.* 25. pp.553-568

Ereshefsky, M. and Reydon, T. Forthcoming. "Scientific kinds." *Philosophical Studies.*

Fang, F. and He, S. 2005. "Cortical responses to invisible objects in the human dorsal and ventral pathways." *Nature Neuroscience.* 8 (10), pp.1380-1385.

Fodor, J. 1974. "Special sciences: Or, the disunity of science as a working hypothesis." *Synthese.* 28, pp.97-115.

Fodor, J. 1981. "The present status of the innateness controversy." In *representations: Philosophical essays on the foundations of cognitive science.* (Cambridge, MA: MIT Press). pp.257-316.

Fodor, J. 1997. "Special sciences: still autonomous after all these years." *Philosophical Perspectives.* 11, pp.149-63.

Fodor, J. 1998. *Concepts: where cognitive science went wrong.* (New York: Oxford University Press).

Fodor, J. 2008. *LOT 2: the language of thought revisited*. (New York: Oxford University Press).

Fougnie, D. and Marois, R. 2007. "Executive load in working memory induces inattentional blindness." *Psychonomic Bulletin and Review.* 14, pp.142-147.

Freeman, J. and D. Pelli. 2007. "An Escape from Crowding." *Journal of Vision.* 7.2, pp.1-14.

Gennaro, R. 2014. "Review of *Attention: Philosophical and Psychological Essays.* Mole, C., Smithies, D. and Wu, W. (eds.)" *Mind* 123 (490), pp.623-628.

Gettier, E.L. 1963. "Is knowledge justified true belief?" *Analysis.* 23 (6), pp.121-123.

Ghiselin, M. 1987, "Species concepts, individuality, and objectivity" *Biology and Philosophy*, 2: 127–143

Ghiselin, M. 1989. "Sex and the individuality of species: a reply to Mishler and Brandon." *Biology and Philosophy.* 2, pp.127-143.

Gibb, S. C. 2012. "Trope simplicity." In Garcia, R. (ed.) *Substance: new essays.* (Philosophia Verlag).

Gobell, J. and Carrasco, M. 2005. "Attention alters the appearance of spatial frequency and gap size." *Psychological Science.* 16 (8), pp.644-651.

Gopnik, A. 2007. "Why babies are more conscious than we are." *Behavioural and Brain Sciences.* 30 (5/6), pp.503-504.

Goodman, N. 1978. *Ways of Worldmaking*. (Indianapolis: Hackett).

Griffiths, P. E. 1997. *What Emotions Really Are: the problem of psychological categories.* (Chicago: University of Chicago Press).

Griffiths, P. E. 2002. "What is innateness?" *The Monist.* 85 (1), pp.70-85.

Griffiths, P. E. 2004a. "Emotions as natural and normative kinds." *Philosophy of Science.* 71 (5), pp.901-911.

Griffiths, P. E. 2004b. "Is emotion a natural kind?" In Solomon, R. (ed.) *Thinking about Feeling: Contemporary philosophers on emotion.* (New York: Oxford University Press). Available at: http://philsci-archive.pitt.edu/566/1/Is_Emotion_a_Natural_Kind.PDF. Accessed 07/02/2014. Citations from this version.

Grush, R. 2007. "A Plug for Generic Phenomenology." *Behavioural and Brain Sciences.* 30, pp.504-505.

Haines, R. F. 1991. "A breakdown in simultaneous information processing." In G. Obrecht and L. Stark (eds.) *Presbyopia Research* (171-75). (New York: Plenum).

Heil, J. 2003. *From an Ontological Point of View.* (New York: Oxford University Press). Reprinted 2009.

Heil, J. 2005. "Kinds and Essences." *Ratio.* 18 (4). pp.405-419.

Hein, E., Rolke, B. and Ulrich, R. 2006. "Visual attention and temporal discrimination: differential effects of automatic and voluntary cueing." *Vision Research.* 33 (9), pp.1219-1240.

Hendry, R. 2005. "Lavoisier and Mendeleev on the elements" *Foundations of Chemistry* 7, pp. 31-48.

Hendry, R. Forthcoming. "Natural kinds in chemistry." *The Oxford Companion to the Philosophy of Chemistry.*

Hill, C. 2009. *Consciousness.* (New York: Cambridge University Press).

Huang, L. and Pashler, H. 2007. "A Boolean map theory of visual attention." *Psychological Review.* 114, pp.599-631.

Hull, D. 1965. "Effect of essentialism on taxonomy: two thousand years of stasis." *British Journal for the Philosophy of Science* 15, pp.314-326.

Intriligator, J., and Cavangh, P. 2001. "The spatial resolution of visual attention." *Cognitive Psychology* 43, pp.171-216.

Irvine, E. 2013. *Consciousness as a Scientific Concept: a philosophy of science perspective.* (Dordrecht: Springer)

Jackendoff, R. 1992. *Languages of the Mind.* (Cambridge, MA: MIT press).

James, W. 1890/1981. *The Principles of Psychology.* (Cambridge: Cambridge University Press). Reprinted 1981.

Jennings, C. D. 2012. "The subject of attention." *Synthese.* 189, pp.535-554.

Jiang, Y., Costello, P., Fang, F., Huang, M., He, S., & Purves, D., 2006, "A gender- and sexual orientation-dependent spatial attentional effect of invisible images", *Proceedings of the National Academy of Sciences* (USA), 1003(45): 17048–17052.

Kandel, E. 2002. "The molecular biology of memory storage: a dialog between genes and synapses." *Bioscience Reports.* 21 (5), pp.565-611.

Kelly, S. D. 2004. "Reference and attention: a difficult connection." *Philosophical Studies.* 120 (1/3), pp.277-286.

Kentridge, R. 2011. "Attention without awareness: a brief review." In Mole, C., Wu, W. and Smithies, D. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch. 10, pp.228-246.

Kentridge, R., Heywood, C. and Weiskrantz, L. 1997. "Residual vision in multiple retinal locations: Implications for blindsight." *Journal of Cognitive Neuroscience.* 9, pp.191-202.

Kentridge, R., Heywood, C.A. and Weiskrantz, L. 1999. "Attention without awareness in blindsight." *Proceedings of the Royal Society (London) Series B.*Vol.266, pp.1805-1811.

Kentridge, R., Heywood, C. A. and Weiskrantz, L. (2004). Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia.* 42, 831-5.

Kentridge, R., de-Wit, L.H. and Heywood, C.A. (2008a) What is attended in spatial attention? *Journal of Consciousness Studies.* 15 (4) pp.105-111.

Kentridge, R., Nijober, T. C. W. and Heywood, C. A. (2008b) Attended but unseen: Visual attention is not sufficient for visual awareness. *Neuropsychologia.* 46 (3) pp.831-69.

Kirk, Robert. 2005. *Zombies and Consciousness.* (New York: Oxford University Press).

Kitcher, P., 1984, "Species", *Philosophy of Science*, 51: 308–333.

Knobe, J. 2012. "Experimental Philosophy." In Margolis, E., Samuels, R. and Stich, S. (eds.) *The Oxford Handbook of Philosophy of Cognitive Science.* (New York: Oxford University Press). Ch. 22.

Koch, C. and Tsuchiya, N. 2007. "Consciousness and attention: two distinct brain processes." *Trends in Cognitive Sciences.* 11 (1), pp.16-22.

Koralus, P. 2014a. "Attention, consciousness and the semantics of questions." *Synthese.* 191, pp.187-211.

Koralus, P. 2014b. "The erotetic theory of attention: questions, focus, and distraction." *Mind and Language.* 29 (1), pp.26-50.

Kouider, S., Dehaene, S., Jobert, A. and Le Bihan, D. 2007a. "Cerebral Bases of Subliminal and Supraliminal Priming During Reading." *Cerebral Cortex.* Vol. 17, No.9. pp.2019-2029.

Kouider, S., de Gardelle, V. and Dupoux, E. 2007b. "Partial Awareness and the Illusion of Phenomenal Consciousness. " *Behavioral and Brain Sciences*, 30, pp.510-511.

Kouider, S., de Gardelle, V., Sachur, J. and Dupoux, E. 2010. "How Rich is Consciousness? The Partial Awareness Hypothesis." *Trends in Cognitive Sciences.* 14 (7), pp.301-307.

Kriegel, U. 2009. *Subjective Consciousness: A Self Representational Theory.* (New York: Oxford University Press).

Kripke, S. 1980. *Naming and Necessity.* (Tyne and Wear: Blackwell). Reprinted 2007.

Kripke, S. 2011. *Philosophical Troubles.* (Oxford: Oxford University Press).

Ladyman, J. and Ross, D., with Spurrett, D. and Collier, J., 2007. *Every Thing Must Go: Metaphysics Naturalised*. (Oxford: Oxford University Press).

Lamme, V. 2003. "Why visual attention and awareness are different." *Trends in Cognitive Sciences.* 7(1), pp.12-18.

Lamme, V. 2004. "Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness." *Neural Networks* 17, pp. 861-872.

Lamme, V. 2007. "Sue Ned Block! Making a better case for P-consciousness." *Behavioral and Brain Sciences.* 30 (5-6), pp.511-512.

Lamme, V. 2010a. "How neuroscience will change our view on consciousness." *Cognitive Neuroscience.* 1, pp.204-220.

Lamme, V. 2010b. "Reply to commentaries: What introspection has to offer, and where its limits lie." *Cognitive Neuroscience.* 1, pp.232-240.

Landman, R., Spekreijse, H and Lamme, V. A. F. 2003. "Large capacity storage of integrated objects before change blindness." *Vision Research.* 43 (2), pp.149-164.

Laurence, S. & Margolis, E. 1999. "Concepts and Cognitive Science." in E. Margolis & S. Laurence (eds.), *Concepts: Core Readings*. (Cambridge, MA: MIT Press). Ch.1, pp. 3–81.

Lavie, N. 2005. "Distracted and confused?: Selective attention under load." *Trends in Cognitive Science.* 9 (2), pp.75-82.

Lavie, N. 2007. "Attention and consciousness." In Velmans, M. and Schneider, S. (eds.) *The Blackwell Companion to Consciousness.* (Singapore: Blackwell). Ch.38, pp.489-503

Lavie, N. and Yehoshua, T. 1994. "Perceptual load as a major determinant of the locus of selection in visual attention." *Perception and Psychophysics* 56 (2), pp.183-197.

Lee, M., 2003, "Species concepts and species reality: salvaging a Linnaean rank." *Journal of Evolutionary Biology*, 16: 179–188.

Leopold, D. and Logothetis, N. 1996. "Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry." *Nature,* 379 (6565), pp.549-553.

Levin, D. 2002. "Change blindness blindness as visual metacognition." *Journal of Consciousness Studies.* 9, pp.111-130.

Levine, J. 2001. *Purple Haze: the puzzle of consciousness.* (New York: Oxford University Press).

Levine, J. 2007. "Two kinds of access." *Behavioral and Brain Sciences,* 30, 514-515

Levine, J. 2010. "Demonstrative thought." *Mind and Language.* 25 (2), pp.169-195.

Lewis, D. 1966. "An argument for the identity theory." *The Journal of Philosophy,*63 (1), pp.17-25.

Lewis, D. 1970. "How to define theoretical terms." *The Journal of Philosophy,* 67 (13), pp.427-446.

Lewis, D. 1986. *On the plurality of Worlds.* (Malaysia: Blackwell)

Lewis, D. 1996. "Elusive knowledge." *Australasian Journal of Philosophy.* 74(4), pp.549-567.

Li, F. F., VanRullen, R., Koch, C. and Perona, P. 2002. "Rapid natural scene categorization in the near absence of attention." *Proceedings of the National Academy of Sciences of the United States of America.* 99, pp.9596-9601.

Ling, S. and Carrasco, M. 2006. "Sustained and transient covert attention enhance the signal via different contrast response." *Vision Research.* 46 (8-9), pp.1210-1220.

Logothetis, N. and Schall, J. 1989. "Neural correlates of subjective visual perception." *Science,* 245 (4919), pp.761-763.

Long, G. M. and Sakitt, B. 1980. "The retinal basis of iconic memory: Eriksen and Cllins Revisited." *American Journal of Psychology.* Vol.93, No.2. pp.195-206.

Lowe, E. J. 1995. "There are no easy problems of consciousness." In Shear, J. (ed.) *Explaining Consciousness: The Hard Problem.* (USA: MIT Press). pp.117-123.

Lowe, E. J. 2006. *The four category ontology: a metaphysical foundation for natural science.* (New York: Oxford University Press).

Lowe, E. J. 2012. "What is the source of our knowledge of modal truths?" *Mind.* 121 (484), pp.919-950.

Lycan, W. G. 1987. *Consciousness.* (Cambridge, Mass. MIT Press).

Lycan, W/ G. 1996. *Consciousness and Experience.* (USA: MIT Press).

Lycan, W. G. 2004. "The superiority of HOP to HOT." In Gennaro, R. J. (ed.) *Higher Order Theories of Consciousness: An Anthology.* (USA: John Benjamins Publishing). pp.93-113.

Lycan, W. G. 2008. "Phenomenal intentionalities." *American Philosophical Quarterly.* 45 (3), pp.233-252.

Machamer, P., Darden, L. and Craver, C. 2000. "Thinking about mechanisms." *Philosophy of Science.* 67 (1), pp.1-25.

Machery, E. 2005. "Concepts are not a natural kind." *Philosophy of Science.* 72 (3), pp.444-467.

Machery, E. 2009. *Doing Without Concepts.* (New York: Oxford University Press).

Machery, E. 2010a. "Precis of *Doing without concepts.*" *Behavioral and brain sciences.* 33 (2/3), pp.195-206.

Machery, E. 2010b. "The heterogeneity of knowledge representation and the elimination of *concept*." *Behavioral and brain sciences.* 33 (2/3), pp.231-244

Machery, E. 2012. "Why I stopped worrying about the definition of life… and why you should as well." *Synthese.* 185, pp.45-164.

Mack, A. and Rock, I. 1998. *Inattentional Blindness.* (Cambridge, MA: MIT Press).

Mack, A. 2002. "Is the visual world a grand illusion? A response." *Journal of Consciousness Studies.* 9, pp.102-110.

Magnus, P. D. 2012. "What SPECIES can teach us about THEORY" unpublished manuscript

Matthen, M. 2006. "On visual experience of objects: comments on John Campbell's *Reference and Consciousness." Philosophical Studies.* 127, pp.195-220.

Margolis, E. & Laurence, S. 2010. Concepts and Theoretical Unification, *Behavioral and Brain Sciences*, 33, pp.219–220

Margolis, Eric and Laurence, Stephen, "Concepts", *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2012/entries/concepts/>.

Marr, D. 1982. *Vision.* (New York: Freeman Press).

Martin, C. B. 1993. "Power for Realists." In Bacon, J., Campbell, K. and Reinhardt, L. *Ontology, Causality and Mind: Essays in Honour of D. M. Armstrong.* (USA: Cambridge University Press). Ch.8, pp.175-186.

Martin, C. B. 2007. *The Mind in Nature.* (New York: Oxford University Press).

Mayr, E. 1982. *The Growth of Biological Thought*. (Cambridge, MA: Harvard University Press).

Medin, D. L. and Schaffer, M. M. 1978. "Context theory of classification learning." *Psychological Review.* 85, pp.207-238.

Mellor, D.H. 1993. "Predicates and Properties." In Bacon, J. Campbell, K. and Reinhardt, L. (eds.) *Ontology, Causality and Mind: Essays in Honour of D. M. Armstrong.* (USA: Cambridge University Press). Ch.5, pp.101-118.

Milner, D. and Goodale, M. 1995. *The Visual Brain in Action.* (Oxford: Oxford University Press).

Milner, A.D. 2012. "Is visual processing in the dorsal stream accessible to consciousness?" *Proceedings of the Royal Society.* 279, pp.2289-2298.

Milner, A. D. and Goodale, M. 1995. *The Visual Brain in Action.* (Oxford: Oxford University Press).

Mole, C. 2008. "Attention and consciousness." *Journal of Consciousness Studies.* 15, pp.86-104.

Mole, Christopher, "Attention", *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/attention/>.

Mole, C. 2011a. *Attention is Cognitive Unison: an essay in philosophical psychology.* (New York: Oxford University Press).

Mole, C. 2011b. "The metaphysics of attention." In Mole, C., Smithies, D. and Wu, W. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch.3, pp.60-77.

Mole, C. 2012. "Attention." In E. Margolis, R. Samuels and S. Stich (eds.) *The Oxford Handbook of Philosophy of Cognitive Science.* (New York: Oxford University Press). pp.201-221.

Mole, C. forthcoming a. "Attention to unseen objects." *Journal of Consciousness Studies.*

Mole, C. forthcoming b. "Attention and cognitive penetration." In J. Zembeikis and T. Raftopoulos (eds.) *Cognitive Penetration.* (Oxford University Press).

Montagna, B., Pestilli, F. and Carrasco, M. 2009. "Attention trades off spatial acuity." *Vision Research.* 49 (7), pp.735-745.

Moore, C., Yantis, S. and Vaughan, M. 1998. "Object-based visual selection: evidence from perceptual completion." *Psychological Science.* 9 (2), pp.104-110.

Muller, F. A. 2007. "Inconsistency in classical electrodynamics." *Philosophy of Science,* 74, pp.253-277.

Murdoch, I. 1970. *The Sovereignty of the Good.* (Cornwall: Routledge). Reprinted 2002.

Nagel, T. 1974. "What is it like to be a Bat?" Reprinted in Heil, J. (ed.) *Philosophy of Mind: A Guide and Anthology.* 2004. (New York: Oxford University Press). Ch.29, pp.528-538.

Nanay, B. 2010. "Attention and Perceptual Content." *Analysis.* 70, pp.263-270.

Needham, R. 1972. *Belief, Language and Experience.* (Oxford: Blackwell).

Noë, A. 2002. "Is the visual world a grand illusion?" *Journal of Consciousness Studies.* 9, pp.1-12.

Noë, A. 2007. "Inattentional blindness, change blindness and consciousness." In Velmans, M. and Schneider, S. (eds.) *The Blackwell Companion to Consciousness.* (Singapore: Blackwell). Ch.39, pp.504-511.

Norman, L.J.; Heywood, C.A. & Kentridge, R.W. 2013. "Object-based attention without awareness." *Psychological Science.* 24 (6), pp.836-843.

Nussbaum, M. 2001. *Upheavels of Thought: the intelligence of emotions.* (Cambridge: Cambridge University Press).

O'Reagan, J. K. 1992. "Solving the "real" mysteries of visual perception: the world as an outside memory." *Canadian Journal of Psychology,* 46. pp.461-488.

O'Regan, J. K. and Noë, A. 2001a. "What it is like to see: a sensorimotor theory of perceptual experience." *Synthese.* 129, pp.79-103.

O'Regan, J. K. and Noë, A. 2001b. "A sensorimotor account of vision and visual consciousness." *Behavioral and Brain Sciences.* 24, pp.939-973.

O'Shaughnessy, B. 2002. *Consciousness and the World.* (New York: Oxford University Press).

Overgaard, M., Fehl, K., Mouridsen, K., Bergholt, B. and Cleermans, A. 2008. "Seeing without seeing? Degraded conscious vision in a blindsight patient.' *PLoS ONE* 38, pp.1-4.

Overgaard, M. 2006. "Introspection in science." *Consciousness and Cognition.* 12, pp.629-633.

Overgaard, M. 2011. "Visual experience and blindsight: a methodological review." *Experimental Brain Research.* 209, 473-479.

Overgaard, M., Rote, J., Mouridsen, K. and Ramsøy, T. Z. 2006. "Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task." *Consciousness and Cognition.* 15, pp.700-708.

Papineau, David. 2002. *Thinking about Consciousness.* (New York: Oxford University Press).

Peelen, M. V., Heslenfeld, D. J. and Theeuwes, J. 2004. "Endogenous and exogenous attention shifts are mediated by the same large-scale neural network." *NeuroImage* 22, pp.822-830.

Petrov, Y. and Popple, A. V. 2007. "Crowding is directed to the fovea and preserves only feature contrast." *Journal of Vision.* 7 (2) pp. 8. 1-9.

Phillips, I. 2014. "Cetacean semantics: a reply to Sainsbury." *Analysis.* doi: 10.1093/analys/anu052.

Piccinini, G. and Craver, C. 2011. "Integrating psychology and neuroscience: functional analyses as mechanism sketches." *Synthese.* 183, pp.283-311.

Piccinini, G. and Scott, S. 2006. "Splitting concepts." Philosophy of Science. 73, pp.390-490.

Posner, M. 1980. "Orienting of attention." *Quarterly Journal of Experimental Psychology.* 32, pp.3-25.

Posner, M. 1994. "Attention: the mechanisms of consciousness." *Proceedings of the National Academy of Science.* 91, pp.7398-7403.

Posner, M. 2012. Attentional networks and consciousness. *Frontiers in Psychology.* 3 (64), pp.1-4.

Posner, M. and Rothbart, M. 2007. "Research on attention networks as a model for the integration of psychological science." *Annual Review of Psychology.* 58, pp.1-23.

Price, H. 1992. "Metaphysical pluralism." *Journal of Philosophy.* 89 (8), pp.387-409.

Price, H. 2003. "Truth as convenient friction." *Journal of Philosophy.* 100 (4), pp. 167-190.

Price, H. 2011. *Naturalism Without Mirrors.* (New York: Oxford University Press).

Prinz, J. 2002. *Furnishing the Mind: concepts and their perceptual basis*. (Cambridge, MA.: MIT Press).

Prinz, J. 2004. *Gut Reactions: A perceptual theory of emotion.* (Oxford: Oxford University Press).

Prinz, J. 2007a. "Accessed, accessible and inaccessible: where to draw the phenomenal line" *Behavioral and Brain Sciences.* 30, pp.521-522.

Prinz, J. 2007b. "The intermediate level theory of consciousness." In *The Blackwell Companion to Consciousness.* Velmans, M. and Schneider, S. (eds.) (Singapore: Blackwell). Ch.19, pp.247-260.

Prinz, J. 2010a. "When is perception conscious?" In Nanay, B. (ed.) *Perceiving the World.* (New York: Oxford University Press). Ch.11, pp.310-332.

Prinz, J. 2010b. "Can concept empiricism forestall eliminativism?" *Mind and Language.* 25 (5), pp.612-621.

Prinz, J. 2011. "Is attention necessary and sufficient for consciousness." In Mole, C., Smithies, D. and Wu, W. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch.8, pp.174-203.

Prinz, J. 2012. *The Conscious Brain: How Attention Engenders Experience.* (New York: Oxford University Press).

Prinz, J. 2013a. "Siegel's get rich quick scheme." *Philosophical Studies.* 163 (3), pp.827-835.

Prinz, J. 2013b. "Attention, atomism, and the disunity of consciousness." *Philosophy and Phenomenological Research.* 86 (1), pp.215-222.

Putnam, H. 1975. "The meaning of 'meaning'" In K. Gunderson (ed.) *Language, Mind, and Knowledge.* (Minneaplois: University of Minnesota Press).

Quine, W. V. O. 1969. "Natural kinds" In *Ontological Relativity and Other Essays*. (New York: Columbia University Press).

Raftopoulos, A. 2009. "Reference, perception, and attention." *Philosophical Studies* 144, pp.339-360.

Ratcliffe, M. 2007. *Rethinking Commonsense Psychology: A critique of folk psychology, theory of mind and simulation.* (Basingstoke: Palgrave Macmillan).

Ratcliffe, M. 2008. "Farewell to folk psychology: response to Hutto." *International Journal of Philosophical Studies.* 16 (3), 445-451.

Reddy, V. 2010. "Engaging Minds in the First Year: The Developing Awareness of Attention and Intention", in G. Bremner (ed.), *Handbook of Infant Development*, Oxford: Wiley Blackwell. pp. 365–393.

Remington, R.W. and Folk, C.L. 2001. "A dissociation between attention and selection." *Psychological Science.* 12, 511-515.

Rensink, R. 2013. "Perception and attention." In D. Reisberg (ed.) *The Oxford Handbook of Cognitive Psychology* (New York: Oxford University Press). pp.97-116.

Rensink, R. 2014. "A function-centered taxonomy of attentional processes." In Coates, P. and Coleman, S. (eds.) *Phenomenal Qualities: Sense, Perception and Consciousness.* (Oxford: Oxford University Press).

Richards, B. 2013. "Identity crowding and object seeing: a reply to Block." *Thought: a journal of philosophy.* 2 (1), pp. 9-19.

Rieppel, O. 2005. "Modules, Kinds and Homology." *Journal of Experimental Zoology.* 304B, pp.18-27.

Rock, I and Gutman, D. 1981. "The effect of inattention on form perception." *Journal of Experimental Psychology: Human Perception and Performance.* 7, pp.275-285.

Rosch, E. 1978. Principles of Categorization, in E. Rosch & B. Lloyd (eds.), *Cognition and Categorization*. (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 27–48

Rosenthal, D. 1986. "Two concepts of consciousness." *Philosophical Studies.* 49, pp.329-359.

Rosenthal, D. 1997. "A theory of consciousness." In *The Nature of Consciousness: Philosophical Debates.* Block, N., Gulzeldere, G. and Flanagan, O. (eds.). 1997. Reprinted 2002. (USA: MIT Press). 46, pp.729-753.

Rosenthal, D. 2004. "Varieties of higher-order theory." In Gennaro, R. J. (ed.) *Higher Order Theories of Consciousness: An Anthology.* (USA: John Benjamins Publishing). pp.17-44.

Rosenthal, D. 2005. *Consciousness and Mind.* (New York: Oxford University Press).

Sainsbury, M. 2014. "Fishy business." *Analysis.* 74 (1), pp.3-5.

Samuels, R. and Ferreira, M. 2010. "Why *don't* concepts constitute a natural kind?" *Behavioral and Brain Sciences.* 33, pp. 222-223.

Sauret, W. and Lycan, W. 2014. "Attention and internal monitoring." *Analysis.* 74 (3), pp.363-370.

Schwitzgebel, E. 2007. "Do you have a constant tactile experience of your feet in your shoes? Or is experience limited to what's in attention?" *Journal of Consciousness Studies.* 14 (33), pp.5-35.

Schwitzgebel, E. 2011. *Perplexities of consciousness.* (USA: MIT Press).

Shanahan, M. and Baars, B. 2007. "Global workspace theory emerges unscathed." *Behavioral and Brain Sciences.* Vol.30, pp.524-525.

Shoemaker, S. 1998. "Causal and Metaphysical Necessity." *Pacific Philosophical Quarterly.* Vol.79. pp.59-77.

Sidelle, A. 2007. "The method of verbal dispute." *Philosophical Topics,* 35, 83-113.

Sider, T. 2011. *Writing the Book of the World.* (New York: Oxford University Press).

Siegel, S. (2006). "How does visual phenomenology constrain object-seeing?" *Australasian Journal of Philosophy.* 84.3, pp.429-41.

Siegel, S. 2006. "How does visual phenomenology constrain object-seeing?" *Australasian Journal of Philosophy.* 84 (3), pp.429-441.

Siegel, S. 2010. *The Contents of Visual Experience.* (USA: Oxford University Press).

Siegel, S. and Silins, N. Forthcoming. "Attention and perceptual justification." In D. Stoljar and A. Pautz (eds.) *Themes from Block.* (MIT press).

Siewert, C. 2002. "Is visual experience rich or poor?" *Journal of consciousness studies.* 9, pp.131-140.

Silins, N. and Siegel, S. Forthcoming. "Consciousness, attention and justification." In Dodd, D. and Zardini, E. (eds.) *Contemporary Perspectives on Scepticism and Perceptual Justification.* (Oxford University Press).

Simons, D. J. and Chabris, C. F. 1999. "Gorillas in our Midst: Sustained Inattentional Blindness for Dynamic Events." *Perception.* Vol.28. pp.1059-1074.

Simons, D. J. 2000. "Current approaches to change blindness." *Visual Cognition.* Vol.7, pp.1-16.

Sligte IG, Scholte HS, Lamme VAF. 2008. "Are There Multiple Visual Short-Term Memory Stores?" PLoS ONE 3(2): e1699. doi:10.1371/journal.pone.0001699

Sligte, I.G., Scholte, S. and Lamme, V.A.F. (2009). "V4 Activity Predicts the strength of Visual Short term memory representations." *The Journal of Neuroscience.* 29(23), pp.7432-7438.

Smithies, D. 2011a. "Attention is rational access-consciousness." In Mole, C., Smithies, D. and Wu, W. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch.11, pp.247-273.

Smithies, D. 2011b "What is the role of consciousness in demonstrative thought?" *Journal of Philosophy.* 108 (1), pp.5-34.

Sober, E. 1980. "Evolution, population thinking, and essentialism." *Philosophy of Science.* 47 (3), pp.350-383.

Soteriou, M. 2007. "Content and the Stream of Consciousness." *Philosophical Perspectives.* 21, pp.543-568.

Soto, D., Mäntylän T. and Silvanto, J. 2011. "Working memory without consciousness." *Current Biology.* 21. pp.R912-R193.

Sperling, G. 1960. "The information available in Brief Visual Presentations." *Psychological Monographs.* 74 (11), pp.1-29.

Speaks, J. 2010. "Attention and intentionalism." *The Philosophical Quarterly.* 60 (239), pp.325-342.

Stanford, P. 1995. "For pluralism and against realism about species." *Philosophy of Science* 62, pp.70-91.

Stanford, K. and Kitcher, P. 2000. "Refining the causal theory of reference for natural kind terms." *Philosophical Studies.* 97, pp.99-129.

Stazicker, J. 2011. "Attention, visual consciousness and indeterminacy." *Mind and Language.* 26 (2), pp.156-184.

Stoerig, P. 2010. "Cueless blindsight." *Frontiers in Neuroscience.* 3, pp.1-8.

Strawson, Galen. 2006. *Consciousness and its place in nature: does physicalism entail panpsychism?* (Exeter: Imprint Academic).

Strohminger, N. and Moore, B. 2010. "Banishing the thought." *Behavioral and Brain Sciences.* 33 (2/3): 225-226.

Styles, E. 1997. *The Psychology of Attention.* (Cornwall: Psychology Press).

Taylor, J. H. 2013a. "Is attention necessary and sufficient for phenomenal consciousness?" *Journal of Consciousness Studies.* 20 (11-12), pp.173-194.

Taylor, J. H. 2013b. "Is the grain of vision finer than the grain of attention? Response to Block." *Thought: a journal of philosophy.* 2 (1), pp.20-28.

Taylor, J.H. 2013c. "Physicalism and phenomenal concepts: bringing ontology and philosophy of mind together." *Philosophia.* 41, pp.1283-1297.

Taylor, J. H. 2013d. "Is consciousness science fundamentally flawed?" *Journal of Consciousness Studies.* 20, pp.203-221.

Taylor, J. H. 2014. "Against unifying accounts of attention." *Erkenntnis.* Available online. DOI: 10.1007/s10670-014-9611-3.

Teichman, J. 1985. "The Definition of Person." *Philosophy* 60 (232), pp.175-185.

Treisman, A. 1988. "Features and objects: the fourteenth Bartlett memorial lecture." *Quarterly Journal of Experimental Psychology.* 40: 201-237.

Treisman, A. 1996. "The binding problem." *Current Opinion in Neurobiology.* 6: 171-178.

Treisman, A. and Gelade, G. 1980. "A feature-integration theory of attention." *Cognitive Psychology.* Vol.12, pp.97-136.

Tse, P. 2005. "Voluntary attention modulates the brightness of overlapping transparent surfaces." *Vision Research.* 45 (9), pp.1095-1098.

Tsuchiya, N. and Koch, C. 2005. "Continuous flash suppression reduces negative afterimages." *Nature Neuroscience* 8 (8), pp.1096-1101.

Tsuchiya, N., Block, N. and Koch, C. 2012. "Top-down attention and consciousness: comment on Cohen et al." *Trends in Cognitive Sciences.* Vol.16, p.527.

Tye, M. 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind.* (USA: MIT Press).

Tye, M. 2009a "A New Look at the Speckled Hen." *Analysis.* 69, pp.258-63.

Tye, M. 2009b. *Consciousness Revisited: Materialism without phenomenal concepts.* (Hong Kong: MIT Press).

Tye, M. 2010. "Attention, seeing and change blindness." *Philosophical Issues.* 20, pp.411-437.

Tye, M. 2014. "Does conscious seeing have a finer grain than attention." *Thought: a journal of philosophy.* 3 (2), pp.154-158.

Uidhir, C. M. and Magnus, P.D. 2011. "Art concept pluralism." *Metaphilosophy.* 42 (1-2): 83-97.

Uttal, W. R. 2011. *Mind and Brain: a critical appraisal of cognitive neuroscience.* (USA: MIT Press.)

Van Boxtel, J. J. A., Tsuchiya, N. and Koch, C. 2010. "Consciousness and attention: on sufficiency and necessity." *Frontiers in psychology.* 1(217), pp.1-13.

Vecera, S. P. and Rizzo, M. (2003) "Spatial attention: normal processes and their breakdown." *Neurologic Clinics* 21, pp.575-607.

Vickers, P. 2008. "Frisch, Muller and Belot on an inconsistency in classical electrodynamics." *British Journal for the Philosophy of Science.* 59, (4), pp.1-26.

Vickers, P. 2013. *Understanding inconsistent science.* (Oxford: Oxford University Press).

Vickers, P. 2014. "Scientific Theory Eliminativism." *Erkenntnis.* 79, pp.111-126.

Watzl, S. 2011a. "Attention as Structuring the Stream of Consciousness." In Mole, C., Smithies, D. and Wu, W. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch.7, pp.145-173.

Watzl, S. 2011b. "The nature of attention." *Philosophy Compass.* 6/11, pp.842-853.

Watzl, S. 2011c. "The philosophical significance of attention." *Philosophy Compass.* 6/10, pp.722-733.

Watzl, S. 2013. "Review of *Attention is Cognitive Unison: an essay in philosophical psychology.*" Notre Dame Philosophical Reviews. Available at: http://ndpr.nd.edu/news/27026-attention-is-cognitive-unison-an-essay-in-philosophical-psychology. Accessed 29/May/2013.

Watzl, S. 2014. "Attentional organization and the unity of consciousness." *Journal of Consciousness Studies.* 21 (7-8), pp.56-87.

Weinberg, J., Nichols, S., and Stich, S. 2001. "Normativity and epistemic intuitions." *Philosophical Topics.* 29: 429-60. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.174.4748&rep=rep1&type=pdf. Accessed 6/November/2013. Citations from this version.

Weiskopf, D. 2009. "The plurality of concepts." *Synthese.* 169, pp.145-173.

Weiskopf, D. 2010. "The theoretical indispensability of concepts." *Behavioral and brain sciences.* 33(2/3), pp.228-229.

Wiggins, D. 1998. *Needs, Values, Truth.* (Oxford: Oxford University Press).

Wilkins, J. 2009. *Species: The history of the idea*. (Berkeley: University of California Press).

Williamson, T. 2000. *Knowledge and its Limits.* (New York: Oxford University Press).

Wilson, M. 2006. *Wandering Significance*. (Oxford: OUP).

Wilson, R. A. 1996. "Promiscuous realism." *British Journal for the Philosophy of Science.* 46, pp.303-316.

Wilson, R. A., Barker, M. J. and Brigandt, I. 2007. "When traditional essentialism fails: Biological natural kinds." *Philosophical Topics.* 35, pp.189-215. Available at:

http://joelvelasco.net/teaching/167win10/BioNatKinds.W.B.B.mar08.pdf. Accessed 11/Jul/2013. Citations from this version.

Wittgenstein, L. 1953. *Philosophical Investigations.* Trans. Anscombe, E. (Oxford: Blackwell). Reprinted 1968.

Wolfe, J. M. and Horowitz, T.S. 2004. "What attributes guide the deployment of visual attention and how do they do it?" *Nature Reviews: Neuroscience.* 5(6), pp.495-501

Wu, W. 2008. "Visual attention, conceptual content, and doing it right." *Mind.* 117, pp.1003-1032

Wu, W. 2010. "What is conscious attention?" *Philosophy and Phenomenological Research.* 83, pp.93-120

Wu, W. 2011a. "Attention as selection for action." In Mole, C., Smithies, D. and Wu, W. (eds.) *Attention: Philosophical and Psychological Essays.* (New York: Oxford University Press). Ch.5, pp.97-116.

Wu, W. 2011b. "Confronting many-many problems: attention and agentive control." *Noûs.* 45, pp.50-76.

Wu, W. 2013. "Mental action and the threat of automaticity." In Clark, A., Kiverstein, J. and Vierkant, T. (eds.) *Decomposing the Will* (Oxford University Press), pp.244-261

Wu, W. 2014. *Attention.* (New York: Routledge).

Zagzebski, L., 1994. "The Inescapability of Gettier Problems," *The Philosophical Quarterly*, 44(174): 65–73.