

**On the Plausibility of a  
Strong Transcendental  
Response to Scepticism**

**Rosemary Jane Smith**

**PhD**

**University of York**

**Philosophy**

**May, 2015**

## Abstract

I argue that the strong transcendental strategy can offer us a serious and robust response to the sceptic who doubts that we can have knowledge of the extra-mental.

This sceptic is motivated by sceptical hypotheses to worry that I could have all the thoughts and experiences I do in fact have if the world were radically different to how I take it to be. Transcendental arguments start from a premise about our thoughts or experience and move on to show that something must be the case because it is a necessary condition of our having such thoughts or experience. As such, transcendental arguments are particularly well-positioned to answer this sceptic, as the premises of a transcendental argument are drawn from the mental propositions the sceptic accepts.

Some philosophers have argued that the same concerns that drive the sceptic to doubt the extra-mental should also give her cause to doubt the mental. To prevent the sceptic retreating to this thought, I show that these arguments are only effective against propositions that were unlikely to form the basis of a transcendental argument.

Strong transcendental arguments (STAs) are differentiated from weak transcendental arguments as being those that move from mental premises to conclusions about the extra-mental world. I defend STAs against Stroud's objection that this is not possible, on the basis that the objection rests on an illicit assumption of dualism about mind and world. I argue for the plausibility of supervenience physicalism as a metaphysical picture upon which such inferences would be possible.

I show how a dispositionally essentialist understanding of the laws of nature would plausibly support a metaphysically necessary psychophysical law, from which we could draw the bridging premise of an STA. This changes the dialectic, forcing the sceptic to defend specific metaphysical positions, such as resemblance nominalism, and to engage substantially with philosophy. The plausibility of strong transcendental arguments tells us something of what must be true of the world for the sceptic's arguments to even get started.

ABSTRACT.....	2
ACKNOWLEDGEMENTS.....	7
AUTHOR’S DECLARATION .....	8
CHAPTER 1: INTRODUCTION.....	9
<b>1.1 SCEPTICISM DEFINED .....</b>	<b>9</b>
<b>1.2 SCEPTICISM, APPEARANCE, AND REALITY IN LITERATURE .....</b>	<b>13</b>
<b>1.3 A POTTED HISTORY OF PHILOSOPHICAL SCEPTICISM .....</b>	<b>16</b>
1.3.4 SIXTH CENTURY BC—JAINIST EPISTEMOLOGY .....	19
1.3.2 FIFTH CENTURY BC—SOCRATIC SCEPTICISM .....	19
1.3.3 FOURTH OR SIXTH CENTURY BC—LAOZI .....	20
1.3.4 FOURTH CENTURY BC TO THIRD CENTURY AD—PYRRHONIAN SCPETICISM .....	21
1.3.5 THIRD TO FIRST CENTURY BC—ACADEMIC SCEPTICS .....	22
1.3.6 FOURTH CENTURY AD—AUGUSTINE .....	23
1.3.7 MEDIEVAL SCEPTICISM .....	23
<b>1.4 SCEPTICISM AS AN ENDURING CONCERN.....</b>	<b>25</b>
<b>1.5 TRANSCENDENTAL ARGUMENTS AS A RESPONSE TO SCEPTICISM.....</b>	<b>26</b>
CHAPTER 2: TRANSCENDENTAL ARGUMENTS.....	30
<b>2.1 THE TRANSCENDENTAL DEDUCTION.....</b>	<b>30</b>
<b>2.2 KANT’S COPERNICAN TURN .....</b>	<b>32</b>
<b>2.3 THE REFUTATION OF IDEALISM .....</b>	<b>35</b>
2.3.1 TRANSCENDENTAL IDEALISM AND MATERIAL IDEALISMS.....	35
2.3.2 THE PROOF.....	36
<b>2.3.3 THE KANTIAN AND CARTESIAN ‘I AM’ .....</b>	<b>38</b>
2.3.4 TRANSCENDENTAL IDEALISM, BERKELEYAN IDEALISM, AND SCEPTICISM .....	41
<b>2.4 AUSTIN .....</b>	<b>45</b>

<b>2.5 STRAWSON: <i>INDIVIDUALS</i></b> .....	<b>48</b>
<b>2.6 STRAWSON'S TRANSCENDENTAL DEDUCTION AND <i>THE BOUNDS OF SENSE</i></b> .....	<b>50</b>
<b>2.7 STROUD AND TRANSCENDENTAL ARGUMENTS</b> .....	<b>52</b>
2.7.1 TRANSCENDENTAL ARGUMENTS AS ANTI-SCEPTICAL .....	53
2.7.2 GOALS FOR TRANSCENDENTAL ARGUMENTS.....	55
2.7.3 STRAWSON: NATURALISM AND SCEPTICISM .....	56
2.7.4 STROUD AND WEAK TRANSCENDENTAL ARGUMENTS.....	57
2.7.5 STROUD'S TAXONOMY OF TRANSCENDENTAL ARGUMENTS .....	59
<b>2.8 WHY A STRONG TRANSCENDENTAL ARGUMENT IS NEEDED TO ADDRESS MY SCEPTIC</b> .....	<b>60</b>
2.8.1 AGAINST WEAK TRANSCENDENTAL ARGUMENTS .....	61
2.8.2 STRAWSONIAN WEAK TRANSCENDENTAL ARGUMENTS .....	62
<b>2.9 CONCLUSION</b> .....	<b>63</b>
CHAPTER 3: WHAT'S SO SPECIAL ABOUT THE MENTAL? .....	64
<b>3.1 INTRODUCTION</b> .....	<b>64</b>
<b>3.2 DOUBTING ONE'S OWN MIND</b> .....	<b>65</b>
<b>3.3 EMOTION</b> .....	<b>66</b>
3.3.1 ANTICIPATED AND REMEMBERED EMOTION .....	67
3.3.2 LINGUISTIC VAGUENESS .....	68
3.3.3 DISTINGUISHING MOODS AND EMOTIONS .....	69
3.3.4 PRESENTLY FELT EMOTIONS.....	70
<b>3.4 VISUAL EXPERIENCE</b> .....	<b>73</b>
3.4.1 'I THINK I CAN PROTRUDE MY TONGUE WITHOUT ITS COMING OUT' .....	73
3.4.2 'I SEE A SEAL AS MY SISTER' .....	74
3.4.3 'I THINK I SEE A RED CARPET THAT'S NOT RED' .....	75
3.4.4 INTROSPECTION, INTROSPECTIVE JUDGEMENTS, AND NEUROSURGEONS FROM ALPHA CENTAURI.....	78
3.4.5 BLINDSIGHT .....	80
<b>3.5 PERIPHERAL AND FOVEAL VISION</b> .....	<b>80</b>
<b>3.6 THE PHENOMENOLOGY OF THOUGHT</b> .....	<b>83</b>
<b>3.7 THE GENERAL ARGUMENT</b> .....	<b>84</b>
<b>3.8 CARTESIAN INFALLIBILITY VS INCORRIGIBILITY</b> .....	<b>85</b>
<b>3.9 THE ARGUMENT FOR INCORRIGIBILITY</b> .....	<b>87</b>

<b>3.10 THE SUPERSTITIOUS MAN .....</b>	<b>89</b>
<b>3.11 CAPGRAS SYNDROME.....</b>	<b>91</b>
<b>3.12 CONCLUSION.....</b>	<b>95</b>

CHAPTER 4: MIND IN WORLD—STRONG TRANSCENDENTAL ARGUMENTS WITHOUT A CHASM OF NECESSITY .....	96
---------------------------------------------------------------------------------------------	----

<b>4.1 STROUD, TRANSCENDENTAL ARGUMENTS, AND THE VERIFICATION PRINCIPLE.....</b>	<b>97</b>
<b>4.2 THE BRIDGE OF NECESSITY .....</b>	<b>99</b>
<b>4.3 TRUE OF THE MIND VS TRUE OF THE WORLD .....</b>	<b>100</b>
<b>4.4 AN OBJECTION: THE PSYCHOLOGICAL AND NON-PSYCHOLOGICAL ARE DIFFERENT CATEGORIES OF FACT.....</b>	<b>102</b>
<b>4.5 MIND IN WORLD .....</b>	<b>103</b>
<b>4.6 SUBSTANCE DUALISM .....</b>	<b>103</b>
<b>4.7 MONISM AND IDEALISM .....</b>	<b>105</b>
<b>4.8 TRANSCENDENTAL IDEALISM.....</b>	<b>106</b>
<b>4.9 PHYSICALISM .....</b>	<b>107</b>
4.9.1 DEFINING PHYSICALISM .....	107
4.9.2 DEFENDING SUPERVENIENCE PHYSICALISM: DIFFERING MENTALLY WHILST NOT DIFFERING PHYSICALLY.....	108
4.9.3 DEFENDING SUPERVENIENCE PHYSICALISM: INDETERMINISM .....	111
<b>4.10 SUPERVENIENCE, THE COMPLETENESS OF PHYSICS, AND COLOURED FILTERS .....</b>	<b>113</b>
<b>4.11 WHAT IS ‘PHYSICS’ AND WHY IS IT COMPLETE? .....</b>	<b>115</b>
<b>4.12 SUPERVENIENCE VIA ENTAILMENT .....</b>	<b>116</b>
<b>4.13 MULTIPLE REALISEABILITY .....</b>	<b>118</b>
<b>4.14 CONCLUSION.....</b>	<b>119</b>

CHAPTER 5: METAPHYSICAL NECESSITY AND THE LAWS OF NATURE.....	121
---------------------------------------------------------------	-----

<b>5.1 THE SECOND PREMISE.....</b>	<b>122</b>
<b>5.2 LAWS OF NATURE AND THE HUMEAN DICTUM .....</b>	<b>123</b>
<b>5.3 NECESSITY, ANALYTICITY, AND THE A POSTERIORI.....</b>	<b>126</b>
<b>5.4 DISPOSITIONAL ESSENTIALISM AND SCIENTIFIC REALISM .....</b>	<b>128</b>
<b>5.5 LAWS AND PROPERTIES.....</b>	<b>133</b>
<b>5.6 BIRD’S ESSENTIALISM .....</b>	<b>137</b>

<b>5.7 DISPOSITIONAL ESSENTIALISM AND REALISM ABOUT UNIVERSALS .....</b>	<b>139</b>
<b>5.8 DISPOSITIONS .....</b>	<b>144</b>
<b>5.9 FINKS, ANTIDOTES, AND MIMICS .....</b>	<b>146</b>
<b>5.10 REPAIRING THE CONDITIONAL ANALYSIS: <i>CETERIS PARIBUS</i> LAWS .....</b>	<b>147</b>
<b>5.11 FUNDAMENTAL DISPOSITIONS CANNOT BE FINKISH.....</b>	<b>148</b>
<b>5.12 THE METAPHYSICAL NECESSITY OF LAWS OF NATURE.....</b>	<b>151</b>
<b>5.13 FINKISH DISPOSITIONS, PSYCHOPHYSICAL LAWS, AND THE SCEPTIC.....</b>	<b>154</b>
<b>5.14 NECESSITARIANISM ABOUT LAWS.....</b>	<b>158</b>
<b>5.15 REGULARITY AND DISPOSITIONAL ESSENTIALISM .....</b>	<b>161</b>
<b>5.16 CONCEIVABILITY AND THE ILLUSION OF NOMIC CONTINGENCY.....</b>	<b>161</b>
<b>5.17 CONCLUSION.....</b>	<b>163</b>
<b>CONCLUSION.....</b>	<b>165</b>
<b>BIBLIOGRAPHY .....</b>	<b>167</b>

## Acknowledgements

Acknowledgement is due to a great many people who have joined me on this eight year Odyssey into the transcendental.

To my supervisor, Tom Stoneham, for his insightful guidance, questioning, and support. To David Efird, for introducing me to Donald Davidson and helping me realise that what excited me about the argument from triangulation was its transcendental form.

To the members of my thesis advisory panel. To Andrew Ward, for his probing questions and challenging me to a deeper understanding of Kant. To James Clarke, for his support and encouragement, for sharing my interest in transcendental arguments, and for offering me a different perspective on them.

Special thanks to Julie Kay, who has helped me in ways innumerable and been a bastion of support and knowledge in the battle against university systems.

To those kind friends and strangers whose donations kept a roof over my head in these final months: Marcus Smith, Edward Allison, Shelley MacGregor, Steve Cooke, Alice Pember, Heath Graham, Amanda Snyder, Jessica Meats, Lee Harris, Greg Eden, Paul Spirito, and all those who wished to remain anonymous.

To my parents for their support, financial and personal, and to the rest of my family, for their tolerance.

Finally, to Steve Cooper, Clare Wall, and Fred Prior, without whose friendship and support I would not have made it.

## **Author's Declaration**

I declare that this thesis is a presentation of original work of which I am the sole author. It has not been previously presented for an award at this university or any other. All sources are acknowledged as references.



## Chapter 1: Introduction

And now, after all paths (as people are persuading themselves), have been tried in vain, there prevails in the sciences a weariness and utter *indifferentism*, which is the mother of chaos and night—yet is also the source, or at least the prelude, of their approaching reform and enlightenment after ill-applied diligence has left them dark, confused, and useless.

For it is futile to try to feign *indifference* concerning inquiries whose object *cannot be indifferent* to human nature. (Kant A x<sup>1</sup>)

Kant refers to the plight of metaphysics, for which he charts a rocky history, from which, he tells us, philosophers ultimately gave up, viewing all paths as exploited and found wanting, becoming indifferent in the face of apparently irresolvable difficulties (Kant A ix–x).

A similar indifferentism towards scepticism has arisen in modern times. Metaphysics is only considered lost by certain factions, but there is a general belief that in the question of what is to be *known* about the nature of reality, we shall only begin to make progress if the sceptic is ignored. Thus, epistemology, when it deals with the sceptic at all, focuses its energies on giving us reasons to dismiss her, instead of answering her.

I do not think the sceptic's incessant questioning can be so easily dismissed. I find that she still wants an answer, and I still want to give it to her.

### 1.1 Scepticism Defined

There is more than one type of scepticism, so I will start by defining the kind of scepticism that interests me. I am perfectly willing to concede that there are some kinds of scepticism we are entitled to ignore, and some which it would be wrong to suggest the philosophical mainstream pretends any kind of indifferentism towards. Moral and religious scepticism have become very popular, and it is not only regarded as common, but relatively unproblematic to be sceptical about moral absolutes or the existence of God. Scepticism about objective

---

<sup>1</sup> References to Kant will be to the 1781 (A) and 1787 (B) editions of the *Critique of Pure Reason*. Translation is from the Pluhar (1996).

standards for beauty or art are similarly respectable positions. Which is not to suggest that they do not have their detractors. Rather, they are not subject to the same kind of scorn that the kind of scepticism that concerns me is subjected to. I speak of the sceptical worry that we cannot have any knowledge of non-mental objects. This is a sceptic who examines the evidence of her sensory experience and concludes that it might have been caused by a world greatly different from that which appears to be present in her experience, or that she believes to exist. Having considered certain sceptical scenarios, she finds that all the evidence available to her could just as easily have existed had she been dreaming, or the victim of a malicious demon's fabrication, or, most worrying of all, that she might have been an isolated solipsistic consciousness who, in her madness, hallucinates the experience of a non-mental world that does not exist at all.

It is this last hypothesis that most directly concerns me. I think it is the most worrying, but also the most plausible. Although others doubt its success, I am satisfied with the *cogito*, or some version of it at least, and that I cannot intelligibly doubt my own existence. I do not think my thoughts are in fact being had by an evil demon, even if I am otherwise the benighted fool of such a creature; without the ability to generate doubt for my own existence, positing an evil demon would mean that there is at least one other being in existence beyond me, even if they are evil. It is rather more worrying that I might be entirely alone. Similarly, a dreamer who is sometimes awake and has a physical body in some real world is better off than the deluded solipsistic consciousness.

Nevertheless, I think these other hypotheses can be valuable in allowing us to explore the capabilities for fabrication of the solipsistic, purely mental consciousness. As Descartes realised, dreams are an altered state of mind we all experience regularly, informing our understanding of the great variety of falsehoods we can be brought to believe.

Ernest Sosa (2007), has argued that we cannot have beliefs in dreams, that all we have are imaginings, but I hold, with Ichikawa (2008: 523), that even if this is so, the 'quasi-affirmation' that takes the place of belief in imaginings is subjectively indistinguishable from beliefs *when one is dreaming*. Similarly, for the general disjunctivist thought that there is something fundamentally different about veridical perceptions and the content of dreams, imaginings, and hallucinations. One may insist as one likes that veridical perception is significantly different from dreaming, or that one knows one is not dreaming or hallucinating when one is under normal or optimal waking conditions, but when I am dreaming I have no

such knowledge. Some have experienced lucid dreaming—a state in which they are asleep and dreaming and they know that this is the case. I congratulate them on their good fortune; I have never had this pleasure. I have, however, dreamt that I had woken up when I was still dreaming. I have even run certain tests upon myself and either believed that I had passed them or made some excuse for why they failed.

In fact, I think I very much do have beliefs in dreams, but for the purposes of argument, I think it may be stipulated. The sceptic who concerns me thinks I might have all the beliefs and experiences I do in fact have even if the world were radically different. I think she deserves a robust response. I think her sceptical scenarios are *prima facie* plausible, and cannot be dismissed solely on the basis that I can tell, really, when my experiences are real.

Nothing hangs on the mechanism of dreaming or hallucination. What concerns me is that there seems to have been persistent doubt, throughout history, about the relationship between our sensory experience and the external world. A concern that such experience might be generated by one's own mind, or some other mechanism via which that which appears to our senses is not that which it appears to be.

I avoid the term 'external world' scepticism, although I think what is generally meant by that term matches the doubt that concerns me. My discomfort with the term 'external' lies in the ambiguity concerning what is and what is not 'external'. Does it mean 'external to the self'? Defining the self is a complex quest with oft disputed results. If one identifies the self with one's body there is a clear conflict with the sort of scepticism that doubts the existence of a physical world. Does it mean 'external to the mind'? Well, then we run the risk of begging the question against cognitive extension advocates, like Andy Clark (see his 2008), who maintain that many of our cognitive resources may be located outside the body. For example, some of my memory might be said to be stored 'external' to my body, on my computer, where I keep my notes. The scepticism that concerns me is worried that there might not be any computer existing distinct from my mind; this is only a concern for the cognitive extension advocate if he already accepts that the computer is not a purely mental object. I would expect most cognitive extension theorists to have such a commitment, but to force them into a question begging position if they are worried by the sort of scepticism that concerns me seems unnecessarily restrictive. Words like 'external' and 'internal' are evocative of physical distinctions, or a distinction between the mind and the physical, and are therefore confusing.

I shall thus talk of the ‘mental’ and the ‘non-mental’ or ‘extra-mental’, rather than the ‘internal’ and the ‘external’. This should clarify that the concern is specifically that propositions concerning non-mental objects<sup>2</sup> are not known with the same degree of certainty that seems available to at least some propositions concerning mental objects or properties. The worry will encompass doubts about the existence of non-mental entities without begging the question for theories of mind that assume mind-body identity, or even that the mind may extend to objects external to the body.

This is the form of scepticism towards which there seems to be a peculiar kind of indifferentism. This is the position against which we are told we argue without an opponent; anyone who claims to be a sceptic of this kind is accused either of insanity or insincerity. Hume (ECHU 5.1.34)<sup>3</sup> is right that it is not a doubt that is evinced in my everyday behaviour. When I make a cup of tea I seem to assume that the kettle, mug, water, and teabag are all there; that the laws of physics hold, that the electricity in the kettle’s element will heat the water and that the application of hot water to the teabag will cause the chemical reaction that we call an infusion of tea. Trust—in a very great deal—is a necessary part of survival, and hence continued existence in the world as a human being. Yet I am convinced that it is as much a part of human nature to worry quite seriously about scepticism from time to time as it is to operate as though those doubts were unjustified at every moment for which they are not called to mind. Trust is rarely an attitude consciously avowed in a continuous manner. Indeed, trust is only a sensible attitude where there is reason to doubt. Otherwise we there would be no need for trust, as we would have knowledge.

The ‘sceptic’ is almost always talked about in inverted commas. No one, we are told, is really a sceptic. She is either someone we falsely believe we must defend our positions against, or she is a hypothesised opponent against whose imagined critiques we test our theories. However, I do not think this reflects reality as well as it has been the habit of modern philosophers to assert.

---

<sup>2</sup> Or, objects matching non-mental descriptions, or objects with non-mental properties. It is my hope that there are at least some objects matching non-mental descriptions or possessed of non-mental properties, that also have mental properties or match mental descriptions.

<sup>3</sup> References to the Hume (1748 and 1777), *An Enquiry Concerning Human Understanding*, are given by ‘ECHU’.

I argue both that the sceptic can be found in *real human beings* who have claimed to be sceptics in the past, and that, for many of us, consistently and throughout philosophical history, the sceptic is a live and serious *part of ourselves*. In this second guise, the sceptic is not an external opponent, real or imagined. I do not claim that as we go about our daily business we fail to believe or make knowledge claims in a normal manner. Rather, in our daily lives we *trust* that we are in a good epistemic position to make knowledge claims or declare our beliefs reasonable. This trust is natural, as Hume suggests, and sensible, for we seem to have no alternative basis upon which to make decisions about how to act. It is not inconsistent with possession of a general disposition to think and act as if there is a world outside one's own mind for one to worry that one does not actually have good reason for doing so, from time to time. Moreover, such worry can be found *both* with the layperson who finds herself momentarily provoked by such worry, *and* with the serious philosopher.

## **1.2 Scepticism, Appearance, and Reality in Literature**

Evidence that laypeople share these moments of worry is not merely anecdotal. It is evident in the continued popularity of films like *Inception* (Nolan 2010), *The Matrix* (Wachowski and Wachowski 1999), *eXistenZ* (Cronenberg 1999), *The Truman Show* (Weir 1998) etc., which capture the popular imagination by offering us literal presentations of sceptical hypotheses, asking how we know that we are not currently dreaming or that our experiences are not being fabricated by a computer or an evil scientist. Moreover, concern with the relationship of appearance to reality is a recurring theme in literature, and especially in film and theatre, where the audience is asked to 'suspend disbelief' and accept as true the fictions that play before them, even as each viewer must know that before them are actors merely playing at being the characters they present, effects that merely represent the events the writer and director sought to evoke.

It is no coincidence that the phrase 'Cartesian Theatre' has come to represent the theory of mind in which mental images are displayed to a watching subject, as the theatre provides a familiar example of a person who willingly takes themselves to be absorbed in something they know to be a mere imitation of reality. (Although, one should not suppose from this that having a concern about discrepancies between appearance and reality entails acceptance of the Cartesian Theatre model of mind.)

The phrase ‘willing suspension of disbelief’ originally comes from Coleridge (2004) and refers to an act of ‘poetic faith’ in relation to lyrical ballads. It is today much more familiar in the discussion of film and theatre, often quoted in discussions of theatre practitioner Stanislavski’s naturalism<sup>4</sup>, which advocates rendering ‘what ifs’ as ‘as ifs’, see, for example, Bailey (2011: 139). Stanislavski (1961: 247) describes ‘the great difference between nature and postcards... between a thought image and physical contact’, and yet he also asserts ‘in every psychological act there is a great deal of the physical’ (Stanislavski (1961: 225)), indicating that although he marks the difference between mental presentation and physical reality, he also sees the two as inextricably linked. One who practices Stanislavski’s naturalism (which has become the norm for most of our film and television) is in a constant attempt to reduce the barrier between appearance and reality.

Yet, the drive to resist absorbing oneself within illusion as though it were reality also has a respectable theatrical tradition. The practitioner, Brecht, rejects such ‘naturalism’ as not only failing to bridge the gap between the world of the play and the real world, but as allowing for deceit and manipulation of an audience that should be alert and critical, alienated from the text by a variety of distancing techniques (*Verfremdungseffekt*), conscious that the play is a mere representation, allowing for analysis and judgement (see Brecht 1961).

Always in the theatre is this question of whether the artifice should be allowed to interfere with the absorption; whether the audience can ever really be a part of the world of the play. The ‘play within the play’ is often used to confront the audience with their own relationship to illusion. *Hamlet* is a particularly striking example. Its ostensive purpose within the play is to echo the real events of Hamlet’s father’s murder by his uncle, prompting Claudius, ‘frighted by false fire’ to reveal his guilt (Shakespeare 1997, *Hamlet*: III.ii). However, its tropes of illusion and reality are enmeshed in the play’s themes of trust and deceit; appearance and reality; existential doubt; and death, the unknown country. Hamlet’s metaphor, ‘frighted by false fire’, underscores how we can be prompted to react to illusion just as we would to veridical experience. Hamlet’s great soliloquies are exquisite explorations of human doubt, as Hamlet hesitates to act when he does not know what to believe. While few of us are faced with the choice of whether to believe the word of a ghost, or challenged to avenge

---

<sup>4</sup> ‘Naturalism’ here refers to the theatrical practice of creating a performance that is as similar to reality as possible, and should not be confused with philosophical naturalism.

our fathers, the agony of acting on imperfect information is a fundamental human experience that has kept the interest of audiences through centuries and across continents<sup>5</sup>.

Fear of madness and an inability to tell appearance from reality is also a classic theme of horror, and one upon which Hitchcock drew heavily. His film, *Vertigo* (Hitchcock, 1958), is named for a dizzying experience produced by a distortion of the senses and deals directly with the unsettlingly blurred line between sanity, dreams, nightmares, and madness: the inability to tell appearance from reality.

In this way we can see that even films and plays that do not deal directly with sceptical hypotheses show a repeated concern with our inability to tell whether the world matches our beliefs about it. The fact that this worry is used by masters of their crafts in some of their most seminal works to evoke memorable scenes of existential angst (*Hamlet*) and psychological horror (*Vertigo*) should suggest to us that something central to the human psyche is being tapped into, and not merely an esoteric doubt restricted to the overly philosophical thinker. The evidence of such sustained concern shows that scepticism is not a ‘merely’ philosophical worry of little relevance to ‘ordinary’ people.

On the other hand, it may be tempting to think that such evidence is no reason to think that the *philosopher* should be concerned by such things. After all, a film like *Inception* is not really a good exploration of the dreaming argument. It might be thought that the ‘common man’ does not seriously worry about such things, but rather enjoys titillating his faculties in the same way that he enjoys watching a horror film, which provokes fear, but from which he knows he is never really in any danger. Perhaps this is true for some, although in general it seems to me a rather patronising view of the non-philosopher (or rather, the untrained philosopher, as I am inclined to think that all people engage in philosophy with varying degrees of rigour). Nevertheless, it would be well to demonstrate that there has been sustained *philosophical* concern for scepticism, too.

---

<sup>5</sup> Whilst an exact figure for the number of translations of *Hamlet* is hard to come by, its cross-culture popularity is remarkable, with translations including Arabic, Japanese, Romanian, Yiddish, Zulu, Korean, and Indian languages, as well as Esperanto and even Klingon, suggesting that we project an enduring relevance for the play and its themes into our future as well as relating it to the present. (See Litvin 2011, Anzai 1999, Cinpoeðs 2010, Berkowitz 2002, Huang 2011, and Shakespeare 2000)

### 1.3 A Potted History of Philosophical Scepticism

Scepticism has a long history of being taken quite seriously by very serious thinkers, and an almost equally long history of those thinkers being scoffed at for doing so. Thus Hobbes writes in his first objection to Descartes:

But since Plato and other ancient philosophers discussed this uncertainty relating to the objects of the senses, and since the difficulty of distinguishing the waking state from dreams is common knowledge, I am sorry that Descartes, who is an outstanding original thinker, should be publishing this old stuff. (Hobbes in Bennett 2007: 42)

It is tempting to regard outlying individuals as exceptions, like Peter Unger (see his 2003 *Ignorance: A Case for Scepticism*), who stands in near isolation<sup>6</sup> in the modern discourse, arguing that no one knows anything or can even be said to reasonably believe anything (Unger 2003: 1). It is easy to dismiss an individual as performing some error that leads them to believe themselves a sceptic when they are not *really*. Thus, C H Whiteley (1975: 490) chides Unger for insisting on a falsely absolutist standard for mental states like ‘knowing’ and ‘believing’. Unger argues that we must devise alternate terms for less absolute states, but Whiteley replies: ‘Don’t the required non-absolute modes of speech already exist?’ (*ibid.*).

It is easy to tell one man his attempt to revive scepticism is merely a redefinition of terms—that scepticism is not really worrying as we never demanded certainty for knowledge in the first place. It is often said that the illusory nature of scepticism is evident in the fact that no sane person could be said to believe it. Thus, Whiteley (1975: 489) asks of Unger ‘How can anyone sanely hold these views?’. I would turn this on its head and ask: if Unger sanely holds these views, can we still judge them mad? Or unholdable by the sane? Ah, but Unger is not *really* a sceptic, it will be said, he *does* believe in knowledge as we commonly understand it, he has just been illicitly redefining knowledge as something much stricter than what we commonly mean.

Let us call this the ‘Argument from Sanity’, made explicit, it runs like this:

1. A sane person is able to live successfully

---

<sup>6</sup> Bryan Frances has also made a foray into scepticism in his 2005, *Scepticism Comes Alive*; I could find no reviews or responses to this, so have focused on the discussion of Unger.



2. Living successfully involves making judgements, and taking action based on these judgements, to preserve one's life<sup>7</sup>
3. One could not make and act on judgements that preserved life unless one takes at least some of our beliefs to be justified and true<sup>89</sup>
4. Therefore, in order to live successfully, I must take myself to have at least some justified, true beliefs (arguably quite a few)
5. Therefore, in order to be sane, I must take myself to have at least some justified, true beliefs
6. Therefore, anyone who does not take themselves to have at least some justified, true beliefs is not sane
7. The sceptic denies that any of her beliefs are justified
8. Therefore, the sceptic cannot be sane
9. I live successfully
10. Therefore, I am sane
11. Therefore, I am not a sceptic

It is worth noting that either living successfully *or* being sane suffices as proof of not being a sceptic on this argument. I think often it is tacitly supposed that one cannot live successfully if one is not sane, but this is not the case. One in four people live with some kind of mental illness (McManus et al 2009), yet only 23,531 people (Care Quality Commission 2015: 34) were detained under section three of the UK Mental Health Act, as being a danger to themselves or others, in 2013–14, suggesting that far fewer than one in four people are unable to live their lives successfully due to mental illness (0.0004% of the population). Obviously the criteria for being sectioned under the Mental Health Act are not the criteria I have given here for unsuccessful living, so the point is not as powerful as it could be, but I think this simply underscores how unhelpful it is for philosophers to dismiss each others' opinions in terms of mental illness. I would suggest that the philosophical habit of dismissing positions as impossible to sanely hold is a vague, misleading, and unproductive practice as a whole.

---

<sup>7</sup> For added impact, one might add further conditions, such as that the actions I take should preserve my comfort and promote my happiness.

<sup>8</sup> If we are good Humeans we might say that the actions of a successful life are motivated by the desire to preserve one's life and the belief that an action will preserve one's life; the desire for the action to facilitate successful life will make us desire to have beliefs that are justified and true.

<sup>9</sup> I am running this argument on justified, true belief, rather than knowledge, as that is the minimal claim. If the argument works for justified, true belief, it should work for knowledge if the requirements for knowledge include true, justified belief. However, the definition of knowledge is contentious, and depending on one's take on 'justification' a true belief might be justified and not count as knowledge.

Let us limit ourselves to talking of living one's life successfully—I imagine that Unger's apparent success is a significant reason that Whiteley assumed him to be sane. In this way one may run Whiteley's criticisms of Unger:

1. Living successfully involves making judgements, and taking action based on these judgements, to preserve one's life
2. One could not make and act on judgements that preserved life unless one takes at least some of our beliefs to be justified and true
3. Therefore, in order to live successfully, I must take myself to have at least some justified, true beliefs (arguably quite a few)
4. The sceptic denies that any of her beliefs are justified
5. Therefore, the sceptic cannot live successfully
6. Unger is not only preserving his life but publishing books and pursuing a career in philosophy at a leading university
7. Therefore, Unger is living successfully
8. Therefore, Unger is not really a sceptic

Premise 2 appears to be the key, but it is by no means obviously true. Much depends on what it means for a belief to be taken as justified and true—which is the point of dispute between Whiteley and Unger—but it is also not obvious that one needs to take beliefs to be justified and true (whatever that means) in order to act upon the judgements that preserve life. One might *make* judgements and *act on them* on the basis that certain beliefs are justified and true whilst harbouring some doubt that they really are. Sometimes I judge that the milk has not turned and put it in my tea even though I am not completely sure I am justified in doing so—I have a cold and I am not sure the sniff I gave the milk is enough to determine whether it was bad or not.

Moreover, numerous philosophers throughout history have urged us to avoid making judgements altogether, as is shown below. It has not always been the case that anyone who claimed to believe the assertions of scepticism was deemed either mad or insincere. Many seem to have lived successfully whilst avoiding judgements, or at least *taking themselves* to not be justified in committing to any firm belief.

In the next few sections I will chart a history of scepticism that reflects persistent doubts about the reliability of our senses; serious philosophers who advocate suspension of judgement and avoidance of knowledge claims; and other, equally serious philosophers who have found these doubts about our ability to have knowledge of this kind deserving of a substantial direct response, and not a mere dismissal.

### 1.3.4 Sixth Century BC—Jainist<sup>10</sup> Epistemology

In the sixth century BC, Jainist epistemology developed as a theory of qualified assertion in order to fit the requirements of Jainist metaphysics. As Koller writes: ‘[B]ecause reality is many-sided and cognition true only from a limited perspective, all ordinary knowledge claims need to be qualified... by adding the *syāt* qualifier’ (Koller 2000: 400–1). In this way the metaphysics of *anekāntavāda* (which holds that existence is many-sided) is rendered compatible with everyday talk via the epistemic theory of qualified assertion (*syādvāda*). Because humans, with our limited perspective, can never achieve full knowledge, the Jainist adds a caveat, ‘*syāt*’ (meaning ‘in some ways’ or ‘from a certain perspective’) to everyday assertions. Whilst this incompleteness should not be simply equated with, say, Cartesian doubt, it does show that Jainists thought seriously about the limitations of knowledge and the fact that what we can know is limited by our perspective.

### 1.3.2 Fifth Century BC—Socratic Scepticism

Socratic scepticism is described in Plato’s *Apology*, where Socrates supposes the oracle at Delphi to have named him the wisest man in the world because he ‘understands that his wisdom is worthless’ (Plato, *Apology*, 23b<sup>11</sup>). Socrates contrasts himself, in Plato’s dialogue, with skilled artisans and politicians. He initially considers himself less knowledgeable than these men because he does not know that which they claim to know. Yet he finds, on questioning them, that their reasons for the knowledge they claim are insufficient. The wisdom of Socrates thus lies in him knowing that he does not know these things: ‘About myself I knew that I know nothing’ (22d).

We can understand Socratic scepticism as an approach to the quest for knowledge, a method of withholding the claim to knowledge so as to avoid making assumptions and encouraging further questioning. Whilst knowledge is not explicitly ruled out, it is made clear

---

<sup>10</sup> Jainism is generally thought to have been founded by Mahāvīra, a sixth century BC contemporary of the Buddha. See Kumar (2001: 5).

<sup>11</sup> References are to the 2002 Grube and Cooper edition.

that *living* with a complete lack of knowledge, and preserving a state of doubt, is perfectly possible and, indeed, to be advocated for philosophical endeavour.

### **1.3.3 Fourth or Sixth Century BC—Laozi<sup>12</sup>**

At a similar time in China Laozi writes: ‘To know and yet (think) we do not know is the highest (attainment); not to know (and yet think) we do know is a disease’ Laozi<sup>13</sup> (1891: 71). Scepticism is advocated as a cure for a disease, the disease from which we live in pain is in thinking that we know that which we do not. To be diseased is the norm; the sage who knows he does not know is the exception, his status a goal to be obtained. Here, the disquiet that prevents one from living a good life lies not in doubt, but in baseless knowledge claims. As with Socrates, this is a warning against dogmatism. The recommendation is that the philosopher should risk not recognising his own knowledge by promoting doubt, rather than insisting upon knowledge that is not justified.

In case the reader fears there may be a contradiction between this supposition that the untrained mind is dogmatic and my assertion that she is prone to sceptical doubt, I would urge you to dismiss it. Firstly, it is very possible to be dogmatic at some times and sceptical at others. One should not be surprised that contradictory attitudes can be found in untrained minds (or in trained minds, for that matter; we are none of us perfectly consistent reasoners). Secondly, to experience doubt does not prevent one later coming to a conclusion and becoming dogmatic about it. Indeed, we may suppose that one may prescribe training to prevent the adoption of dogmatic positions in response to the spectre of sceptical doubt. Thirdly, as I have already noted, though I think many of us do experience periods of (sometimes quite intense) sceptical doubt, I did not suppose that this is often a sustained position, rather, I suggested that doubt primarily occurs against a background of trust. For many, firm beliefs—in political ideologies, in the affection of one’s partner, and so forth, in addition to more foundational beliefs—undergo moments of doubt before returning to their

---

<sup>12</sup> Also rendered ‘Lao-tsu’.

<sup>13</sup> The edition cited is an 1891 translation, which assumed a single author. Tradition dates the text to a single author in sixth century BC (see Loewe and Shaughnessy 1999: 746), but the date of the original text is unknown and there is some suggestion that the Laozi is actually a collection of texts gathered over a period of time between 340 and 249 BC (see Brooks and Brooks 1998: 151).

former assurance. We should not be surprised that sceptical philosophers have found dogmatism amongst untrained minds despite a human tendency towards at least occasional scepticism.

Moreover, to suppose that the trained and untrained form completely separate camps is not only misguided, but a somewhat pessimistic view of the power to philosophy to affect everyday lives. We should not dismiss lightly the way in which the Socratic claim that ‘the wisest man in the world is the one who knows that he knows nothing’ has struck a chord across the centuries. Versions of this thought are repeated throughout literature. The wise fool is a common trope in literature, thus, Touchstone says in *As You Like It*: ‘I do now remember a saying: “The fool doth think he is wise, but the wise man knows himself to be a fool”’ (Shakespeare 1997, AYLI, V.i). The intertwining of popular and philosophical scepticism is symptomatic of the longevity of its interest for human beings.

### **1.3.4 Fourth Century BC to Third Century AD<sup>14</sup>—Pyrrhonian Scepticism**

Pyrrhonian sceptics found tranquillity in their scepticism. The arguments of Pyrrhonian sceptics like Sextus Empiricus are constructed by presenting appearances and judgements alongside antithetical counterparts to provoke precisely the suspension of judgement that most modern epistemologists are so concerned to avoid. The suspension of judgement that is deemed impossible for the sane.

It seems the Pyrrhonians found ways to live successfully without knowledge, or at the very least believing that they did not have knowledge. From our earliest philosophy it has been argued that a full life is not wedded to the *possession* of knowledge, or to believing that one possesses knowledge, or that one should be able to *assert* knowledge. Indeed, some argue that life is fuller and more satisfying where no knowledge claims are made.

One might think that this argues against my purpose. The Pyrrhonians claimed to be at peace with (or at least sought to attain peace with) their lack of knowledge. However, it is clear, nonetheless, that knowledge and lack thereof was an issue that Pyrrhonians struggled with and viewed as centrally important. Their arguments sought to break us out of the very

---

<sup>14</sup> Pyrrho is thought to have lived from 365/60–275/70 BC (Svavarsson 2010: 36), his thought inspired the founding of the Pyrrhonian school by Aenesidemus in the first century BC (Bett 2010: 5); the most complete account of Pyrrhonian scepticism is found in the works of Sextus Empiricus, in the second and third centuries AD (Pellegrin 2010: 120)

*natural* inclination to seek after the knowledge they believed impossible to attain. Unlike the modern movement that seeks to convince us that ‘ordinary’, non-philosophically trained folk are content to believe that they have a wide variety of knowledge, the whole Pyrrhonian philosophy is premised on the thought that people are commonly involved in a fruitless struggle for knowledge that *philosophy* can wrest us free from.

### 1.3.5 Third to First Century BC—Academic Sceptics

Academic sceptics like Cicero argued that nothing can be known, as we cannot distinguish false presentation from true (Cicero, *Academica*, II, xxiii, 76-II, xxviii, 90<sup>15</sup>), and reason (or dialectic) cannot provide certainty (Cicero, *Academica*, II, xxviii, 90-xxx, 98). Yet, unlike the Pyrrhonian sceptic, Cicero argues that:

[E]ven many sense percepts must be deemed probable if only it be held in mind that no sense presentation has such a character as a false presentation without differing from it at all. Thus the wise man will make use of whatever apparently probable presentation he encounters... if he were not to approve them life would be done away with... (Cicero, *Academica*, II, xxxi, 99)

Thus, the academic sceptic takes something akin to the modern pragmatist position (although, he does not take the pragmatist position of dismissing the sceptic). Cicero is adamant that sceptical arguments are effective and nothing can be known, and that it is important to recognise this to avoid dogmatism: ‘the wise man will never assent’ (Cicero, *Academica*, II, xxiv, 78), he will merely form probable beliefs. This concern to avoid assertions echoes the Jainist argument that one should not assert ‘X is Y’. Advocating probability is notably different to advocating the perspectival caveat of *syāt*, but both seem motivated by an acknowledgement of the limitation of our ability to know absolutes from the subjective perspective.

Acting on probable beliefs without ever claiming knowledge does not seem to have caused any problems for Cicero. One might, as with Unger, call this merely a refusal to use the term ‘knowledge’ in the way that it is commonly employed with regard to non-certain, but highly probable, beliefs. Nevertheless, if our concern is to reflect genuine language usage, it is

---

<sup>15</sup> References are to the 1933 Rackham translation.

clear that from ancient times there has been a consistent current of thought that shows individuals as unwilling to ascribe the term ‘knowledge’ to their probable beliefs.

### **1.3.6 Fourth Century AD—Augustine**

As noted by Klima, Allhoff, and Jayprakash Vaidya (2007: 1) there is still a tendency to assume that nothing much of philosophical interest happened between ancient times and Descartes—that scepticism, along with other philosophical issues, ceased to be of interest or subject to substantial thought. This is false. For example, Augustine (*Confessions* 5.10.19<sup>16</sup>) confesses that he briefly fell sway to scepticism in 383–4 AD<sup>17</sup>: ‘The thought came to my mind that the philosophers whom they call Academics were shrewder than others. They taught that everything is a matter of doubt, and that an understanding of truth lies beyond human capacity’. Yet, within a relatively short period of time (386–7 AD) he rejected the scepticism he had embraced and attacked it in his *Contra Academicos*. Arguing that happiness lies in wisdom and that Academic sceptics are not wise.

Augustine’s argument rests on a questionable reading of Cicero to the effect that wisdom and happiness might be found in the search for truth, even if there can be no knowledge: “‘why can’t someone who searches for truth be happy, even if he cannot find it?’” (*Contra Academicos* 1.3.7<sup>18</sup>). Augustine’s attack takes a fruitless search for knowledge to be inducing despair as opposed to tranquillity, powerfully feeling that a defence of the attainability of knowledge was required.

### **1.3.7 Medieval Scepticism**

In c. 1100, Al-Ghazali, an Islamic philosopher and theologian, wrote *Deliverance from Error*. In it he outlines a project that will sound familiar to readers of Descartes’s meditations. He writes:

---

<sup>16</sup> References are to the 1991 Chadwick translation.

<sup>17</sup> Date estimates comes from O’Daly (2001: 159)

<sup>18</sup> References are to the 1995 *Against the Academics* translation by King.

My aim is to perceive the deep reality of things; I wish to seize the essence of knowledge. Certain knowledge is that in which the thing known reveals itself without leaving any room for doubt or any possibility of error or illusion, nor can the heart allow such a possibility. One must be protected from error... However, when I examined what I know, I found myself lacking this kind of certain knowledge (Al-Ghazali 2002: §§25–8)

Compare:

Some years ago I was struck by the large number of falsehoods that I had accepted as true... Reason now leads me to think that I should hold back my assent from opinions which are not completely certain and indubitable (Descartes *Meditations*: I.18–19)

Descartes's standard of certainty—immunity from doubt—is often cast by modern philosophers as a peculiarity of Cartesian thought that has infected our discourse and from which we must be cleansed. Yet, Al-Ghazali's position that 'perceiving the deep reality of things' is dependent on leaving no 'room for doubt' and being 'protected from error' predates the *Meditations* by five hundred years.

Both texts move on from this starting point to question first the reliability of the senses, and then 'intellectual truths' (Al-Ghazali 2002: §31) or 'simpler and more universal things' (Descartes *Meditations*: I.20), like the truths of mathematics or geometry. Their conclusions are very different. Where Descartes hits on the *cogito* as a first principle, from which he believes all knowledge can be derived, Al-Ghazali is unable to establish a first principle and concludes the reverse, that 'no construction of a proof is possible' (Al-Ghazali 2002: §34). Al-Ghazali dwelt for two months as a sceptic, 'though neither in theory nor in outward expression' (*ibid.*) before he was 'cured' by God's revelation:

I recovered my health and mental equilibrium. The self-evident principles of reason again seemed acceptable; I trusted them and in them felt safe and certain. I reached this point not by well-ordered or methodical argument, but by means of a light God the Almighty cast into my breast, which light is the key to most knowledge (Al-Ghazali 2002: §§ 34–5)

It is tempting to see in this a religious version of Hume's naturalism. Where Al-Ghazali's return to trust in self-evident principles is attributed to God, Hume attributes this to human nature. However, where Hume's naturalism argues for contentment with less than certainty,



there is no such concession in Al-Ghazali. Rather, I would argue that we see the proto-Humean element in Al-Ghazali's statement that his theory and outward expression did not reflect his inner doubt: i.e. Al-Ghazali continued to think and act as though he believed, even though he was philosophically consumed by doubt. Al-Ghazali's restoration to trust in 'self-evident' principles does not mark an acceptance that certainty is not required for knowledge. Rather, it lies in the conclusion that certainty about deep reality is founded on faith in God.

The exclusion of doubt as a requirement for knowledge was also advocated by Henry of Ghent<sup>19</sup>, who, drawing on Aristotelian principles, held that 'there can be knowledge only by apprehending what is certain and determinate' (*Summa Quaestionum Ordinarium*, a.1 q.1, 1.3<sup>20</sup>). Given this, Ghent constructs an argument that 'a Human Being Cannot Know Anything' (*Summa Quaestionum Ordinarium*, a.1 q.1, 1), on the basis of the untrustworthiness of the senses, 'The senses apprehend nothing of what is certain concerning a thing' (*Summa Quaestionum Ordinarium*, a.1 q.1, 1.3); combined with our complete reliance upon them for knowing contingent objects, 'All human intellectual cognition has its knowledge in the senses' (Ghent, *Summa Quaestionum Ordinarium*, a.1 q.1, 1.2).

#### **1.4 Scepticism as an Enduring Concern**

The above is a truncated and simplified account of scepticism through the ages, but what it reveals are concerns that have persisted across time, in diverse cultures, and which existed long before Descartes and the Cartesian doubt. We see that philosophers such as the Jainists, Socrates<sup>21</sup>, Laozi, and Cicero not only did not claim knowledge, or 'full' knowledge, but advised against it. The Pyrrhonians advocated a way of life founded upon accepting the impossibility of knowledge, and other philosophers, such as Henry of Ghent, Augustine, and Al-Ghazali, seem to have lived under sustained doubt, or belief in the impossibility of knowledge, for at least some period of time. Moreover, we see again and again arguments for doubt motivated by the uncertainty of the senses alongside their central role in our knowledge claims. 'All human intellectual cognition has its knowledge in the senses' but, 'The senses

---

<sup>19</sup> 1217–93 AD (Klima, Allhoff, and Jayprakash Vaidya 2007).

<sup>20</sup> References are to the Pasnau 2002 translation, reprinted in Klima, Allhoff, and Jayprakesh Vaidya 2007.

<sup>21</sup> Or Plato in the guise of Socrates, I take no position on this point.

apprehend nothing of what is certain concerning a thing' (Ghent, *Summa Quaestionum Ordinarium*, a.1 q.1, 1.2–3). Cicero argues that we cannot distinguish false presentation from true (Cicero, *Academica*, II, xxiii, 76-II, xxviii, 90), and Sextus Empiricus articulates extensively the conflicting deliverances of the senses in the Ten Modes, particularly modes one to seven (see *Outlines of Scepticism*, I.xiv<sup>22</sup>). More loosely, there is Jainist thought that the human perspective is fundamentally limited.

From some of our earliest recorded philosophy we see the same concern that there is, or might be, a mismatch between how the world seems to us and how it really is. The responses to this concern differ. Some believe that knowledge, though difficult to attain, is not out of reach. Some philosophers find contentment in accepting the problem as irresolvable. Some abandon knowledge for probability. Some turn to a higher power to guarantee our knowledge. However, although the responses may differ, the worry that appearance may not match reality, and that our senses may not be a reliable source of information, is a clear undercurrent through time.

Moreover, we see that the concern that that which we perceive may not match reality is recurrent in literature, which reflects the cultural interests of the mind untrained by philosophy. We question our senses. We question what evidence suffices for knowledge. We are disquieted and fascinated by illusion and we unsettle ourselves by exploring sceptical hypotheses in our literature.

It would be wrong to say that only one sceptical interest arises from this milieu of doubt, but I think my sceptic, the sceptic who is concerned that she might have just the thoughts and experiences she has now if the world were radically different—even if there were no world outside of her own mind at all—articulates a prominent thread. I think she demands a robust, direct answer to her concern, and I think that such an answer would offer substantial reassurance to those who have been plagued by this doubt through the ages.

## **1.5 Transcendental Arguments as a Response to Scepticism**

I want to give the sceptic a robust answer. I think her questions reflect valid and genuine worries and deserve a serious response. By 'robust' and 'serious', I mean that I want to give her an answer that acknowledges the common ground between us. Rather than dismissing her

---

<sup>22</sup> References are to the 2005 Julia Annas and Jonathan Barnes edition.

worry as misguided or nonsensical, I want to accept her framing of the argument, but show that her conclusion is false. I take her argument to be:

1. I have a complex mental life that includes experience which seems to present an extra-mental world
2. I could have *just these* thoughts and experiences in sceptical scenario S where there is no extra-mental world
3. I cannot get outside my own perspective—the only evidence I have to justify my beliefs are my own thoughts and experiences
4. Therefore, I have *just as much* justification to believe that there is an external world that is much as it appears as to be as I have to believe that I am in S
5. Therefore, I do not know that I am not in S

In saying that I take her argument seriously, I mean that I accept her standards for evidence and do not take her worry to be unintelligible. A serious and robust response must therefore be one that starts from the evidence she accepts: the thoughts and experience that make up my complex mental life which seems to present an extra-mental world. It is therefore my challenge to show that there is not '*just as much*' justification to believe the sceptical scenario is could be the cause of this complex mental life as there is to believe that the world is much as it seems.

The key to answering this type of sceptic is to only use premises the sceptic would accept. She is not a sceptic content to simply keep asking 'Yes, but why?' in the manner of a small child. She is principled. She is ourselves when we contemplate our own thoughts and experiences and worry that there might not be anything else. In particular, she is concerned that there seems to be a discrepancy between her assurance for her own thoughts and experiences and her beliefs about the world outside those thoughts and experiences.

Transcendental arguments are particularly suited to the purpose of offering a robust and serious response to this type of sceptic. There is more than one type of transcendental argument, and I will look in more detail at what exactly it is for an argument to be transcendental in chapter two, but as a generalisation, we can take an argument to be transcendental if it starts from a premise about our thoughts or experience, and moves on to show that something must be the case because it is a necessary condition of our having thoughts or experience, or of our having a certain type of thoughts or experience. Given that the transcendental argument starts from a premise concerning precisely the type of proposition that the sceptic *does not doubt*—indeed, a type of proposition about which she has such

assurance that she holds in question anything for which she does not have equal assurance—if it can be shown that that which she doubts is a *necessary condition* of this, she will have to abandon her doubt.

I distinguish two broad categories of transcendental argument: strong and weak. A strong transcendental argument is one that shows that the truth of some proposition concerning the mental has as a necessary condition a proposition concerning the extra-mental. A weak transcendental argument is one that focuses on elucidating the necessary connections between mental concepts, the conclusion of which will concern what we should believe to be the case. In recent years, the preference has overwhelmingly been to focus on weak transcendental arguments. I argue that no weak transcendental argument can offer reassurance against my sceptic, and that therefore I will go on to defend the plausibility of strong transcendental arguments.

It may be wondered why the sceptic prizes propositions concerning her thoughts and experience. In recent years, some have come to question whether the same sceptical scenarios that lead us to doubt the extra-mental might not also give us cause to doubt the mental. In chapter three, I examine these arguments. I argue that although in some respects it must be conceded that we do not know our own minds perfectly, not all propositions are brought into doubt; moreover, the type of propositions brought into question (e.g. our knowledge of our own emotions) are not of the sort that the sceptic was inclined to prize in the first place.

Having secured that there are at least some propositions concerning the mental to form the first premise of a transcendental argument, in chapter four, I go on to examine the criticisms that have been raised against the plausibility of strong transcendental arguments. In particular, I consider Stroud's challenge that we cannot build a 'bridge of necessity' from psychological premises to conclusions about the world. I argue that Stroud's criticisms rest on the illicit supposition that the mind is not a part of the world. I show that supervenience physicalism is a viable metaphysical picture, and one which would explain how propositions concerning the mental can have the truth of extra-mental propositions as a condition of *their* truth. I argue that no metaphysical picture of the relationship between mind and world that could generate the restrictions on reasoning that lead to the impossibility of reasoning from premises about the mind to conclusions about the extra-mental world can be assumed. Given that the sceptic cannot rule out the supervenience physicalist picture of mind and world, she

cannot simply assert that the world is such that reasoning from premises concerning the mind to conclusions concerning the world is implausible.

In chapter five, I explore what form the second premise of a strong transcendental argument could take and propose that it should be a psychophysical law of nature, and that a dispositional essentialist view of properties would guarantee the metaphysical necessity of such a law so that if it could be found it would be true at all possible worlds where the mental property existed. I do not give a strong transcendental argument myself, but I show how one could be constructed on the basis that some mental proposition prized by the sceptic will only be true at worlds where some extra-mental proposition is also true. This will be the case if there is a metaphysically necessary psychophysical law, for the truth of dispositional essentialism will mean that any world at which the prized mental property is present will be a world at which it is in a stimulus and/or manifestation relation to certain physical properties. Which is to say that *it could not be the case* that the truth of propositions concerning the mental life we do have could be equally well explained if we were subject to a sceptical scenario in which there was not physical world.

A full defence of dispositional essentialism and watertight argument for a specific strong transcendental argument would be a much larger project than there is space for in this thesis. However, I will show that it is entirely plausible that such an argument could be given. Moreover, I show that any response from the sceptic will involve her taking up and defending specific philosophical positions, such as nominalism about the laws of nature or a Cartesian dualist picture of the relationship between mind and world. This, in itself, is no small achievement. As such, the sceptic becomes not merely the natural, recurrent human worry I have outlined above, but a concrete position with theses to defend. This is a shift in the dialectic that can itself offer comfort. We can know something of what must be true of the world for the sceptic's arguments to get started.

## Chapter 2: Transcendental Arguments

It is my contention that transcendental arguments, and especially strong transcendental arguments, are particularly suited to responding to the sceptic. To understand why this is, we must first understand what a transcendental argument is. The purpose of this chapter is to articulate what is meant by the term ‘transcendental argument’ broadly, and to provide a finer account of different types of transcendental argument, understanding their similarities and differences as they have arisen through their historical context.

Transcendental arguments are various in their form and the term is best understood as a kind of family resemblance concept, for which Kant’s transcendental deduction should be regarded as a historical anchor and point of origin. It would not be right to say that the Kantian project provides a paradigm, as the development of transcendental arguments has progressed as often by misunderstanding Kant as it has by direct inspiration. Nevertheless, the name is owed to Kant, and the branches of modern discourse that I am particularly interested in trace their origins to his work.

I distinguish transcendental philosophy as an epistemological theory of what can be known and of the structure of knowledge, from a transcendental argument, being an argument structured and defined by a specific methodological approach. The distinction between the two is not always carefully maintained in the literature, and this is to a certain extent due to a failure to make the distinction in reading Kant. I shall therefore begin this chapter by providing an account of Kant’s transcendental project and what we can learn from it in order to develop a satisfactory response to scepticism.

### 2.1 The Transcendental Deduction

The name, ‘Transcendental Deduction’, has often misled scholars into conflating Kant’s transcendental philosophy with we would now call ‘transcendental argument’. There *are* transcendental arguments within the *Critique of Pure Reason*, but ‘Deduction’, here, is not being used in our modern sense to mean ‘proof’. Rather, Kant offers us a defence of certain indispensable propositions in the manner described in ‘The Principles to any Transcendental Deduction’ by analogy to an eighteenth century Prussian legal practice, whereby the rights to a

property are established by ‘deduction’ in the form of tracing the claim back to its origin.<sup>23</sup> The ‘Transcendental Deduction’, therefore, does not try to *prove* or *justify* any particular claim in the modern sense of ‘deduction’, in which one offers an argument whereby the warrant of premises transfers by entailment to a conclusion. Rather, it seeks to elucidate where our concepts come from. The following is worth quoting at length to understand the full meaning:

When teachers of law talk about rights and claims, they distinguish in a legal action the question regarding what is legal (*quid iuris*) from the question concerning fact (*quid*), and they demand proof of both. The first proof, which is to establish the right, or for that matter the legal entitlement, they call the deduction [This term also applies in philosophy] We employ a multitude of empirical concepts without being challenged by anyone. And [...] neither from experience, nor from reason can we adduce any distinct legal basis from which the right to use them emerges distinctly.

But there are [...] some that are determined for pure a priori use [...] proofs based on experience are insufficient to establish the legitimacy of using them in that way; yet we do need to know how these concepts can refer to objects even though they do not take these objects from any experience. Hence when I explain in what way concepts can refer to objects a priori, I call that explanation the *transcendental deduction* of these concepts. (Kant, A 84–5, B 116–17; comment in square brackets is the translator’s, ellipses are mine.)

He starts from the *supposition* that certain claims constitutive to our understanding of the world are known to hold, and then sets about providing the ‘justification’—not in our modern sense of providing the reasons to believe, but rather, having noted that we *do* believe he seeks to show why we are right to do so.

When Stroud first characterised transcendental arguments in his influential 1968 paper, ‘Transcendental Arguments’, his account was skewed by a failure to distinguish the Transcendental Deduction and the Refutation of Idealism as discrete projects within Kant’s Transcendental Aesthetic. Stroud writes:

The transcendental deduction (along with the Refutation of Idealism) is supposed to provide just such a proof [against scepticism] and, thereby, to give a complete answer to the sceptic about the existence of things outside us. (Stroud 2000a: 9–10)

There is no doubt that Kant considered scepticism as deserving of a response. He felt that it was: ‘a scandal to philosophy and to human reason in general that the existence of things

---

<sup>23</sup> See Kitcher’s ‘Introduction’ to the Pluhar 1996 edition: xlii; also Henrich 1989: 29–46.

outside us [...] must be accepted merely on faith' (B xl). However, the scandal to which Kant refers is not that we do not *know* that things exist outside us. On the contrary, he is convinced that we do know this. The scandal is that we have to take truths such as this on trust, as no one has been able to articulate our justification.

It is in this sense that I understand the Transcendental Deduction to be an articulation of Kant's *transcendental philosophy*, rather than as a transcendental argument. To be transcendental in this sense is not to seek knowledge of things as they are in themselves (independently of us), but rather to note what we *do* know, of the existence of things outside us, and to justify this knowledge claim via an articulation of the necessary connections between the nature of our thoughts and experience and the things we know. This transcendental philosophy is not a *proof*, but an articulation of a metaphysical picture.

## 2.2 Kant's Copernican Turn

The insight that enables Kant's Transcendental Deduction is that which Kant characterised as his 'Copernican turn': a way of solving a fundamental problem via a radical shift in point of view. Just as Copernicus realised that we were failing to understand the movements of the stars because we supposed that they were revolving around the Earth, Kant's realisation is that we have been considering fundamental epistemological questions from the wrong perspective. He writes of the 'turn':

We here propose to do just what Copernicus did in attempting to explain the celestial movements. When he found that he could make no progress by assuming that all the heavenly bodies revolved round the spectator, he reversed the process, and tried the experiment of assuming that the spectator revolved, while the stars remained at rest. We may make the same experiment with regard to the intuition of objects. If the intuition must conform to the nature of the objects, I do not see how we can know anything of them a priori. If, on the other hand, the object conforms to the nature of our faculty of intuition, I can then easily conceive the possibility of such an a priori knowledge. (B xvi)

Kant is proposing that our mistake has been trying to find proof of the existence of things outside us by supposing that our experience (intuition) of them is caused by the objects. As we can thus only know of the objects by that experience, anything we know of them must be a posteriori. Kant supposed that only a priori truths can be necessary, so no argument based on a



posteriori truths can entail a necessary conclusion. In other words: if we suppose that our experience is caused by objects outside us, we cannot have a proof of the existence of objects outside, as that experience can only contain a posteriori knowledge, which cannot be necessarily true, and is thus open to doubt. His ‘Copernican turn’ is to suppose that ‘the object conforms to the nature of our faculty of intuition’—i.e. the nature of the object we experience is in some sense determined by our experience of it. If this is the case, then we can know the nature of objects outside us by considering what objects could be formed by our ‘faculty of intuition’. As we can know about our own minds and experience a priori, we should, Kant supposes, be able to work out what kinds of things we can have knowledge of, and thereby know the nature of the objects our experiences are of.

Kant attributed this awakening from his ‘dogmatic slumber’ to Hume, whom he felt had clarified the problem, but failed to identify the solution it required. Hume’s insight is that there are some propositions that cannot be justified either via the evidence of the senses (for he supposed that there was no necessary connection discoverable by the senses<sup>24</sup> (ECHU 7.8)), or by analytic a priori argument, yet which are still such that we must believe them. For Hume (T 1.4.1.8<sup>25</sup>) it is simply a fact about our nature that we cannot believe the sceptical conclusion: ‘as experience will sufficiently convince anyone [...] that tho’ he can find no error in the foregoing [sceptical] arguments, yet he still continues to believe’. It is this thought of there being certain psychological facts about beliefs that we cannot relinquish that intrigued Kant, but it is in their response to this realisation that Hume and Kant differ.

Hume supposes that imagination draws a connection between one object and another on the basis of an effect’s constant conjunction with its cause, and thus we are able to get on with our lives, supposing that such a connection exists. However, this is a contingent ‘happy’ fact about how we are constituted, about our nature (see T 1.4.1.12). ‘This sceptical doubt,’ he writes (T 1.4.2.57), ‘is a malady that can never be radically cured’. We would be consumed by it, if we did not have some way of setting it aside: ‘Carelessness and in-attention alone can afford us any remedy’ (*ibid.*). I argued in the introduction that the persistence of disquiet in the face of perennial doubt shows that sceptical doubt is just as natural as our careless

---

<sup>24</sup> I will discuss Hume’s sceptical arguments in more detail in 4.2. For now the concern is how Kant responded to Hume’s thought.

<sup>25</sup> References to the Hume (1739–40), *A Treatise of Human Nature*, are given by ‘T’.

inattentiveness to such doubt. Inattention to doubt, after all, does not mean that doubt is not present nor worth attending to. I think it is for this reason that many, like Kant, are not content to be fortuitously, if naturally, inattentive to the doubt. Our anti-sceptical convictions require some justification. Something more substantial is required, and this is what Kant hopes to offer—what he thinks is scandalous that we have not offered already. An explanation of how it is that we can know the things we know.

The Transcendental Deduction is Kant's account of the metaphysics that explains our right to the use of the 'categories' in knowledge claims about objects of experience. These categories are given as unity, plurality, allness (categories of quantity); reality, negation, and limitation (categories of quality); inherence and subsistence, causality and dependence, and community (categories of relation); possibility, existence, and necessity (categories of modality) (see A 80/B 106). Kant does not argue for each of these specifically in the Transcendental Deduction, but, rather, provides an account of our mental processing (synthesis) that accounts for how the application of these concepts (the categories) is necessary for our cognizing of the objects of experience in the way that we do. It is justificatory in that its aim is to show that this is the only possible explanation for the necessary application of these concepts. Its purpose is not to defend that the concepts *are* necessarily applied. It is transcendental because it demonstrates the dependence of objects of experience upon the application of the categories to the manifold of intuition to form an experience that we, as human beings, can understand. Kant's psychological model requires that there be a 'unity of apperception', in which we are able to attribute unity to an object of experience by relating it to ourselves in a way that reflects that oneself is a unified being: 'the possibility of appearances lies in ourselves, and their connection and unity (the presentation of an object) is to be met with merely in us)' (Kant A 130).

It's important to understand this in the context of Kant's transcendental idealism. Objects of experience are not things as they are in themselves, independent of the human mind: '[For] what is merely in us cannot determine the character of an object distinct to our presentations' (Kant A 129). These are objects of the 'phenomenal realm' formed by the application of the categories to the manifold of intuitions. It is the metaphysical supposition that the objects of our experience are not completely independent of our mental faculties and what we are *capable* of experiencing that allows for an account to be given of why they are necessarily applicable.

## 2.3 The Refutation of Idealism

It will no doubt strike the reader that the transcendental philosophy outlined above will not be of great use against my sceptic. This sceptic's worry is precisely that we cannot have knowledge of objects that exist independently of the mind. Moreover, she will not be satisfied with a deduction in the quasi-legal/historical sense that Kant gives where the application of the categories to in space is assumed and thus merely to be explained. Hers is the disquiet that comes upon Hume in his study, though he is able to set it aside in his daily life. To her there is more than one equally plausible explanation for why I have just these thoughts and experiences, and therefore, she insists, we must prove her doubt wrong, or she will not be satisfied.

However, the Transcendental Deduction is not the only kind of transcendental argument articulated in the *Critique of Pure Reason*. Kant also gives us the Refutation of Idealism, which can be understood as a transcendental argument in a way that more closely aligns to how the term is used in modern discussions. I stress that Kant does not draw this distinction himself. Rather, it is a meaningful distinction for us because a) the second kind of argument is more relevant to the needs of my sceptic, and b) this distinct kind of argument that has had a significant impact upon philosophers of the twentieth and twenty-first centuries.

Over the next few sections I will examine the Refutation of Idealism for its form as a transcendental argument, and also for its relevance as a response to my type of sceptic. This second point is particularly significant as a refutation of idealism is *exactly* what she is worried we cannot have, but in order for such a refutation to be given, one must be clear as to what *kind* of idealism one is refuting.

### 2.3.1 Transcendental Idealism and Material Idealisms

Kant added the 'Refutation of Idealism' to *The Critique of Pure Reason* for its second edition. He explains (Kant B 274–5) that he sought to distance himself from what he called 'material idealism'; of which he identified two sorts: 1) problematic idealism, and 2) dogmatic idealism. Kant associates problematic idealism with the philosophy of Descartes, and defines it as the

claim that the existence of external objects in space is doubtful and indemonstrable<sup>26</sup>. He associates dogmatic idealism with Berkeley, and takes it to be the position that the existence of external objects in space is false and impossible. The purpose of the refutation, then, is not simply to show these positions to be false, but to demonstrate the way Kant's own philosophy differs from them.

The theorem to be established as a refutation of these positions is given as '*The mere, but empirically determined, consciousness of my own existence proves the existence of objects in space outside of me*' (Kant B 275).

Kant (*ibid.*) takes Descartes's position to be that the only empirical assertion for which one can have certainty is that of one's own existence (that 'I am'): the empirically determined consciousness of my own experience. Kant accepts Descartes's criterion of not forming a 'decisive judgement' before it is shown that the truth of a proposition cannot be put into doubt, and aims is to show that this singularly indubitable proposition (the claim that 'I am') has as a necessary condition that one has already had some genuine experience of spatial objects external to oneself (the proposition that he takes Descartes to doubt and Berkeley to hold as false).

In the next section I will outline Kant's 'Proof', before going on to show how one's understanding of what 'I am' means and what 'objects in space outside of me' are affects one's understanding of and response to this argument.

### **2.3.2 The Proof**

Kant (B 275–6)'s proof runs as follows:

1. I am conscious of my own existence as determined in time
2. All determination in regards to time presupposes the existence of something permanent in perception
3. Something cannot be that permanent something (which determines my existence in time) if it is within me
4. So, this permanent something must be external to me

---

<sup>26</sup> It is perhaps unfair of Kant to say that for Descartes the existence of the material world remains in doubt; Descartes offers his own argument for the existence of the material world from what he takes to be indubitably certain premises (see Descartes 'The Principles of Philosophy', II.i (Descartes (1931) in my bibliography).

5. Hence, this perceived permanent thing cannot be a mere representation, and must be a genuinely external thing
  6. So, the determination of my existence in time has as a necessary condition, the existence of genuine external objects
  7. Necessarily, consciousness of my own existence in time requires the possibility of determination in time
  8. So, consciousness of my existence in time necessarily requires the existence of things external to me
- (See Kant B 275–6)

Having set out this argument in an unusually succinct paragraph, Kant then proceeds to provide a number of ‘Comments’, which serve to clarify some of the more opaque premises. Comment I clarifies the first premise (see Kant B 276–7). Kant starts from the claim (which he acknowledges to be indubitable) that ‘I am’, here interpreted as consciousness of one’s own experience as ‘determined’ in time. As Kant understands it, the expression ‘I am’ can be granted to accompany all thoughts, which can be read to mean that for Kant there is no consciousness without self-consciousness, no conscious thought without an awareness that there is a person, me, who thinks.

Additionally, he writes that this expression (‘I am’) ‘directly includes the existence of a subject’, but not ‘a cognition of that subject’ (Kant B 277). What is it for an expression to ‘include’ the existence of a subject but not any cognition of it? Well, Kant (*ibid.*) goes on to say that without any cognition of the subject there is no empirical cognition, by which he means, experience. This is because, for Kant, experience involves both the thought of something existent and ‘intuition’, and the subject must be determined in relation to that intuition. The form of ‘intuition’ here must be time, as Kant has argued that this is an internal intuition. So, when Kant says that the expression ‘I am’ includes the existence of a subject, but not the cognition of it, he means that knowing or recognising that I am (rather than simply having the self-conscious awareness that is involved in any thought) requires the subject to be ‘determined’ in time.

In what sense ‘determined’? In the sense that the subject’s thought or perceptions must be so ordered such that one happens after another one, as this is the only way that we (human beings) can proceed intelligibly. To know that I am, I must be able to think ‘I am’ at two different times, and to relate these two thoughts such that I appreciate that I am the same being that existed at  $t_1$  and thought ‘I am’, and then at  $t_2$ , with that same thought ‘I am’. Otherwise one simply does not have enough information to make sense of what this ‘I’ is of which we are

to assert its existence when we think—at least, we cannot understand ourselves as a unified being, possessed of both those thoughts and persisting through time (what Kant calls the ‘unity of apperception’).

However, Kant (B 277–8) goes on to argue in Comment II, this representation cannot be determined in time without the existence of external things, as we only understand time through change in something permanent. Why should this permanent thing be external to me, though? Well, Kant (*ibid.*) argues that as the presentation ‘I’ is not itself an intuition, there will be no predicate of ‘I’ that could ‘serve as correlate for the time determination in internal sense’. Kant’s meaning here is not entirely clear, but I interpret it as the thought that as ‘I’ is a mere ‘spontaneous’ representation, that itself can only be made sense of when determined by time, it can have no characteristic—and no permanence, in and of itself—against which time could be determined. There is not that stability in self-awareness in and of itself such as to provide the relation that this thing is the same thing thinking that ‘I am’ at  $t_2$  as the thing that thought ‘I am’ at  $t_1$ . This must be so, as what permanence ‘I’ has, it derives from time, and so it cannot be that which in turn provides the permanence time needs to be determined by. Hence the ‘something permanent’ must be external, and, Kant thinks, material.

### **2.3.3 The Kantian and Cartesian ‘I am’**

The ‘I am’ of problematic idealism, or Cartesianism, is the conclusion of the *cogito*. It is established in a dialectical context in which the only direct experience is inner experience. From this perspective the Cartesian problematic idealist will accept that ‘I am’, but will hold that belief in external spatial objects can only be justified indirectly.

Kant argues that outer experience is direct—inner experience is mere presentation until ‘determined’. The Kantian ‘I’ accompanies all thoughts and experiences, but is merely formal. As he describes it, all thought and experience directly include the existence of a subject, but not the cognition of that subject. This is why he holds that there is no consciousness without self-consciousness—cognition requires recognition—and that my consciousness of my existence *is of* that existence being determined in time.

On Kant’s account of experience there are always two factors. There is something, ‘I am’, that accompanies all thought and experience, but it is empty unless it can be related to an ‘intuition’. What Kant understands by the term ‘intuition’ is somewhat ambiguous, but what

he says of it is this: ‘thoughts without content are empty, intuitions without concepts are blind’ (Kant A 51/B 75). This is enigmatic, to say the least, but I take the implication to be that intuitions are the content of our thoughts, but without some sort of conceptualisation they can be nothing to us—they will be content that lacks meaning and usefulness to us because we will not know how to process it. Hence, if the ‘I am’ is an empty thought without determination, then it seems that determination is going to involve some sort of conceptualisation of ‘intuitions’—the content of thought or experience. ‘Time’ is what Kant calls the ‘form’ of inner intuition. It is a ‘form’ in the sense that it imposes form upon our thoughts. It has a kind of organisational capacity. So, the thought is that time in some way helps conceptualise the content of thought or experience so as to enable one to recognise that ‘I am’. How does this work?

Kant is working from a Humean bundle picture of identity:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. (Hume, T 1.4.6.3)

For Hume, there is no ‘I’ to find on introspection, only a bundle of experiences. Similarly, the Kantian ‘I’ is empty because it is nothing over and above its thoughts and experiences. So Kant’s thought is that, rather than supposing that ‘I’ am something to be found upon introspecting, that which I do find upon introspecting—the experiences and thoughts that I have—are all accompanied by the ‘presentation’ ‘I am’. The thought that ‘I am’ brings an appreciation that I am the same being whose having of those two thoughts ties them together. Making sense of a series of thoughts or experiences requires the recognition that the ‘I am’ that accompanied E1 is related to the ‘I am’ that accompanied E2 in the sense that it is the same ‘I am’ that accompanies both. This relation can only be recognised, however, if E1 and E2 can be distinguished. So they must be conceived of as distinct, and the organising factor (or form) that will distinguish them must be time, as the material idealist wants to insist that our inner knowledge is prior, so we cannot have access to the form of outer sense (i.e. space). I determine that ‘I am’ on re-encountering/recognising the same ‘I’ accompanying E1 and E2 at different times.

Thus, Kant reaches the conclusion that ‘I am’ is necessarily true for self-conscious individuals by different means than the Cartesian problematic idealist, and one might question whether it is the same ‘I am’ that he uses as his first premise: ‘I am conscious of my own existence as determined in time’. Determination in time is necessary because of the bundle theory of identity, to which Descartes is not committed. For Descartes, the ‘I’ that exists is only known to do so for as long as it is thinking, it is not recognising itself over disparate moments of time as accompanying experiences. Everything is in doubt, including experiences, until the realisation that doubt, as a form of thought, is only possible if there is one who is doubting:

I have just said that I have no senses and no body... am I not so bound up with a body and with senses that I cannot exist without them? But I have convinced myself that there is absolutely nothing in the world, no sky, no earth, no minds, no bodies. Does it now follow that I too do not exist? No: if I convinced myself of something then I certainly existed. But there is a deceiver of supreme power and cunning who is deliberately and constantly deceiving me. In that case I too undoubtedly exist, if he is deceiving me. (Descartes *Meditations*, 2.24–5<sup>27</sup>)

As we shall look at again in chapter four, it is precisely the bareness of the Cartesian ‘I’ and the sharp division between mind and world in the Cartesian picture that make it difficult for a transcendental argument to work on such a view. However, we can also see Kant’s strategy. He has tried to start from a premise his opponent accepts and to take the problematic idealist’s criterion of knowledge (indubitability) seriously. The argument then moves on:

If we accept that to be conscious of my own existence that consciousness must be determined in time, there is still a question as to what determines time. Why should we think that the experiences of  $t_1$  and  $t_2$  are separate such that they can be joined by me? Kant’s thought is this: we understand time through change in something permanent; ‘I’ is a mere ‘spontaneous representation’ determined itself by time. The intuitions that ‘I’ accompanies are similarly fleeting. Hence, the ‘something permanent’ must come from outer sense—i.e. it must be external. Moreover, outer sense must be so arranged that it contain things distinct from me that persist over time. In order to understand them as being distinct from me they are

---

<sup>27</sup> References to Descartes *Meditations on First Philosophy* are to the John Cottingham (1996) translation and are abbreviated to ‘*Meditations*’.



conceived of as persisting even when not perceived. We need another dimension along which they can be separated from us—to appreciate that things can persist in time although not perceived. We need to think of them as *being* somewhere else. These things are external spatial objects.

Thus, Kant believes he has proven the existence of material objects in a way that the problematic idealist would be forced to accept: to be capable of recognising that I am, there must be a material world. We have reason to think that the argument is not well-aimed if considered as a direct response to Descartes, who has his own argument for the existence of the material world and whose understanding of ‘I am’ can be interpreted as significantly different to the ‘I am’ of Kant’s first premise. However, one can certainly see what Kant is trying to do: taking a psychological premise held immune from doubt by his opponent and arguing that that which his opponent doubts (the existence of external spatial objects) is a necessary condition of the truth of that premise. The strength of the justification of the first premise should carry over to the justification of the conclusion, removing it from doubt. It is this pattern of argument that, more than anything, has influenced the more recent philosophers who interest me.

#### **2.3.4 Transcendental Idealism, Berkeleyan Idealism, and Scepticism**

I move now to consider the similarities and differences between Kant’s transcendental idealism and the Berkeleyan idealism that he also supposes to be refuted by this argument. This is significant because the differences between the two are not as great as Kant believes, and that this gives us reason to think that the Refutation will not be reassuring to the sceptic.

It is tempting to look at the Refutation of Idealism as an anti-sceptical argument, for it seeks to establish the existence of external spatial objects. However, what is meant by ‘external’ and ‘spatial’ for the transcendental idealist is significantly different than what is meant by my sceptic. Dogmatic idealism, as Kant presents it, is the position that the existence of external objects in space is impossible. Kant (B 274) supposes that the mistake that leads the dogmatic idealist to this position is the supposition that space is a property of things-in-themselves. For Kant, space is the form of outer intuition. It orders and arranges the manifold of intuition such that is it intelligible for us. Objects in space are material, but they are also distinguished from things-in-themselves. Berkeley rejects matter because it is supposed to be

some property of things-in-themselves, and he finds it, as such, unintelligible. *However*, it is not nearly so clear that he rejects the existence of space *as Kant understands it*.

In the *Three Dialogues between Hylas and Philonous*, Philonous, challenges Hylas to define what matter is, and to argue for its existence, and what they find is that there is actually no intelligible way of understanding what matter is, and, indeed, no need to suppose its existence. Philonous says of the word ‘matter’: ‘it is plain, by your own confession, there was included no idea at all’ (Berkeley TD II.226<sup>28</sup>). However, Berkeley (TD I.190) clearly differentiates matter from extension, saying: ‘no figure or mode of extension... can be really inherent in matter’. This is not necessarily to say that spatial objects do not exist, but rather that properties (such as extension) which we attribute to space are a feature of the way our ideas relate to one another, or, as Kant would say, the form that determines our (outer) intuition. Extension is a property of sensations, and varies with them, as Berkeley (TD I.189) writes: ‘as we approach or recede from an object, the visible extension varies’. It is ‘acknowledged to have no existence without the mind’ (Berkeley TD I.191), but this is not to say that it does not exist at all.

Moreover, in the ‘Third Dialogue’, Berkeley (TD III.261–3) also raises a possible differentiation between ‘matter’ or ‘material substance’, as the philosophers talk about them, and ‘matter’ as a species of sensible thing. The latter Berkeley is perfectly happy to allow exists. So, the question, then, is whether it is the philosopher’s ‘matter’ and ‘material substance’ that Kant sees the material idealism as denying or doubting.

On the one hand, Berkeley (TD III.261) has Philonous say: ‘With all my heart: retain the word *matter*, and apply it to objects of sense, if you please, provided you do not attribute to them any subsistence distinct from their being perceived’. This is perfectly in accord with Kant, and could almost be taken to express his own point of view. Matter can be said to be a property of objects of sense—are these that different from what Kant calls ‘objects of experience’, ‘objects of sensibility’, ‘objects as such’, and even directly: ‘objects of sense’? (See, for example, Kant 1996: B 41, A 92–3/B 125–6, A 115–6, B 148–9). What Philonous says appears to accord well with Kant’s assertions that: ‘Space is nothing but the mere form of all appearances of outer senses... Only from the human standpoint, therefore, can we speak of

---

<sup>28</sup> References indicated by ‘TD’ are to the 1734 edition as reprinted by Jonathan Dancy in 1998; I therefore use the pagination of the 1734 edition, as included by Dancy.

space, extended beings, etc. ... This predicate is ascribed to things... only insofar as they are objects of sensibility' (Kant A 26–7/B 42–3). Thus, Kant and Berkeley seem very much in accord: matter, space, extension—these are properties of objects, but objects of sense, not the transcendental object, or thing-in-itself.

However, Kant does specify that his project for the 'Refutation of Idealism' is to prove 'the existence of objects in space outside me'—there is an externality to Kant's objects of sense that would appear to not sit so well with Berkeley. For Berkeley (TD III.261) writes: 'there is no matter, if by that term you meant an unthinking substance existing without the mind'. That said, there sometimes seems to be in Berkeley a blurring of the distinction between a particular mind (as in the class of minded existents) and all minds. He writes:

sensible things cannot exist otherwise than in a mind or spirit. Whence I conclude, not that they have no real existence, but that seeing they depend not on my thought, and have an existence distinct from being perceived by me, *there must be some other mind wherein they exist...* an infinite, omnipresent spirit. (Berkeley TD II.212)

I. e. sensible things (or at least, some sensible things) exist in the mind of God. Not external to mind, in that it is not external to all minds, but certainly external to a mind, specifically, one's own.

So it is at least arguable that Berkeley understands objects of sense to be external, yet at the same time, I think we must allow that this is still something different to what Kant means. Kant does not argue, as Berkeley does, that there is no such thing as a thing-in-itself, but rather that: 'The transcendental object... remains unknown' (Kant A 46/B 63). It does not seem that the sort of externality we might claim for Berkeley would satisfy Kant. There is something much more akin to the 'philosophical' sense of matter in the materialism Kant thinks it is wrong to reject. Furthermore, Kant sees no reason to rule out that which is unknown, or even unknowable; but the fact that 'matter' in the philosophical sense, is unintelligible to us, is taken by Berkeley as a reason to suppose that it does not exist. Where Berkeley states that there is nothing but the minded and sensible things, Kant argues that we should not claim to know anything about things-in-themselves, and that admitting this does not limit our knowledge of the existence of external objects in space.

What emerges from this discussion is that Berkeleyan idealism and Kantian transcendental idealism are not entirely different in their belief that there exist objects of sense

which have extension and object permanence rooted in something external to an individual's mind. For Kant it is the outer sense of space, for Berkeley, God. They differ in that Berkeley denies the existence of things-in-themselves, and although Kant does believe that there are things-in-themselves, his *Refutation of Idealism* does not address this. Things-in-themselves are unknowable.

My sceptic is concerned that she might have all the experiences she now has without there being a mind-independent world causing those experiences. In this context, Berkeleyan idealism represents a sceptical hypothesis. Extension is a mere relation of sensible things, and our understanding of words like 'matter' and 'space' must be adjusted to match this. Berkeley would say that this is how people untrained in philosophy speak anyway, but in my introduction I showed that the possibility that what we experience is not caused by something *beyond* the senses is regarded as a worry in many cultural items enjoyed popularly by those untrained in philosophy.

Where Berkeleyan idealism expresses the sceptical hypothesis, transcendental idealism can be understood as expressing the doubt. There *is* something out there (things-in-themselves) that causes the intuitions we then form into spatiotemporal objects of sense, but we cannot know anything about those things. The *Refutation of Idealism* offers comfort to Kant because all he is concerned with are external objects in space—he has severed from his concepts of space and externality the mind-independence I require to assuage the sceptic.

For this reason, transcendental idealism and the *Refutation of Idealism* cannot provide me with a solution to this problem. However, I do find in this style of argument the kernel of an approach, taken up by twentieth and twenty-first century analytic philosophers, wherein we may find a viable approach.

So this, then, would seem to be the form of idealism Kant is anxious to distance himself from. It is the theory that it is either false or indemonstrable that there should be external objects in space; where 'external' is taken in a more robust sense than Berkeley's assertion that at least some perceived ideas exist in the mind of God, and yet is not so robust as to refer to these objects as things-in-themselves. Moreover, Kant would reject an idealism that asserted matter to be a property of things-in-themselves, and hence moved from a rejection of things in themselves to a rejection of matter.

## 2.4 Austin

Barry Stroud (2000d: 203)<sup>29</sup>, gives the first use of the term ‘transcendental argument’ as coming from J L Austin in his 1939 paper ‘Are There A Priori Concepts?’<sup>3031</sup>. Austin’s primary purpose is not to give a definition of transcendental arguments, so what he means must be extracted from his discussion.

Austin first raises the term in analysing an argument for universals, which he characterises as being based on these assumptions: we sense multiple things that are different to one another, we call the different ‘sensa’ by the same ‘name’ (‘we say “This is grey”, and “That is grey”’), this practice is considered ‘justifiable’ in the sense that it is indispensable. It is then argued, firstly, that because we use the same name for different things, there must be *something* that is identical that is so named and is in common to all the sensa to which it is applied: a universal. Secondly, because the things sensed are ‘many and different’, the universal, as single and identical, must not be sensed (see Austin 2003: 33–4).

Austin (2003: 34) gives two ways in which we can understand the above argument to be transcendental, one which is owed to Kant, and another which he says is not. It is ‘Kantian’ sense in that it argues that: ‘if there were not in existence something other than sensa, we should not be able to do what we *are* able to do’. It is considered to be transcendental in *another* sense in that it aims to ‘prov[e] the existence of a class of entities different in kind from sensa (viz. name things [universals])’ (*ibid.*).

He also examines another argument, which he holds to be transcendental:

A true statement is one which corresponds with reality: the statements of the scientist are true: therefore there are realities which correspond to those statements. Sensa do not correspond to the statements of the scientist... therefore there must exist other

---

<sup>29</sup> References to Stroud are to the 2000 edition of his collected papers, *Understanding Human Knowledge: Philosophical Essays*. Individual papers are distinguished as 2000a, 2000b, 2000c, and 2000d relative to their positioning within that work.

<sup>30</sup> References here refer to the 2003 edition.

<sup>31</sup> Stern (2000: 8, n. 8) notes that there are preceding instances. Kant uses the term ‘transcendental arguemnt’ at A 627/B 655, but he uses it to mean something that ‘goes beyond the legitimate employment of understanding’ (*ibid.*), this is therefore a different use to both Kantian transcendental deduction and what is meant by the modern term ‘transcendental argument’. Additionally, C S Pierce used the term in 1902; however, as his use has had no significant impact on the development of the use of the term, I leave it to one side.

objects, real but not sensible, which do correspond to the statements of the scientist. Let these be called 'universals'. (2003: 35)

Austin says that this argument is transcendental in that:

The 'universal' is an *x*, which is to solve our problem for us: we know only that it is non-sensible, and in addition must possess certain characters, the lack of which prohibits *sensa* from corresponding to the statements of the scientist. (*ibid.*)

Austin does not say how this relates to either of the characterisations of what it is to be transcendental that he gave before, but I think we can draw two senses in which the argument may be considered transcendental from this. Firstly, the argument starts from something assumed to be true: 'the statements of the scientist are true'. Then we are given a problem: *Sensa* do not correspond to the statements of the scientist. Given this, an explanation is required of how both statements about *sensa* and scientific statements can be true. The explanation is that there are real things which correspond to the statements of scientists and to which *sensa* are appropriately related to make sense of them as well. This is transcendental in the sense that the transcendental *deduction* is transcendental. It is not so much an argument as an *accounting* for the justification we already have to hold certain statements to be true. The first argument also reflects a flavour of this—one of the assumptions was that we were already justified in the indispensable practice of referring to multiple *sensa* by the same name, and the existence of universals provides an accounting for this.

One can see this sense of being transcendental as both similar and different to the senses given above. In the first sense, we are able to know scientific statements to be true because there is something in existence other than *sensa*. In the second sense, we could read it as an attempt to prove the existence of a class of entities different in kind from *sensa*. In another sense, it is like neither, for if it were a transcendental deduction in the true Kantian sense, it would not be trying to give a truth, but rather an accounting for justification of something that is already known.

Stroud reads this differently. On Stroud's (2000: 204) analysis, there is nothing special about the first argument as Austin summarises it, it is just an instantiation of *modus tollens*: if there did not exist a different kind of entity from *sensa*, we would not be able to name things; we do name things, so there is a different kind of entities from *sensa*. *Modus tollens* is not a

particularly special or new form of argument that we owe to Kant, so whatever it is that makes transcendental arguments special must be something else.

Looking to the second argument for help, Stroud characterises it thusly:

attempting to prove something by showing that it is required as a solution to a problem: in this case, how is it possible for science to make true statements about reality, given that they are not statements about “sensa”. Stroud (2000d: 205)

Which sounds similar to my own reading, but what Stroud identifies as the characteristic exhibited by this strategy is that the argument is to prove that one thing is a necessary condition of another. As we shall see, this conforms to Stroud’s overall conception of transcendental arguments, but differs from the more Kantian reading. Moreover, as Stroud (*ibid.*) states, this cannot be a sufficient condition for an argument’s being transcendental. It is much too broad: all valid deductions would thereby be transcendental. They also share the characteristic of trying to show that something is a necessary condition of something we do that is justified or indispensable. This, also, Stroud argues, is not enough: on this definition we could transcendently prove the existence of bananas on the basis of the fact that we eat them. I wonder if Stroud has not got a bit muddled, here. It is not indispensably true that I eat bananas. I think he is thinking of the indispensability of there being bananas on condition that I have eaten some, which would be a much more ordinary form of necessity. However, to be like Austin’s examples, the thing known to be true is that which is indispensable, and this is why the indispensability requires explaining. For the example to be relevantly similar it would need for my eating of bananas to be indispensable, which it clearly is not.

Stroud goes on to note that Kant was specifically concerned with our thought and experience of things, the *objects* of our thought or experience. Something that is required for thought or experience has a special status, he argues (above what is required for the eating of bananas). Stroud notes that Kant saw thought and experience as the starting point of metaphysics, and a starting point that would make metaphysics a secure science—but *why*? Well, we quite clearly *do* think and experience things. As Kant notes in the *Refutation*, even the problematic and dogmatic idealist must grant *that*; and we might add: even the *sceptic* does. Hence: anything that is found to be a necessary *condition* of thought or experience *must* be true. As Stroud (2000d: 207) says: ‘they cannot fail to hold if we think at all’.

In this, I think Stroud is correct that this is an essential part of the kind of transcendental argument we find in the Refutation of Idealism, precisely because the Refutation is looking to a) start from a premise accepted by the problematic and dogmatic idealists, and b) meet the high standard set for justification by the problematic idealist. It is also true that this is a primary concern to Austin, who is dealing with the relationship of sense to universals. However, it is important to understand that it is not a transcendental argument in the sense that the transcendental deduction is. Just as Kant would not have recognised the Refutation as a transcendental deduction. Because for Kant the indispensability lies in the assumption that requires explaining. I suspect this is why Austin differentiates both a Kantian and a non-Kantian sense in which an argument may be a transcendental. One is an accounting for the justification of some already accepted phenomena, the other is a proof of something based on it being a necessary condition of some fact about thought or experience. In the latter the nature of the proof is ‘transcendental’ in that it seeks to ‘transcend’ what Strawson would later call the ‘bounds of sense’ by starting with a premise concerning thought or experience and moving to a conclusion concerning something ‘beyond’ that boundary. In the former, the form of deduction is not transcendental, it is simply the accounting for our justification to know something we *do* in fact know, it is the subject matter (the application of the categories to objects of sense) of the deduction that is transcendental.

A failure to differentiate these two kinds of transcendental approach has led to both innovation and confusion. In a sense, it does not *matter* if one is reading Kant correctly, but as philosophers have continued to reference their understanding of transcendental arguments back to Kant, this confusion has influenced how transcendental arguments have developed.

## 2.5 Strawson: *Individuals*

Perhaps the most influential figure in the development of transcendental arguments since Kant is P F Strawson. His two seminal works, *Individuals* (1959<sup>32</sup>) and *The Bounds of Sense* (1966<sup>33</sup>), mark two significant transcendental paths of his early thought. The subtitle of *Individuals*, ‘*An Essay in Descriptive Metaphysics*’, clearly demarcates his approach as

---

<sup>32</sup> References refer to the 1987 edition.

<sup>33</sup> References are to the 1976 edition.



specifically descriptive, as opposed to ‘revisionist’. It is a style of conducting metaphysics that he attributes to Kant and Aristotle, and which is distinguished by seeking to clarify the necessary connections between our concepts and describing the fundamental concepts that underlie the varying metaphysical views of different times: the ‘permanent relationships’ described in ‘impermanent idiom’ (see Strawson, 1987: 10–11).

In particular, Strawson argues that material bodies and persons occupy a central place as fundamental particulars (Strawson 1987: 11). There is no space here to examine his arguments in any great detail, but the key is to note Strawson’s methodology. The descriptive approach, which Strawson attributes to Kant, is not exactly akin to the ‘accounting for our justification’ approach that I have ascribed to the Transcendental Deduction. Yet this mapping out of conceptual relations *is* strongly reminiscent of Kant’s description of the categories and their relation to our ability to have the kinds of thoughts and experiences we do in fact have. Moreover, Strawson’s arguments take the form of looking at what we are capable and considering what must be the case for this to be so. Thus he moves from noting our ability to identify particulars in speech to the necessity of there being objective particulars located in space and time, distinct from the individual (see Strawson 1976: 9–30). From our ability to reidentify particulars to the necessity of our supposing that there is some dimension along which they continue to exist when not present to our minds, and via which they can again be found by returning to the same ‘place’—a point along that dimension at which the particular was supposed to have been when last we encountered it, and at which it remained when we moved to some other place along that dimension (see Strawson 1976: 31–8). Thus we move from that which we certainly do know about our conceptual capacities to the discovery of how we are required to conceive of the world in order to have these conceptions we do in fact have.

This, too, bares similarities to Kant, especially with regard to the importance of space and time as organising factors in enabling us to conceptualise our experiences in the way we do. Although the similarity should not be overstressed. The metaphysics Strawson describes is not one of transcendental idealism, but one of material bodies located in dimensions of space and time, rather than space and time as the forms imposed on inner and outer intuition in order to ‘form’ intelligible experiences for us. Nevertheless, it’s never clear that what we get from *Individuals* is an argument *for* the existence of persons and material objective particulars, or if merely the *concept* of such things is essential to our human conceptual schemes.

## 2.6 Strawson's Transcendental Deduction and *The Bounds of Sense*

In *The Bounds of Sense*, Strawson attempts to reconstruct a version of the Transcendental Deduction that is accessible from an analytic approach and which does not rely upon Kant's transcendental idealism and what Strawson calls Kant's 'transcendental psychology' (see Strawson 1976: 88–89). By 'transcendental psychology' Strawson means those elements of the deduction that he sees as having a more explanatory role, detailing the way in which Kant understands our faculties as they relate to experience as it functions under his transcendental idealism.

As such, this is a deliberate departure from what I have characterised as an explanatory kind of transcendental philosophy and a move towards a kind of transcendental *argument*, one that rejects the transcendental idealism that Kant sees as a solution and that my sceptic regards as a worry. It is this *kind* of transcendental argument that I believe may be fruitful in answering the sceptic; although I hasten to add that I am not defending this argument in particular, and shall therefore not dwell on defending its details.

Strawson's interpretation of the Transcendental Deduction is articulated over sections five to eight of Part II of *The Bounds of Sense*. It may be summarised thusly. It is a necessary condition of the possibility of experience that it be possible for one to self-ascribe diverse experiences to a numerically identical self (Strawson 1976: 110). This self-ascription is taken minimally as a kind of self-reflexiveness that is to be a necessary pre-condition of self-consciousness (Strawson 1976: 107). The claim is not intended as a sufficient condition for self-consciousness, but as a necessary precondition of it. Kant's acceptance of the bundle theory of identity means that there is no empirical self to be encountered in experience; Strawson cites in particular the passage where Kant writes:

In the synthetic original unity of apperception I am conscious of myself not as I appear to myself, nor as I am in myself, but only that I am. This representation is a thought, not an intuition. (Kant B 157; quoted in Strawson 1976: 95)

Which is understood to mean that one neither encounters a representation of oneself in experience, nor oneself as a thing-in-itself, merely the accompanying thought 'I am' that is with all other experiences, linking them as a unified presentation. It is a 'transcendental consciousness', rather than an 'empirical self-consciousness', in that while it does not

constitute the recognition of oneself as a conscious being, it is a necessary condition of that recognition, as it provides a unifying character to one's experiences (Strawson 1976: 95).

Strawson reads Kant's argument as being that the *synthesis*, by the imagination, of diverse discrete sense data under the categories determined by the faculty of understanding, is the necessary precondition of ascribing experiences to oneself, thus offering an explanation of the 'unity of diverse representations in a single consciousness' (Strawson 1976: 96; see also 97). He finds this explanation unsatisfactory, and seeks to provide an alternative account of the precondition of self-ascription requiring this unity that does not rely upon such transcendental psychological talk of 'synthesis' and the workings of the faculties of imagination and understanding. Instead, he thinks we can draw on the 'necessary duality of intuition and concept' (Strawson 1976: 97). This notion takes a little unpacking. Strawson characterises 'intuitions' as: '[Kant's] word for awareness in experience of particular instances of general concepts' (Strawson 1978: 20). They are to be distinguished, then, from mere awareness—intuitions are awareness of something, but without *bringing* the intuition *under* a concept, one cannot have a sense of *what* that awareness is of. Recall Kant's statement that: 'thoughts without content are empty, intuitions without concepts are blind' (A 51/B 75). Understood on Strawson's interpretation, this means that intuitions are not really things of which we can be aware unless they are conceptualised as particular instances of a certain kind of thing. At the same time, the *thought* that 'I am' is empty unless it is accompanying an intuition, for the transcendental self is nothing more than the unification of diverse experiences as belonging to the same individual. It is a symbiotic relationship: intuitions are not purely states of awareness that the subject has, and if they were, the subject would not be able to make anything of them.

What, then, is it for an intuition to be brought under a concept? It is to make a judgement about what the intuition is of, for example "'this is a heavy stone'" (Strawson 1976: 101), and to make a judgement is to attribute objectivity to the thing judged, in the sense that it has an existence independent of that which is perceived: 'esse' (existence) is distinct from 'percipi' (perception) (Strawson 1976: 100). Part of judging something to be the case is an acknowledgement that it might not be the case, or it might not have been the case—there must be room for correction (see Strawson 1976: 106). For a judgement to claim 'objective validity' is for it to claim a validity independent of the subject. Judgement requires a distinction between how things are objectively and how things are subjectively. When I apply a concept and judge that an object of my experience is thus and so, I am inherently attributing to the

object an existence as a part of an ordered reality that may differ from my experience of it, thus: ‘experience of objects is possible only if objectively valid judgements are possible’ (Strawson 1976: 98). It is only possible to make judgements if that which the judgements are made of is of a unity distinct from the unity of one’s own mind.

So. Experience is only possible under judgement. Experiences must have an order and arrangement of their own distinct from the order and arrangement as we are aware of them, or there will not be any meaning to the judgement. Therefore experience requires self-ascription *and* a sense of an ordered reality distinct from what is ascribed to the self. It is therefore a necessary condition of my having the experiences I do have that there be an objective, mind-independent reality: ‘our pictures of the objective world and our picture of possible perceptual routes through it cannot be independent of each other’ (Strawson 1976: 105).

Thus, the Strawsonian interpretation of the Transcendental Deduction is one that aims to show that the existence of a mind-independent objective reality is a necessary condition of the possibility experience and self-ascription of experience. Given that we do have experiences and do ascribe them to ourselves in a unified manner, there must be an objective world (or at least, we must have a ‘picture’ of such an objective world).

## **2.7 Stroud and Transcendental Arguments**

Barry Stroud’s attempts to offer a systematic characterisation and criticism of transcendental arguments have been as influential as Strawson’s move to take transcendental arguments seriously in an analytic context. His criticisms have contributed towards a shift in focus towards a less ambitious, weak form of transcendental arguments. I will look in detail at Stroud’s criticisms in chapter four, but for the moment, my purpose is to draw out what he views as the defining characteristics of transcendental argument.

Stroud’s assessment of what it is for an argument to be transcendental is variable. He talks of the significance both of the ‘goal’ of transcendental arguments and of the content of their premises. At times he seems to focus overly on the historical origins in Kant, and at other times to have more concern with how modern transcendental arguments can and do distinguish themselves from this. Of course, it would be surprising if there were no development to be found in Stroud’s thought from 1968 to today, but it is often difficult to untangle the different lines of thought and see what conclusions have actually been drawn. My

purpose over the next few sections is to unpick the varieties of transcendental argument Stroud identifies.

### **2.7.1 Transcendental Arguments as Anti-sceptical**

Stroud initially saw the Transcendental Deduction and the Refutation of Idealism as part of the same project, one aimed at refuting the sceptic. In his seminal 1968 paper, 'Transcendental Arguments' (Stroud 2000a), he begins his characterisation of what a transcendental argument is with a return to Kant, and an outline of what he perceives the motivation of Kant's project to be. At odds with the interpretation I have given, Stroud understands the question of our 'rights' or 'justification' for certain propositions, with which Kant is interested in the Transcendental Deduction, as concerned with 'the challenge presented by the epistemological sceptic' (Stroud 2000a: 10). One can understand why he sees it thus when he quotes a passage from the B text where Kant writes that it is 'a scandal to philosophy and to human reason in general that the existence of things outside us... must be accepted merely on faith' (B xl). Stroud does not define the epistemological sceptic precisely, but one may infer from his concerns that this is one who doubts our knowledge of the external world, and elsewhere he has written that the concern of philosophical scepticism can be conceived of as the worry that it is logically possible that things might seem to us just as they do without the existence of physical or material objects outside the mind. His sceptic is this very similar to my own (see Stroud 1979: 278).

Certainly, as can be seen in the Refutation, Kant is concerned with dispelling doubt in the existence of things outside of us and, in that passage, he is concerned to show that his philosophy can demonstrate the fault of such doubts. However, I think it is mistaken to interpret Kant's understanding of what a transcendental deduction is concerned with in terms of its potential as a tool against the sceptic. As outlined above, I think Kant is interested in doing something subtly different. Kant believes that the *a priori* application of these concepts has already been established; what he is looking to do in the deduction is demonstrate our *right* to them.

To the analytic philosopher, this seems a very strange way to go about things: looking for a justification of concepts that have already been established. I think this is why Stroud feels the pressure to read these two projects within the *Critique of Pure Reason* as aimed at the

same goal: rebutting the sceptic. Moreover, Strawson's work reconceiving the Transcendental Deduction in a form similar to that of the Refutation of Idealism (in which an objective world is established as a necessary condition of our thought and experience) encourages this view.

To do justice to Stroud, he does recognise this difference elsewhere. In 'Kantian Argument, Conceptual Capacities, and Invulnerability' (1994)<sup>34</sup> he distinguishes two aspects of Kantian strategy. One aspect he identifies is establishing 'the invulnerability of certain fundamental ingredients in our conception of the world'; the other is 'the question of how that invulnerability is to be established' (Stroud 2000b: 158). Although Stroud states that this is not a distinction Kant recognised himself, I would argue, rather, that Kant did distinguish between these two aspects, but saw only the latter as being the direct concern of transcendental deduction. True, the 'invulnerability' of certain claims to knowledge is a significant aspect (or at least, consequence) of Kant's transcendental *philosophy*—one that he explores in the Refutation of Idealism and uses to emphasise the distinction between his own form of idealism and that of dogmatic and problematic idealism—but it is distinct from the goal of the Transcendental Deduction, and hence, I think, from what Kant would have thought of as the distinctive kind of *argument* he engaged in there.

The invulnerability of the conclusion of a transcendental argument, 'those broad features of our conception of the world which are themselves necessary conditions of our thinking about or experiencing anything at all' (*ibid.*), remains a focus of Stroud's understanding of transcendental arguments. He returns to this thought again, in 'The Goal of Transcendental Arguments' (1999)<sup>35</sup>, although there his aim is to persuade us that a different goal might be more fruitful. This is because of the criticisms Stroud raises in 'Transcendental Arguments' and later papers which are that: 1. transcendental arguments that aim at defeating an external world sceptic always rely on a suppressed verification principle, and 2. to reason from premises about the mind to conclusions about the world would require an implausible 'bridge of necessity'. A full account and assessment of these criticisms is reserved for chapter four, but for now it is sufficient to note that these thoughts motivate a move away from thinking of sceptical arguments as a tool useful for directly refuting the external world sceptic.

---

<sup>34</sup> References are to the 2000b edition.

<sup>35</sup> References are to the 2000c edition.

## 2.7.2 Goals for Transcendental Arguments

Although the title of the paper highlights the fact that he still thinks that the transcendental *target*, as it were, is of significance, several different options are presented in ‘The Goal of Transcendental Arguments’. On Stroud’s (2000d: 204) analysis of Austin, the arguments considered are transcendental in that entities that ‘transcend’ sense experience—universals—are (intended to be) proven to exist on the basis that they are in some sense ‘required’ for some practice we undertake ‘justifiably’ or indispensably. As Stroud takes these arguments to be structurally no different from *modus tollens*, he declares that the key difference is their *goal*: ‘what makes it transcendental is not its logical form or its subject matter, but its aim or goal’ (Stroud 2000d: 205). The goal here being establishing the existence of something not found directly in sense experience.

Elsewhere, this goal-centric claim is qualified. Stroud writes that: ‘Kant was concerned with the necessary conditions... in particular of our thinking or experiencing things—of there being “objects” of our thoughts and perceptions’, and that: ‘To show that something is required for thought or experience would be to show that it has a very special status... transcendental status’ (Stroud 2000d: 207). This suggests that what is distinctive is not *just* the goal in the sense of demonstrating the truth of something that the sceptic denies or something that transcends sense experience, but to do so in a specific way: by showing that the truth of the conclusion is *a necessary condition of thought or experience*. Hence, the nature of the premises and their subject matter *would* seem to be significant. Furthermore, he writes:

To say that scepticism can be used to draw attention to the special character of transcendental arguments is not to say that the only point of going for transcendental arguments is to refute scepticism, or to suggest that that is mainly what Kant was trying to do. (Stroud 2000d: 211)

This is a distinct departure from his characterisation in ‘Transcendental Arguments’<sup>36</sup>, and it is more useful to us. Being anti-sceptical is no longer considered the ‘goal’, rather, consideration

---

<sup>36</sup> Stroud (2000d: 208–9) claims that it is a mistake to suppose that he had Kant and Kant’s transcendental idealism in mind when, in 1968, he raised ‘those questions’ about whether a transcendental strategy can succeed. Certainly, it is fair to say that he was not considering the potential success of transcendental idealism and Kant’s transcendental deduction as an instantiation of this kind of strategy. However, he was drawing on his

of the anti-sceptical role of many transcendental arguments helps draw out what is distinctive about transcendental projects. Scepticism does this, he claims, because the very propositions doubted by the sceptic are necessary conditions of our having the thoughts and experience necessary to ask sceptical questions. As he puts it: ‘A positive, reassuring answer to the question is shown to be built into our asking it, and so to be guaranteed’ (*ibid.*).

Freed from taking the defeat of the epistemological sceptic as a defining element of transcendental arguments, Stroud moves forward to advocate a different kind of transcendental argument, with a different kind of goal. It is a goal he finds in Strawson’s later work. Strawson, he notes, has become content to explore the necessary connections between concepts along the lines that ‘Given I have concept x, I must also have concept y’ (see Stroud 2000b: 163 and 166, and 2000d: 213). This is because (according to Stroud 2000b: 165) Strawson regards the sceptical question as ‘idle’, in that we could never actually be made to change our minds about the existence of objective particulars and so on. This is a shift towards what we may call ‘weak’ transcendental arguments, as opposed to ‘strong’ transcendental arguments, conceived of as attempting to rebut the epistemological or external world sceptic.

### **2.7.3 Strawson: Naturalism and Scepticism**

In *Skepticism and Naturalism* (1985)<sup>37</sup>, Strawson advocates a more ‘naturalistic’ approach to dealing with the sceptic than the kind of ambitious project he explored in *The Bounds of Sense*. In this he advocates a response to scepticism that ‘does not so much attempt to meet the challenge as to pass it by’ (Strawson 1985: 3). This is a return to, or a reinterpretation of, the Humean response to scepticism. Strawson conceives of it as both naturalistic and transcendental.

Strawson understands Hume as saying that ‘all arguments in *support* of the sceptical position are totally inefficacious; and, by the same token, all arguments against it are totally idle’ (Strawson 1985: 11). The sense in which Nature determines that we cannot accept the

---

understanding of what he took to be Kant’s project as a way of understanding and characterising transcendental projects in general.

<sup>37</sup> Based on his Woodbridge Lectures delivered at Columbia University in 1983.



sceptical conclusion is that belief in the body and in other minds is ‘inescapable’ (*ibid.*). This is transcendental in the sense that:

in order for the intelligible formulation of skeptical doubts to be possible or, more generally, in order for self-conscious thought and experience to be possible, we must take it, or *believe*, that we have knowledge of external objects or other minds. (Strawson 1985: 21)

This is *not* the move that Stroud was so critical of in ‘Transcendental Arguments’—i.e. a move from a psychological fact about what we must believe to a fact about how the world must be—it is what Strawson calls an ‘only connect’ model of transcendental arguments (see Strawson 1985: 21–2). The argument establishes something about how we must conceive the world based on the concepts we must have in order to form sceptical doubts in the first place, or to have self-conscious thought or experience at all. This kind of transcendental argument harks back to the descriptive metaphysics of *Individuals*—transcendental arguments merely elucidate the necessary connections between concepts. It is a natural fact that we cannot accept the sceptic’s conclusion, and therefore the question is idle.

#### **2.7.4 Stroud and Weak Transcendental Arguments**

Stroud, although wanting something weaker than what he calls ‘strong’ or ‘would be’ transcendental arguments, still thinks that there is something more than this ‘natural fact’ about what we can believe that transcendental arguments can uncover for us. He thinks that via this kind of Strawsonian conceptual analysis we might be able to establish a sort of invulnerability to certain beliefs; we might find that some beliefs are required in order for us to have any beliefs at all, and it would be from *this* that we could say that it would be impossible to find that belief false. This he takes to be stronger than Strawson’s thought, because it bestows a kind of ‘invulnerability’ on certain beliefs. There would thus be *some* kind of reassurance in knowing that it would be impossible to find these beliefs to be false. (See Stroud 2000b: 165–6, and 2000d: 216.)

To defend this sense that weak transcendental arguments can deliver invulnerability, Stroud draws particularly on what he thinks of as the ‘weak’ version of Davidson’s transcendental argument from triangulation, the gist of which is that: ‘If we ascribe beliefs to

others, we must assume that they, and, by extension, we, are largely correct'<sup>38</sup>, thus rendering it impossible for us to find that our beliefs about the world are not, for the most part, true. This is outlined particularly in 'Radical Interpretation and Philosophical Scepticism' (1999)<sup>39</sup>.

Stroud writes:

The possibility from which the sceptical line of thinking begins is therefore not one which anyone could consistently find to be actual. It involves the presence of certain

---

<sup>38</sup> To say that this is a very brief account of Davidson's argument is an understatement. Davidson's argument from triangulation arises from his theory of meaning and radical interpretation, unfurled across many papers. The above is a summary sufficient for the discussion of what Stroud takes from Davidson, focusing on the role of charity in the ascription of belief to others. This 'charity' is understood through principles of Coherence and Correspondence, according to which, in order to interpret another, one must attribute a certain amount of logical consistency to the thoughts of a speaker, and a certain amount of correspondence with features of the world that are shared with the interpreter (see Davidson 2002b: 148–9; 2002d: 211). This is more than simply a condition of interpreting others, however, it is a condition it is a condition of knowing out own minds (see Davidson 2002a). For Davidson, meaning is communal, 'Communication begins where causes converge' (Davidson 2002b: 151). Moreover, our very ability to have beliefs is dependent upon an understanding of objectivity (that beliefs can be false), which Davidson holds to only be possible when confronted by error, which in turn is only possible against a background of successful communication with another minded being about a shared external world of which you are both largely correct. In order to understand someone contradicting your own assertions and beliefs, you must have been referring to enough of the same things to interpret one another intelligibly. Davidson thus holds that 'Belief is in its nature veridical' (Davidson 2002b: 146). From this theory of meaning two possible kind of transcendental argument emerge, one that is ambitious/strong, and one that is more modest. The difference turns on whether we actually need to be in successful communication with another minded being about a shared external world of which we are both largely correct, or if the same results can be achieved if we merely *believe* that this is the case. Davidson at first thought that the former would be true, arguing that the coherence required for interpretation would 'yield correspondence' (Davidson 2002b: 137), that interpretation requires making speakers *right* most of the time (*ibid.*), that 'most beliefs are correct' (Davidson 2002c: 168), and that 'a theory of interpretation cannot be correct that makes a man assent to very many false sentences' (Davidson 2002c: 169). However, in 'Afterthoughts' (Davidson 2002b: 154–7) Davidson distances himself from these ambitious, anti-sceptical claims, saying that what his arguments enable is not so much a refutation of the sceptic, but telling him to 'get lost'; we have to attribute truth to a speaker's beliefs and consistency with our own, and to that extent it does not matter whether those beliefs 'correspond' to an independent reality. It is this, more modest argument that Stroud argues in favour of: that we must *assume* that we and those we speak to are largely correct in our beliefs.

<sup>39</sup> References are to Stroud 2000c.

specific beliefs, but attribution of those beliefs requires our finding them to be largely true. Believing a certain set of propositions to be all or mostly false precludes our assigning them as contents of the beliefs of any people we find to have beliefs. (Stroud 2000c: 198)

‘Invulnerability’ is thereby understood in terms of our inability to find something to be actual. Because belief attribution requires the assumption that both myself and the person to whom I attribute beliefs are correct not just in that belief, but in most of our other beliefs. Scepticism, as the thought that all or most of our beliefs might be false, is thus not simply ‘idle’, but in a certain sense impossible. In holding that I have beliefs that might be false the sceptic attributes to me beliefs. Which is only possible if most of my beliefs are believed to be true. A weak Davidsonian argument can thus be used to show that the belief that most of my beliefs are largely true is invulnerable in the sense that I cannot doubt the truth of that belief without believing it to be true. This does not show that it is not *possible* that I might be subject to a sceptical scenario; rather, if this kind of weak transcendental argument is correct, then we cannot *believe* the scenario to be true.

### **2.7.5 Stroud’s Taxonomy of Transcendental Arguments**

We can therefore identify several different types of transcendental argument. 1. those that aim to defeat the sceptic on the basis that it is a condition of sceptical doubt that the sceptical hypothesis not hold. 2. those that aim to establish the existence of entities that ‘transcend’ sense experience on the basis of some practice or belief that is indispensable to us. 3. those that aim to establish that something is the case because it is a necessary condition of thought or experience (the goal is not definitive here, it may be either anti-sceptical or not). 4. those that aim merely to establish ‘natural facts’ by articulate the necessary connections between our concepts (these arguments are not directly anti-sceptical, but can be used to show that the sceptical question is ‘idle’, because we could not believe it). 5. those that aim to establish invulnerability from sceptical doubt on the basis that it is a condition of our having thought or experience that it be impossible to believe the sceptical hypothesis to be possible.

1 and 2 I understand as ‘strong’ transcendental arguments, with 1 as ‘very strong’, in the sense that it not only aims to rebut sceptical doubt, but show that the truth of that which is doubted is a condition of the doubt itself. I understand 4 and 5 to be ‘weak’ transcendental

arguments in the sense that they do not aim to establish that anything be true of the world, but focus on showing what we must believe. I take it that whilst both of these arguments as ‘weak’, Stroud sees 4 as weaker than 5, as its focus is on articulating natural facts about what we must believe, given other beliefs we have, rather than establishing ‘invulnerability’ for any proposition. I take 3 to be more of a general characterisation of what it is for an argument to be transcendental as a whole, not just because it covers both strong and weak transcendental arguments as Stroud understands them, but because it characterises what I, also, find interesting in them: establishing that something must be the case because it is a necessary condition of thought or experience (whether that be merely that we must believe something, or that something extra-mental is true) has value because thought and experience are prized as more certainly known than the extra-mental, and if something is a necessary condition of the truth of some prized proposition concerning the mental, then the justification we have for our belief concerning the mental is transferred to whatever is a necessary condition of the truth of that belief.

## **2.8 Why a Strong Transcendental Argument is Needed to Address my Sceptic**

In the preceding sections I have outlined an understanding of transcendental arguments as these have descended from Kant. It is not possible in this space to articulate all varieties of argument that have been termed transcendental, but these cover the broad categories. Clearly the kind of deduction Kant originally described in the transcendental deduction will not be of use in the context of responding to my sceptic. We cannot simply assume the truth of the propositions the sceptic doubts, and providing an internally consistent account of the psychological that cause us to have the types of thoughts and experiences we do have is not sufficient. The sceptic concedes that one explanation for us having the types of thought and experience we do have is that the world is much as it seems. The issue is that she thinks these thoughts and experiences could be as easily explained by something radically different.

We must then look to either strong or weak transcendental arguments as Stroud describes them<sup>40</sup>. Most recent philosophers have favoured the latter, taking strong transcendental arguments to have been shown to be problematic by Stroud's criticisms. However, I argue that weak transcendental arguments cannot provide a satisfactory answer to my sceptic. In the remainder of this chapter I will articulate why weak transcendental arguments cannot be of use to us, before going on in the rest of my thesis to defend the plausibility of strong transcendental arguments.

### 2.8.1 Against Weak Transcendental Arguments

I am not convinced of the efficaciousness of Stroudian weak transcendental arguments against scepticism. Perhaps one can construct an argument that shows that there are certain beliefs that I simply cannot find to be false, but it does not follow from my being unable to find something false that it is *not* false. Of course, Stroud does not think we should try to show the sceptical scenario *is not true*, merely that it is impossible for us to believe it to be so, but this does not render these belief invulnerable to doubt. Firstly, the Davidsonian style argument Stroud advocates can only establish that *at one time* or *at most times* I must believe something to be the case, or, at its most ambitious, that most of my beliefs are correct. But even if I must *at most times* believe *most* of my beliefs to be correct, or even actively believe the denial of the sceptical hypothesis, this does not rule out my intelligibly wondering if I am justified to such a belief. That it was once necessary for me to conceive of the world in a certain way does not make my doubting it *now* incoherent. Furthermore, even if the argument shows that I have to conceive of it being that way *now*, it can still make sense to ask whether that which I have to believe is so *really* is so.

Moreover, if such an argument were to succeed, I cannot help but concur with Burge (2003: 345) that it would be rather more unsettling than reassuring, as: 'The sceptic can

---

<sup>40</sup> Stroud is not the only philosopher to have made these kinds of distinctions. Robert Stern, in particular, provides a detailed taxonomy of transcendental arguments in his *Transcendental Arguments and Scepticism* (2000) that is rich and interesting. He also distinguishes between strong and weak transcendental arguments, and advocates for a more 'modest' approach (see Stern 2007). I focus on Stroud for the influential role his writings have taken in framing the debate, and because his account sets up what I take to be the main criticism strong transcendental arguments face, which I will respond to in chapter four.

heartily agree with the claim [that we cannot do otherwise than to find some of our beliefs true] and say that that just brings out our pitiful, irrational plight'. If anything, the success of Stroudian weak transcendental arguments heightens the worry.

Strawson could respond to Burge on this front by questioning whether the sceptic has thereby shown our plight to be irrational; although we may be wrong, that does not mean that we cannot have good reason to believe. The role of transcendental arguments under his picture, then, could be to show us why our beliefs are still reasonable.

This is not accessible to Stroud. As we have seen, Stroud wants something stronger than that with which Strawson is prepared to settle. He wants invulnerability. This suggests that Stroud feels that there is *something* about sceptical arguments that requires a more substantial response; some kind of 'ruling out', even if it does not require direct proof of the falsity of sceptical scenarios. Hence, I think we can say that motivation for employing *Stroudian* weak transcendental arguments evaporates: they are at best ineffectual, and at worst, somewhat disturbing.

### **2.8.2 Strawsonian Weak Transcendental Arguments**

The question then becomes one of whether Strawsonian weak transcendental arguments could satisfy our sceptic. I hope it is clear that they could not. Strawson's argument is that key claims hang together well to form a coherent point of view. Perhaps a system of beliefs in which the belief that the extra-mental exists hangs together well with beliefs about induction and other beliefs appropriate for our point of view on the world, but the sceptic's worry is that so could other systems for which these beliefs are fundamentally false. Strawson claims that although there may be some changes to the point of view, it will remain a fundamentally human point of view (Strawson 1985: 26–7), but he offers no reason to suppose that this is the only possible human point of view. Admittedly, it seems massively prevalent. However, given that we can and (at least some of us) do question this point of view, it seems that a human point of view should account for this human questioning, and provide some answer as to why human beings should stop their questioning. This is what the sceptic I have identified demands, and Strawsonian weak transcendental arguments cannot answer her. Strawson holds that the sceptic's questions are idle because it is a natural fact that we cannot sustain her doubt, but it is an equally natural fact that, even if not constantly, many of us *do* still entertain these

doubts, and have done throughout history. If the reasonableness of belief in that which the sceptic doubts stems from its hanging together as a part of a natural, human point of view, it must be equally reasonable to entertain these doubts from time to time, *unless* we invoke something stronger than this natural fact, as with Stroud's argument that we *must* believe some things even it is does not follow from this that those things are true. In which case we are right back in the disturbing situation described by Burge.

## 2.9 Conclusion

In this chapter I have examined the nature of transcendental arguments from their origins in Kant, through their development into strong and weak transcendental arguments during the twentieth and twenty-first century. I have argued that what is useful against the sceptic in transcendental arguments is that they start from a premise concerning thought or experience, which the sceptic does not doubt, and they move to show that something must be the case because it is a necessary condition of thought or experience. Although I argue that transcendental arguments are not necessarily anti-sceptical, they have a clear application if it can be shown that something the sceptic doubts is a necessary condition of thought or experience.

I have furthermore argued that strong transcendental arguments are to be preferred in offering a robust response to the sceptic who concerns me, as weak transcendental arguments have been shown to be either ineffectual, or, if they did work, they would actually land us in a situation even more worrying than the one the sceptic was originally concerned about. Namely, that it might be impossible for us to believe the truth about the world if a sceptical scenario does hold.

In the next two chapters I will go on to defend the plausibility of strong transcendental arguments. In chapter three I will defend the principle that my sceptic must accept at least some propositions concerning the mental to be certainly known. I will then go on in chapter four to argue that Stroud's criticisms of strong transcendental arguments rest on the illicit assumption of a metaphysical picture that is less plausible than the one that would need to be true for a strong transcendental argument to be viable.

## Chapter 3: What's So Special about the Mental?

### 3.1 Introduction

The power of a transcendental argument comes from that fact that its first premise is something prized and undoubted by the sceptic. The hard labour of the strong transcendental argument is to show that something the sceptic doubts is necessarily entailed by that which she already accepts: the first premise, an undoubted mental proposition. Why should she cede this much?

It may seem that even if a transcendental argument works, the sceptic can easily shift her ground to the mental premise. She may say: 'True, *if* you have experiences and thoughts of the sort you describe, I may be forced to grant that only certain sorts of world could give rise to such experiences and thoughts. However,' she will add, 'I do not grant this. How do you know that the content of your thoughts and experiences are just what you think they are? What's so special about the mental?' The first task of this chapter, then, shall be to defend the mental against sceptical attack.

I will examine a number of initially worrying doubts, and I will grant that some of these are successful in their targets. However, I shall argue that although these cast doubt on our confidence in our ability to know the nature and content of some of our thoughts and experiences, they do not motivate doubt that there are such thoughts and experiences, and that at least some broad features of our thoughts and experiences remain untouched by such challenges. It may seem that this is not a very reassuring response—that the broad features of at least *some* of our thoughts and experiences can be known and believed with confidence—but the first premise of a strong transcendental argument need not, in itself, be substantial and reassuring. It is only intended as a starting point, from which an STA can be given so as to provide us with something stronger.

I shall then move on to provide a positive argument for the incorrigibility of my beliefs about what I believe. I take care to distinguish this from Cartesian infallibility. The reasons why I know what I believe with an assurance I do not have for my knowledge of the external world are not connected to any assumptions about transparency of the mind in general or



metaphysical claims entailing the mind is of a fundamentally different kind of *stuff*<sup>41</sup> to other things. Many philosophers have sought to tease apart privileged access from its Cartesian heritage, and I will argue that there is no really compelling reason to suppose that the STA theorist must be lumbered with all the faults of Cartesianism simply because she will need to advocate a privileged access that bears surface similarities to the Cartesian spectre.

### **3.2 Doubting One's Own Mind**

The question of why the sceptic is not interested in casting doubt on first personal mental propositions as well as non-psychological propositions has recently received attention in response to research conducted in the field of psychology and neurobiology. Doubt has been raised about our ability to know our own emotions, our assessment of our own visual experiences, the phenomenology of thought, even our ability to recognise our own experiences as belonging to us. Eric Schwitzgebel, in particular, has taken these doubts as providing plentiful reason for supposing that introspection is in many areas 'highly untrustworthy' (Schwitzgebel 2008: 246), as unreliable as perception, if not more so. If introspection should prove as unreliable as perception it will not affect the possibility of some strong transcendental argument succeeding (at least in so far as one might show some non-mental fact to be as securely known as some mental fact) but it will certainly call into question the value of such an exercise. Perhaps my sceptic has no interest in doubting her own mind, but if I cannot demonstrate that she has some reason for prizing at least some mental propositions above non-mental ones it will be wondered why anyone should be interested in this sceptic at all.

In this sections 3.3–3.7 I will examine the reasons for doubting the trustworthiness of introspection. I will agree that in some cases we have reason to be less certain about what introspection can deliver to us. However, I do not accept the extent of the untrustworthiness Schwitzgebel ascribes, and I shall argue that we are not given sufficient reason to doubt

---

<sup>41</sup> I use the term 'stuff' so as not to limit this objection to talk of different 'substances', definition of which may be controversial. The term should be taken to include any kind of ontological categorisation that would define objects of different types so fundamentally distinct that one could not reason from propositions about one to propositions about another.

introspection where it counts: i.e. where it is needed for the construction of a strong transcendental argument.

I shall start by accepting Schwitzgebel's broad definition of 'introspection' as 'a species of attention to currently ongoing conscious experience', and 'the primary method by which we normally reach judgments about our experience' (Schwitzgebel 2008: 248)<sup>42</sup>. As will be seen later in the chapter, I will need to be much more specific as to the sort of introspection I wish to defend. However, this broad definition will do for the purposes of assessing criticisms of the kind that are intended to make us doubt introspection as a whole.

### 3.3 Emotion

Schwitzgebel's begins his argument against our knowledge of our own emotions by launching a barrage of questions:

Are emotional states like joy, anger, and fear always felt phenomenally—that is, as part of one's stream of conscious experience—or only sometimes? Is their phenomenology, their experiential character, always more or less the same, or does it differ widely from case to case? For example, is joy sometimes in the head, sometimes more visceral, sometimes a thrill, sometimes an expansiveness—or, instead, does joy have a single, consistent core, a distinctive, identifiable, unique experiential character? Is emotional consciousness simply the experience of one's bodily arousal, and other bodily states, as William James (1981 [1890]) seems to suggest? Or, as most people think, can it include, or even be exhausted by, something less literally visceral? Is emotional experience consistently located in space (for example, particular places in the interior of one's head and body)? Can it have color—for instance, do we sometimes literally 'see red' as part of being angry? Does it typically come and pass in a few moments (as Buddhists sometimes suggest), or does it tend to last awhile (as my English-speaking friends more commonly say)? (Schwitzgebel 2008: 249)

He anticipates that we will not be able to give answers to all these questions with equal speed and ease—'one-two-three'. It is tempting to suppose that some of his assumption of difficulty

---

<sup>42</sup> This seems broadly in line with definitions of introspection given by other authors; for instance, Bayne and Spencer (2010: 1) write: 'In this paper we use 'introspection' and its cognates solely to refer to a mode of first-personal access to one's current (or immediately prior) conscious states, episodes or processes, and we leave to one side uses of the term that apply to non-occurrent and non-conscious mental states (background motivations, dispositional states, etc)'.

rests with the fact that the answer to many of these questions (the first one, for instance) was ‘sometimes’ or, ‘in some cases’, and any initial hesitancy arises from the fact that although there are clear answers to be had, the questions need to be refined. However, when it comes to ‘Is emotional experience consistently located in space (for example, particular places in the interior of one’s head and body)?’ I find myself unsure. As a physicalist, I am confident that it *occurs* in space, but do I always *feel* it as specifically spatially located? What about the states, such as unconscious anger, that I want to confidently assert do not always have a phenomenal feel? Perhaps that should be my answer, then: if I do not feel it consciously at all, I cannot feel it as specifically in a certain place. Yet, what about those emotions I *am* sure are conscious. I am *pretty sure* that I think of them as being located in space—joy is in my head, my chest; fear is in my stomach—but is the trembling of my fingers a part of how my nervousness feels, or a mere consequence of it? I am not sure, and I must concede that the answer does not come to me instantly ‘one-two-three’, and it is not just because so many questions have been fired at me at once.

Schwitzgebel (2008: 250) is quick to say that it is not simply that language fails us. The question is whether the trembling of my fingers was really a part of the *phenomenology* of my nervousness at the time when I was nervous. It is not so much concerned with whether I should class the trembling of my fingers as a part of what it was for me to be nervous as whether I *felt* the trembling *as a part of* my nervousness. And that is a question to which I do not feel confident that I could give a swift and sure answer.

### **3.3.1 Anticipated and Remembered Emotion**

Schwitzgebel also calls us to consider the things in life that we really get pleasure or suffering out of (*ibid.*), and how often we misjudge these. He asks the reader whether she really enjoys Christmas. It is important to distinguish two questions so as to prevent a misleading answer. I am sure many people put on a show of enjoying Christmas when they do not, and many people complain of how Christmas is really a miserable season—that there are more suicides at that

time of year than any other<sup>43</sup>, and so forth. However, these would not seem to be examples of people who do not know whether they enjoy Christmas or not; these are examples of people who note a discontinuity between actual enjoyment felt and expectation, hopes, or conventions concerning what we are *meant* to enjoy. Similarly, I often find that I do not mind doing the dishes as much as I thought I would, but that does not mean that I do not know that I do not mind it when I am actually doing it. It means that I misremembered or overestimated how bothersome it would actually be when I was not doing it.

### 3.3.2 Linguistic Vagueness

Schwitzgebel also asks whether one might not be happier weeding than at a restaurant with friends. Again, I think this confuses the issue between conventionally pleasurable activities and awareness of actual pleasure felt. Many introverts feel that they ought to enjoy dinner out with friends more than they do, but this does not mean that they are not aware of how they actually feel when they are out with friends. Also, although I can report that working in my allotment is calming or makes me feel better, I am not entirely sure whether it would be correct to categorise my feelings when I am there as ‘happy’, or ‘pleasurable’. Certainly, it is not the same experience as that which I have when laughing at a good friend’s joke. This is more of a linguistic issue about the vagueness of such concepts. Many things that count as ‘happy’ can be quite different to each other. If I find myself uncertain whether one experience is more enjoyable than another, might not the problem be that I am comparing degrees of ‘happiness’ where the ‘happiness’ in question is in fact two quite different emotions? If I am 40% Happy-A when out with my friends and 80% Happy-B when weeding, does it make sense to say that I am twice as ‘Happy’ when I am weeding? Surely not. Say that Happy-A is ‘having fun’ and Happy-B is ‘contentment’ or ‘satisfaction’. We may think of all of these things as in some sense ‘happy’, but I feel satisfied by doing many things that are not exactly ‘fun’. Weeding would seem to be an obvious example.

---

<sup>43</sup> In fact, there is evidence to suggest that there are *fewer* suicides around Christmas, but the popular belief at least suggests that many people *feel* that it is not the happiest time of the year. See Sparhawk (1987) and Phillips and Wills (1987).

### 3.3.3 Distinguishing Moods and Emotions

In addition to a failure to distinguish different types of happiness, philosophers who employ counter-examples that challenge our self-reports of happiness often ignore that one may have more than one emotion at once. Thus, Haybron's 2007 paper repeatedly uses evidence that a person was stressed or depressed to counter self-reports of happiness as a part of his argument for 'affective ignorance'—or ignorance of one's own emotions.

Some of this relies on a rather slanted interpretation of a fictional man, Glen, who insists that he is happy whilst living a stressful city life, but upon visiting a friend in the country realises from his sense of relief (or 'decompression') that he had been unhappy, 'not because he didn't understand what happiness is, but simply because he was oblivious to the character of his emotional condition' (Haybron 2007: 395). This last is a characterisation of Glen's experience that Haybron surely cannot simply write into the example. It begs the question of how we are to understand Glen's experience. Haybron returns to this example surprisingly often, considering that it is an entirely fictional account and thus represents empirical evidence of nothing but Haybron's preconceived ideas. The entire section, '*Can We Tell Pleasure from Pain?*' (Haybron 2007: 405–6) relies on demonstrating how supposing that we struggle to tell pleasure from pain provides an explanation for Glen's faulty reporting, but the story itself is not enough to suggest real people report falsely in the way that Glen does.

In supporting his thesis that affective ignorance is extensive, Haybron (2007: 409–10) relies on surveys that over-emphasise a binary view of emotion, such as Andrews and Withey (1976), which purports to show only 3% of people reporting negative well-being when asked to pick from a scale of smiley faces how they would rate themselves. Haybron takes this as evidence of poor self-reporting on the basis that the incidence of depression and similar mood disorders is much greater than that. However, there has long been a call for more multifaceted assessments of subjective well-being, as advocated for by Ed Diener (1994). Moreover, Haybron appeals to symptoms of depression or anxiety as signs that a person's self-reports of being happy are false, but, as Beck and Alford (2009: 8) write, 'depression is a complex pattern of deviations in feelings' consisting of a 'cluster of signs and symptoms'. It is not defined as feeling sad all the time or never feeling happy.

Reporting that one feels happy is not necessarily wrong simply because one also has symptoms of depression or anxiety. Well-being, depression, and anxiety are complex experiences that cannot simply be contrasted to being sad-right-now or happy-right-now. Moreover, Haybron himself notes a difference between our reliability concerning focused states with a ‘more or less specific object or (phenomenological) location’ and ‘moods’, which ‘have no particular location or object’. In this sense, he distinguishes moods from ‘emotions’, which he holds to be more object-focused and intense, but, supposing this is correct, he fails to acknowledge that this difference could mean that moods and emotions which he treats as binary opposites *could* co-exist. The depressed person whose serotonin deficiency creates a general low mood, low energy, an inability to act on desires, etc., nevertheless feels joy when laughing at a friend’s joke, watching a favourite movie, and so on for other, specific, object focused events. It is for this reason that the widely-used PHQ-9 (for depression) and GAD-7 (for General Anxiety Disorders) employed by mental health professionals in primary care ask patients to assess their mood over a week or two weeks at the start of a session, rather than asking the patient how they feel at that specific moment (see Robert L. Spitzer, Janet B. W. Williams, Kurt Kroenke, and colleagues, 2010a and 2010b). At times, Haybron seems to recognise this, noting towards the end ‘mood is not a singular entity like that’ (2007: 414); it is puzzling that he does not seem to take this into consideration elsewhere—for instance where those with higher coronary activity are labelled ‘distress deniers’ (2007: 411), as opposed to people whose mood is not a singular entity.

Perhaps our self-reporting of moods is unreliable because it is vaguer and can be misunderstood because of presently felt emotions. Perhaps we are not perfect at assessing how we have felt in the past and how we are likely to feel in the future. This says nothing of presently felt emotions.

### **3.3.4 Presently Felt Emotions**

Schwitzgebel invites one to consider one’s current experience in reading his article, and unleashes another barrage of questions:

Is it completely obvious to you what the character of that experience is? Does introspection reveal it to you as clearly as visual observation reveals the presence of the text before your eyes? Can you discern its gross and fine features through

introspection as easily and confidently as you can, through vision, discern the gross and fine features of nearby external objects? Can you trace its spatiality (or nonspatiality), its viscosity or cognitiveness, its involvement with conscious imagery, thought, proprioception, or whatever, as sharply and infallibly as you can discern the shape, texture, and color of your desk? (Schwitzgebel 2008: 251)

Again, I feel matters of language confuse this. I am not sure, upon reflecting on my present experience of reading his paper, or writing my thesis, what exactly it is that I am feeling, but I feel fairly sure that it is more that I do not have a name for it, than that I am unclear as to what I am feeling.

To avoid this issue, Schwitzgebel invites the reader to contemplate some event that provokes a strong emotion, and this is easily done. I remember some event from my past that ‘riles’ me up (*ibid.*). I firmly report that I feel bitter, sad, angry. I turn to contemplate Schwitzgebel’s question: ‘Can you discern its gross and fine features through introspection as easily and confidently as you can, through vision, discern the gross and fine features of nearby external objects?’, and I must confess that I cannot. However, I am equally sure that this is because I do not quite know what Schwitzgebel means by ‘gross and fine features’ in the way that I discern the gross and fine features of the table. I never expected my emotions to be bounded or clearly demarcated in the way that the table is bounded and clearly demarcated—at least from the perspective of my meagre human vision. I know that from a finer perspective I would find it harder to delimit where the table ended and the air began. I am not sure in what sense my sensory apparatus are proving more reliable in reporting the gross and fine features of the table than my introspection is in reporting mental objects, except in so much as that the table itself is less vaguely delimited than my emotions. There is also an uneasiness in the comparison, as it seems to me that better sensory apparatus reveals the vagueness of the table’s *physical* boundary, whereas no improvement on my introspective capabilities could make the *concept* of calmness or happiness less vague.

Schwitzgebel may have a point in noting the disagreements in the literature on the experiential character of emotion, ‘disputes which at least *seem* to be about the emotional phenomenology itself’ (*ibid.*). I could perhaps insist that this is a mistake of reporting. I have never been convinced by arguments that start from noticing a divergence of opinions and go on to conclude from this that there is no fact of the matter—no one opinion that is true.

However, it is certainly evidence in favour of a lack of introspective *clarity* that my conclusions have not been easily reached by others.

The most convincing argument against the introspective clarity of emotions, however, is a familiar one from psychoanalysis:

My wife mentions that I seem to be angry about being stuck with the dishes again (despite the fact that doing the dishes makes me happy?). I deny it... But I'm wrong, of course, as I usually am in such situations: My wife reads my face better than I introspect' (Schwitzgebel 2008: 252)

This argument is convincing. Who has not at some point come to discover that whilst they thought they were fine, they really were not? One may be brought to see, either by an outsider, or upon later reflection, that, actually, one has been impatient, or raised one's voice, or been overly forceful in setting down an object, and these are signs that, in actuality, one is angry about something, or at least angrier than one thought one was. Schwitzgebel (*ibid.*) concedes that anger is not an ideal example, as it necessarily involves less favourable conditions for reflection than quiet contemplation. However, if one is not aware that one is angry, one may not realise that one is in less than favourable conditions in which to assess one's own reflective capabilities.

Nevertheless, it is worth noting that this point only brings into question emotions or moods that I am not aware that I have, not strong, object-focused, well-defined, presently felt emotions. Arguably, paradigmatic cases are to be found—agony and ecstasy—that we may know ourselves to be experiencing even if our general mood is vague and difficult to define.

For the majority of emotions and moods, I cede that I am not an infallible judge of my emotional experiences. However, I do not think there are many strong transcendental arguments (if any) that start from a proposition concerning these kinds of experience. Furthermore, I suspect that this is at least partially due to the fact that we are generally much less inclined to place emotional experience under the list of things about which we hold ourselves peculiarly sure. A great deal of literature (e.g. *Pride and Prejudice*, or *Gone With the Wind*), has been written centring around the fact that people often do not realise what their 'true' feelings are for quite some time. This suggests that our ability to read our own emotions with confidence is not actually all that commonly assumed.



### 3.4 Visual Experience

A much more worrisome challenge is that which Schwitzgebel levels against our visual claims. Take some familiar claim, such as that I am having a visual experience of redness. Schwitzgebel (2008: 253) is willing to grant that is ‘hard to go wrong’ about such simple visual cases as a patch of bright red, revealed to the ‘inward glance’. However, he maintains that it is equally ‘hard to go wrong’ about much of external perception. He notes a traditional assumption that sceptical hypotheses (such as that I might be dreaming, or that I might be a brain-in-a-vat (BIV)) do not undermine our judgements about our visual phenomenology. He challenges this assumption, noting that our judgements are often incoherent in dreams, and giving several examples to back up this point.

Over the next few sections I shall examine his examples, first from the point of view of what they might mean for a dreamer, and then from the point of view of whether they could be ‘massaged’ by neurosurgeons into the beliefs of a BIV. I argue that Schwitzgebel is confusing *introspection* with *judgements about introspection*, a point that emerges when one presses for clarity as to what exactly he is supposing in his examples. I think there is some sliding here between talk of introspection and judgements based on introspection, which is muddying the discussion. These are his examples:

Example 1: ‘I think I can protrude my tongue without its coming out’

Example 2: ‘I think I see a red carpet that’s not red’

Example 3: ‘I see a seal as my sister’

(See Schwitzgebel 2008: 258)

#### 3.4.1 ‘I think I can protrude my tongue without its coming out’

Example 1, ‘I think I can protrude my tongue without its coming out’, is an incoherent judgement, but it is not clear exactly what this is supposed to show. The example is under-described, and there are a number of ways one might seek to clarify what is happening here, but it is not obvious that they would cast doubt on one’s judgement.

The dreamer can be understood as making a linguistic mistake. Schwitzgebel mentions that such examples can seem quite insightful, and one way that might come to pass is this: I fancy I have come to understand some hidden way in which protruding my tongue could involve keeping it in my mouth. In this case, the incoherence lies in the dreamer either failing

to notice, or to remember, the meaning of the word ‘protrude’. To protrude is simply for something to stick out. I have failed to maintain an understanding of what it is for something to ‘protrude’ in English, so my statement is incoherent in English. The incoherence in terms of judgement, however, is illusory. What I have really done is to amend my understanding of what the word ‘protrude’ means to include not only those things that stick out, but also those things which do not. Perhaps it seems profound because some part of my mind is aware of the English meaning, and knows that it would be impossible for this sentence to be true in English. In my dream language, though, it is fine. My judgement that I have thought something profound is flawed, but then, I do not think that anyone supposed that judging one’s thoughts to be profound was an example of introspection. Rather, it is a case of comparing one’s thoughts to some objective (or, at least, intersubjective) standard of significance.

This is but one way to understand what might be going on in the example. Perhaps Schwitzgebel has in mind a case where the dreamer believes that they have discovered some secret working of the universe that allows protrusion without physical emergence—perhaps my tongue is protruding in some other dimension. However, it seems to me that this would be nothing more than a slightly more elaborate version of the familiar mistakes one makes about the structure and organisation of the non-mental world whilst dreaming. It is not a mistaken judgement about the content of one’s thoughts or experience based upon introspection; it is a mistake about the nature of that which causes those thoughts and experiences. I would be mistaken about my dream-experience, whatever that might be, being caused by my tongue protruding into ‘other dimensions’, but would not be mistaken about the experience I was having.

### **3.4.2 ‘I see a seal as my sister’**

The third example is not dissimilar, and I would argue it is even less convincing. When I see my sister as a seal there is nothing wrong with the introspective element of my judgement. I am having an experience as of a seal being before me, and I think that my sister is that seal. The use of ‘see’ is misleading, because it can refer to both my visual experience and to how I understand something—how I ‘see’ it. I am not incorrect about my visual experience as of a seal. I am incorrect in my understanding, because I believe that my sister is a seal. However, I am not incorrect in my introspection of what my understanding of the situation is. Again, the

introspective element of my judgement is not at fault. I *know* that I believe my sister to be a seal in the dream.

One might seek to make a more subtle point about beliefs here. If a belief is a disposition to assent to certain propositions, then I am wrong that in my waking life I am generally disposed to assent to the proposition that my sister is a seal. However, this does not undermine my correct assessment that at that time, in that dream, I am disposed to assent to the proposition that my sister is a seal<sup>44</sup>. I can introspect on this and deliver the correct assessment of my current disposition, even if I am unable to introspect and discover that *in general* this is not my disposition, that in the past it was not my disposition, and that in the future it is unlikely to be my disposition again. However, this is merely a manifestation of the fact that in dreams we believe many things to be true that we know to be false upon waking; it does not relate to our ability to assess correctly what our present beliefs actually are.

Furthermore, even if it did, it must be conceded that this example has strayed far from what Schwitzgebel was aiming at as regards introspection of visual presentations of bright primary colours. Our beliefs about seals and sisters are far more complicated and prone to error. I might form a belief about seals being fish because they swim in the sea, or that my sister is a Buddhist, and be wrong in a perfectly ordinary way that implies nothing about introspection. That one can make poor judgements of this kind about the nature of creatures outside of oneself in dreams does not affect one's ability to introspect correctly about the content of one's experience.

### **3.4.3 'I think I see a red carpet that's not red'**

What of the second example, then? 'I think I see a red carpet that's not red.' It certainly seems more to the point as regards casting doubt on our visual experience, and it also sounds like the sort of case that would cause profound doubt about the accuracy of our judgements about the colours we experience. However, its phrasing leaves much room for ambiguity. In this section

---

<sup>44</sup> There are those who argue that it is not possible to have beliefs in dreams, but I do not think they will have much interest in this discussion. If the sceptic thinks I might be dreaming and therefore have incorrect beliefs, she is clearly not of this point of view.

I will argue that, as with the other examples, it is from this ambiguity that the worry arises, not the possibilities that might actually be being expressed by this example.

We can understand the quoted sentence as:

- (A) a judgement that you are having a red phenomenology that has been caused by a carpet that is not red
- (B) a judgement that you are *seeing* a red carpet that does not present a red phenomenology
- (C) a judgement that you are seeing a carpet that is both red and not red
- (D) you are actually having a red phenomenology that is not red

A and B are judgements, and involve at least an apparent mistake, but are not incoherent. C is an incoherent judgement, but is more akin to the tongue protruding case, and I shall argue that it does not motivate the doubt Schwitzgebel is aiming at. D is incoherent, but I would argue that this means nothing as regards our dreaming judgements, as it is not a judgement. Its incoherence lies in the fact that it is *impossible*.

In A, what is going on is that the dreamer is having a red phenomenology that they judge has been caused by a carpet that is not red. The dreamer may do this for many reasons: because she believes she is subject to an illusion, because she believes the lighting is poor, and so forth—all the reasons why she might doubt a genuine sensory experience, along with a few others such as that she believes an evil witch has swapped out her eyes for magic ones that deliver false reports about the world. In none of these cases, though, is her introspection faulty. She judges that she is having a red phenomenology, and she is correct. She believes that it was caused by a carpet that is not actually red, but that is irrelevant to the quality of her judgement about her introspected experience.

B is simply the flip side of A. You judge that the carpet you ‘see’ is actually red, even though your phenomenal experience is not red. In other words, you believe that you are subject to an illusion or a witch’s prank, and although the carpet appears blue to you, you think your experience was caused by an object that was actually red. You are not wrong about your introspective experience of blueness.

C is more worrying. Something cannot be both P and not-P, and if I am judging that something is both red and not red I would appear to be judging contrary to the laws of logic. This could be akin to the tongue example, in that I might be mistaking what the laws of the universe allow. I judge that I am having a phenomenology as of that which would be caused

by a red carpet, yet *somehow* that phenomenology is also not red. This would again seem to me to be a mistake about the laws of the wider universe, not of introspection. I grant that it does not bode well for our powers of reasoning when asleep if we make judgements that contradict the laws of logic, but it does not indicate that we are wrong about our phenomenology. I do not see how, even in a dream, we could understand a person to be seeing something that is both red and not red, unless we are either saying that they really do have a phenomenology that is both red and not red, or they have a phenomenology that is red, and they believe that one can (and they do) have a red phenomenology that is also not red<sup>45</sup>.

The first seems to me to be an obvious impossibility, and, if it is not, then the person is not wrong—it would have to be the case that somehow, when asleep, we could have phenomenologies that are impossible to us when awake. In the second, it is as with the protruding tongue: they are not wrong about the phenomenology, they just also think that it is possible to have a red phenomenology that is also not red. The phenomenology they would in fact have would still be of red, they just somehow believe that the meaning of ‘red phenomenology’ or ‘red experience’ can include phenomenologies that are not red, even not-red-at-the-same-time-as-presenting-as-red. Introspection is not delivering to us a *phenomenology* of both-red-and-not-red—it could not—but we may judge that it is because we are understanding the meaning of ‘phenomenology’ somewhat differently in the dream.

This may not be obvious. It may seem that I am simply insisting that one cannot make a mistaken judgement about the content of one’s experience concerning bright red colours. In a sense this is true, it seems to me that I cannot really distinguish between my experiencing a red phenomenology and believing that I see red. However, I do not think it counts as a judgement, as there is no inference. To see red is simply to believe that I see red.

There is a secondary linguistic point, which I believe both supports this, and survives even if one is not convinced of it, which is this: it is not clear how I can be making a judgement that the content of my experience is ‘red’ unless ‘red’ is bound to a distinctive phenomenology. Furthermore, if the dream me is judging that she is experiencing something as red that is not an experience of red and using ‘red’ in such a way that is not tied to a

---

<sup>45</sup> Or the corollary: they have a phenomenology that is not red, but they believe that one can (and they do) have a not red phenomenology that is also red

particular phenomenology, then we revert back to the more basic linguistic point that she has ceased to use the word 'red' in the manner we commonly do in English.

Following on from this discussion, it should be obvious that I concede D's incoherence. I am sure, though, that this is not what Schwitzgebel meant by 'I think I see a red carpet that's not red'. He is seeking to undermine judgements made on the basis of introspection, rather than suggesting that we actually have impossible phenomenologies in dreams. However, I think it is the obvious incoherence of this from which the suggestion that 'I think I see a red carpet that's not red' gains its power to undermine our introspective judgements about visual experience. It would be impossible for introspection to deliver an awareness of both red and not red, yet this says nothing about any judgement that could be meant by 'I think I see a red carpet that's not red'.

#### **3.4.4 Introspection, Introspective Judgements, and Neurosurgeons from Alpha Centauri**

If what I say above is correct, it is not the deliveries of introspection that generate incoherent beliefs about dreams. It is not about the phenomenology that we are wrong. I think the confusion here stems from a slide in Schwitzgebel from talking about introspection as 'a species of attention to currently ongoing conscious experience' or 'the primary method by which we normally reach judgments about our experience' (Schwitzgebel 2008: 248), to talking about 'introspective judgements'. These are judgements that contain an introspective element, but, if what I have said above is correct, it is not solely on the basis of introspection that the judgement is formed, and it is not the introspective element that has gone wrong.

One might suppose that the dreaming example is misleading. Whatever else they may be, dreams are natural occurrences, and maybe elements of the way in which we form introspective judgements are left untouched by dreams that could be overwhelmed by artificial interference. Schwitzgebel (2008: 253) also considers the possibility that 'neurosurgeons from Alpha Centauri [might] message and stroke our brains' to create the judgement that something is red when it is not.

I think the analysis will go down the same. The reason for this is, as suggested above, that the phenomenology and awareness of basic things, such as bright colours, do not come apart. To have the phenomenology of red is to be aware of redness (at least when one is

attending to the redness, as Schwitzgebel suggests we are in introspection). It seems to me that what is different about introspection is perhaps what makes it different from straightforward judgement. To be aware of redness *is just* to believe that one is experiencing redness. It is non-inferential, and, in that sense, is not a judgement in the straightforward sense of being something that one judges to be true based on evidence. There is evidence of the experience of red, and it is as a result of that evidence that one forms the belief that one is experiencing red, but it is not judged in a way such that the poor reasoning of dreams or the messaging of neurosurgeons could interfere with it. The neurosurgeons can cause me to report that I am having an experience of red when I am not, but to make me believe that it is red, they will either have to change what I take the meaning of 'red' to be; or they will have to give me a phenomenology that is different to what should be given by what I am looking at; or they will have to make me believe that I have discovered some hitherto unknown aspect of the way the world is, such that in some unfathomable different aspect, red and (for example) blue can be the same, but this would be a linguistic mistake or a mistake about the *world*, not about my phenomenology.

Schwitzgebel also suggests that the neuroscientists might 'suppress' the phenomenology, whilst creating the judgement. He states that supposing that this is possible does not require that the phenomenology and the judgement be distinct existences, merely that the one is possible without the other. It is puzzling, then, in what sense they are not distinct, if they can exist without one another. This term 'suppressing' seems to be key, but needs fleshing out. It sounds more as though the suggestion is that the judgement could be 'messed' into us, and yet the alien neurosurgeons would *block* the phenomenology that would usually accompany anything that would cause that judgement. This does not undermine introspection directly—indeed, in this case, I might introspect (assuming I am not being forced to say what I do not believe) and report that I had no red phenomenology, yet firmly believe that I was seeing red. Yet it seems to me that if phenomenology and belief are not distinct existences, if, as I say above, to be aware of one's phenomenology is to have beliefs about it, one could not create the belief that *I have a red phenomenology* right now, if the phenomenology were suppressed. One would be aware of what phenomenology one did or did not have, and thus one would have beliefs about that. This example alone is not enough to drive a wedge between experiencing red phenomenology and believing that one does.

### 3.4.5 Blindsight

An important point should be made to distinguish these issues from what is known as ‘blindsight’<sup>46</sup>, whereby a person reports the absence of the phenomenology of something, yet they behave in other ways that suggest that in some sense they really do see the object. Perhaps some case of blindsight could arise wherein the sufferer reports that they do not see a red carpet as distinctly coloured, or claim that they could not say what colour the carpet is, yet they also deliver some suitable answer to a question as to whether my clothes match the carpet. This would be a case of blindsight, but it does not seem to me that it would be blindsight *about introspection*. My retina may be responding as normal to redness, and affecting that part of my brain that identifies what matches what (let us suppose), and the aliens may be suppressing my phenomenology of redness. In this case, though, *I am right* that I have no phenomenology of redness. It is just that it so happens that some other part of my brain is able to continue to operate coherently and produce red-appropriate behaviours that can be exhibited in the absence of red phenomenology. Blindsight is revealing as to how phenomenology can come apart from other brain operations that rely on perception, but it does not tell us anything about the reliability of introspection in reporting phenomenology.

For these reasons, I do not think Schwitzgebel is right to suggest that the sceptical hypotheses of dreaming and brain massaging provide us with reason to doubt our introspection of our experiences concerning bright colours. I concede that reasoning may well go wrong in dreams, and that this may be for a whole variety of reasons, but they are not reasons that relate to our ability to tell accurately what the phenomenal contents of our experiences in these cases are.

### 3.5 Peripheral and Foveal Vision

I now turn to the issue of *clarity* in our visual phenomenology in general. Various examples purport to show that we do not have as clear a presentation of our visual phenomenology as we generally take ourselves to have. Dennett (1991: 53–4) gives an example wherein one takes a playing card, holds it out to the periphery of one’s vision, and then slowly moves it

---

<sup>46</sup> For a detailed account of blindsight, see Weiskrantz (1986).



around towards the centre. He suggests the reader will be surprised at how close to the centre the card must be in order for us to distinguish its colour, suit, and value.

However, though it may be unexpected to the individual, much of what we know about how vision works should make this unsurprising. Our eyes saccade about an image or area of attention, focusing on various spots and presenting a composite; we do not simply *see* a detailed three dimensional picture of the world we inhabit. This results in other familiar oddities, such as ‘change blindness’<sup>47</sup>, wherein we remain oblivious to changes if they occur during an eye saccade, at places other than those at which our foveal attention is directed. It should be of no real surprise to us that very little is as clear as we commonly take it to be outside of points of foveal attention.

Although one’s phenomenal experience may be formed piecemeal and patchwork, this does not mean that the phenomenal experience is not of a (relatively) detailed three dimensional visual field. We are blind to change because our mind is ‘filling in the gaps’ with previously accurate data to form a cohesive image. As I look at my screen now, it is clear to me that the mouse mat on my right is red in colour, even if, in Dennett’s experiment, I might not be able to tell the colour of the card until it is within the area of my foveal attention. Similarly, I can say the same for the bright red bottle-cap on my left—the result does not seem to be simply a matter of size. Plausibly, then, I am not wrong in thinking that what is introspectively present to me is a relatively rich three-dimensional image. It is just that this image is a construct of my current perception and extrapolations from this and the recent past.

This is not to deny that the experience may be increasingly blurry towards the edges of my vision, nor that, *when called to consider the content of my current visual experience*, I may find that I am not nearly so sure about a great many things that I took myself to be sure of before. When called to attend to the periphery of the visual field, the brain may seek new information in preference to extrapolations from previous data and will thus replace the inner impression of completeness with something much vaguer, based more strongly on visual experience had *right now*. None of this is an indictment of my ability to correctly assess what

---

<sup>47</sup> Dretske (2004: 3) notes that ‘change’ blindness might more properly be termed ‘*difference*’ blindness, as part of the set-up of these examples often necessitate that we specifically *do not* see the change. If the change were given to us as a part of the visual field, we would note it (not least because our eyes are drawn to movement). However, he does also concede that we can fail to notice changes when they happen at the certain points during the eye saccade.

is introspectively available to me, or to form beliefs about the content of my introspected experience.

When Schwitzgebel considers this example he *seems* to concede this point. For he writes:

[I]t's one thing to mistake one's abilities and quite another to misconstrue phenomenology. Our visual experience depends on the recent past, on general knowledge, on what we hear, think, and infer, as well as on immediate visual input—or so it's plausible to suppose. Background knowledge could thus fill in and sharpen our experience beyond the narrow foveal center. (Schwitzgebel 2008: 255)

Yet he goes on to argue that when we prize apart foveation from introspective attention by fixing our eyes on a spot and drawing our attention to something on the periphery, most people confess that they have very little clarity beyond the centre of foveation. This, he claims, demonstrates that 'visual experience does not consist of a broad, stable field, flush with precise detail, hazy only at the borders' (Schwitzgebel 2008: 256). He notes that 'My interlocutors—most of them—confess to error in having originally thought otherwise' (*ibid.*); and based on this he concludes: 'most naive introspectors are badly mistaken about their visual phenomenology' (*ibid.*). However, this simply does not follow—Schwitzgebel has overstated his case.

As I have argued above, we can grant that *upon directing our attention in the manner Schwitzgebel describes* it becomes clear that one only has clarity for a few degrees out from the centre of foveation. Equally, I am sure interlocutors may be easily brought to confess that, *assuming the experience of attending away from foveation delivers the same phenomenal experience as naïve introspection*, they were in error in their original assessment of their introspective experience as rich, clear, and three-dimensional. However, it is not clear that we must accept that their ordinary introspection, filled out with the memory and extrapolation from previous eye saccades, delivers the same results as this artificially constrained case, where it is perfectly reasonable to suppose that *because you are trying to see what you can see based only on the present sensory experience when the eye is completely still*, the introspective experience *is different* to the usual case. These are two very different cases, and one simply cannot conclude that 'most naive introspectors are badly mistaken about their visual

phenomenology' when asking them to consider whether their normal visual phenomenology is at all like this special case under very specific and unusual conditions.

Two doubts can be raised against this. Firstly: regardless of whether the conclusion is correct, it is plausible that Schwitzgebel may have convinced his interlocutors that their original introspective beliefs were wrong. If we can be so easily convinced that we are wrong about the nature of our naïve introspection, then it may seem dubious to say that we were that confident in the first place, and it would seem that we are still easily convinced of things that are not true of the contents of our ever day introspection. Secondly, when one's target is a sceptic, stating that 'it is perfectly reasonable to suppose...' simply is not strong enough. The sceptic doubts many things that we prima facie take to be 'perfectly reasonable to suppose', so why should she not reject my supposition that there is, in fact, a difference between ordinary, 'rich' visual experience, and the visual experience one finds when one introspects having deliberately divorced attention from foveation?

On this point I concede. However, I do not think it is a very large victory for the sceptic. I do not think conceding that (outside the foveal centre) my visual experience is hazier than I naïvely take it to be shows that I was 'badly mistaken' about my visual phenomenology, and, more importantly, I do not think it endangers anything any STA I can think of would be likely to rely on. The clarity of my non-foveal vision does not seem to be of any great significance. Moreover, that non-foveal vision becomes blurry when I hold my eye completely still and attend away from foveation does not undercut the great deal of similarities that persist in what one took to be introspectively available before undertaking this exercise: objects appear to be similarly located, or the same size and shape, and even of the same colour (if one has seen them before—Dennett's example relies upon the presence of a novel object). There is a great deal at quite a high level of complexity that remains unchallenged by these cases.

### **3.6 The Phenomenology of Thought**

A similar point can be made regarding Schwitzgebel's arguments concerning the phenomenology of thought. He points out that there is widespread disagreement on whether there is a distinctive phenomenology of thought separate from the imagery that accompanies thought. Citing both the Chalmers and Hoy (2002) seminar for philosophers of mind at Santa

Cruz, as well as disagreement in the literature, he notes that even those of us who have reflected carefully on such things and engaged in research cannot come to an agreement about something as basic as the intentionality of thought. However, I do think the way he puts the case to us is misleading.

Schwitzgebel (2008: 258) asks: ‘the introspection of current conscious experience—that’s supposed to be easy, right?’ Well... yes and no. Certainly, if we assume a Cartesian ideal in which introspection of any aspect of conscious experience is supposed to be easy, then yes, we are likely to fail on a number of issues. Failure of obscure aspects of our conscious experience is cheap, however. As Schwitzgebel himself notes, it is hard to separate our thoughts from accompanying imagery, so why would it be easy to extrapolate whether there is a phenomenology distinctive to thought over and above any imagery that accompanies our thoughts? Surely any phenomenology of thought would be quite subtle in comparison to the rich and overlapping imageries that accompany thought; if it exists, we should not be surprised if it were overpowered by these others.

This is no very significant point in the overall attempt to discredit introspection, and having ‘a distinctive phenomenology of thought’ is unlikely to be the kind of content an STA is looking for to provide its mental premise.

### **3.7 The General Argument**

In addition to questioning specific aspects of our introspective capacities, some philosophers seek to form a gestalt argument from these instances, to a general sense of the unreliability of introspection. Schwitzgebel concludes on the basis of the doubts he raises, that ‘The introspection of current conscious experience, far from being secure, nearly infallible, is faulty, untrustworthy, and misleading—not just possibly mistaken, but massively and pervasively’ (Schwitzgebel 2008: 259). I think this sets up a false dichotomy. We do not need to demonstrate infallibility for all our introspective beliefs in order to show that we are not massively and pervasively mistaken.

There are specific features of our conscious experience about which we take ourselves to introspect regularly and accurately. There are other aspects of our conscious experience about which only philosophers of mind commonly make much attempt to introspect, and for which the deliveries of introspection are much less clear. I do not think it is surprising that

these areas should prove less secure than the simpler, more common ones, and I do not think our failures in these areas do anything to knock our confidence about the paradigm cases of introspection.

I have shown that Schwitzgebel's arguments against the paradigmatic case of bright foveal colour do not succeed, and his arguments against our knowledge of the phenomenology of peripheral vision are weak. He also considers pain as a paradigmatic case of introspection, and tries to throw doubt on this, by noting that we sometimes have trouble distinguishing mild pains from itches, but this is unconvincing (see Schwitzgebel 2008: 259–60). Do I sometimes have trouble explaining to my doctor exactly what it is that I am feeling (i.e. a pain, an itch, an ache; whether it's a '3' or a '4' on a scale of 1–10)? Sure, of course, but this is a linguistic issue. It is a question of whether I have correctly *described* what I am feeling, so that the doctor will know how to treat me. That does not mean I do not know what it *feels* like. I know exactly what it feels like, even if, because of the vagueness of the concepts of pains and itches, I am unsure how to respond when my doctor asks me 'Is it more of a pain, or an itch?'

The arguments against the paradigmatic cases are thus weak. Moreover, even if I remain unsure as to *which* experience, thought, or belief I am enjoying, no argument has been presented that I am wrong about my *having* beliefs, thoughts, experiences in general. Indeed, this generally seems to be the case. The *content* of my thought or experience is rarely deemed as significant as that I have them at all. Or, at most, it is required that I have a *certain sort* of content: that my thoughts concern objective particulars, perhaps (see Strawson 1987). I do not think that Schwitzgebel's general argument from the failure of specific (and often the least secure) elements of our conscious experience has the effect he desires on our general security about naïve introspection. It certainly does not undermine the security needed by STAs about certain minimal, paradigm cases.

The foregoing sections constitute a defence against attacks on our introspective reliability. I now turn towards providing a positive argument for the claim that I do have just those beliefs I believe myself to have.

### **3.8 Cartesian Infallibility vs Incorrigeability**

Few philosophers these days would argue for the full complement of what might be termed 'Cartesian infallibility'—the exact definition of the Cartesian theory of mind varies slightly

from account to account, and it is debatable which aspects of it could truly be attributed to Descartes, but for our purposes, I shall consider Cartesian infallibility to include at least these two theses: that the mind is ‘transparent’, in the sense that I have access to all my thoughts and experiences; and that this access is such that I am able to know and correctly report on all my thoughts and experiences.

The widespread acceptance of theories of unconscious and subconscious thought have rendered Cartesian infallibility deeply implausible in most circles, but this alone does not suggest that we have no special authority to make knowledge claims about our own thoughts and experiences. Furthermore, as Stoneham notes in his 2004 paper ‘Self-Knowledge’, even Descartes conceded a more subtle form of fallibility that should not be neglected:

[M]any people do not know what they believe, since believing something and knowing that one believes it are different acts of thinking, and the one often occurs without the other. (Descartes, *Discourse on Method* iii (AT VI.23), quoted in Stoneham 2004: 659)

We can understand Descartes as meaning that there are likely many propositions that I believe that I have never taken the trouble to reflect on and hence come to know that I believe. The belief is merely the act of holding the proposition true, whilst knowing that one believes something is a second order thought that has the belief as its subject.

So, we must add the concession that to have a belief does not automatically entail that one knows that one has a belief. However, it does not follow from this concession that all sense of authority concerning our own thoughts is lost. Stoneham (2004: 659), following Williams (1978: Appendix 1) and Dennett (1978: 226), suggests that we consider ‘incorrigibility’ instead. ‘Incorrigibility’ can be defined as: necessarily, if you believe that you have an attitude, *A*, to a proposition, *p*, then you do have that attitude to that proposition ( $BAp \rightarrow Ap$ ). This moves the debate away from the thorny issue of transparency—there is no requirement that I know what all my thoughts and beliefs are at all times. The debate now shifts to the more interesting question of whether, if I believe myself (*B*) to possess an attitude (*A*) to a proposition (*p*) (for example, the attitude of holding *p* true) I could possibly be wrong in my belief that I hold *A* to *p*.

The denial of the incorrigibility claim is ‘it’s possible for me to believe that I have attitude *A* to *p*, and not have attitude *A* to *p*’:

$\diamond(BAp \ \& \ \neg Ap)$

This is the claim I must defend against.

### 3.9 The Argument for Incorrigeability

Tom Stoneham presents the following argument for incorrigeability.

The argument has two stages. The first starts from the Davidsonian premise that: one cannot believe that anyone believes that  $p$  unless one also takes them to grasp the proposition that  $p$  (and, indeed, one must grasp the proposition as well oneself, although in the first-person case the point is the same). Second, anything that is a necessary conceptual precondition for having a propositional attitude towards  $p$  will also be a condition upon believing that one has that attitude towards  $p$ . Third, anything that is a consequence of believing that  $p$  will also be a consequence of believing that one believes that  $p$ . It follows from this that any self-attribution of belief has ‘*at least* all the consequences and commitments of the mental state attributed’ (Stoneham, 2004: 663).

Stoneham does add a clarificatory comment on premise three to the effect that the relevant consequences will be in the form of inferences and actions, and they will be *normative* in nature. That is, if I believe that  $p$ , and I believe that  $p$  entails  $q$ , I *should* believe that  $q$  as a consequence. However, I am not perfectly rational, and I may fail to do this. Stoneham notes this fact, but points out that although some non-conformity must be allowed for, we would surely admit that sufficient lack of conformity to the consequences of the state in question would be taken as evidence that I was not actually in the state after all. This argument does not require that I be perfectly rational, merely that, if I attribute a belief to myself I must demonstrate *at least* the consequences of holding that belief that I would if I did indeed hold the belief.

In support of premise three, Stoneham gives the following argument:

- i. To be committed to believing that  $p$  is to think and act as if  $p$  were true
- ii. So, if one believes that  $p$  one may infer anything that follows from  $p$  and act in any way that would help achieve one’s goals if  $p$  were true, and one may not infer anything that is inconsistent with  $p$

It is interesting that Stoneham has shifted from talk of what one ‘ought’ to do to talk of what one ‘may’ do. I presume this is to reflect the permissibility of not inferring all the propositions that follow from  $p$ . He then goes on to argue:

- iii. If one believed that one believed that  $p$  a rational believer would be equally prohibited from inferring anything inconsistent with  $p$  or ‘acting in any way that will only achieve her goals if not- $p$ ’, and ‘similarly for all other constraints’ (*ibid.*)
- iv. So, if one believes that one believes that  $p$ , one is committed to think and act as if  $p$  were true

I.e. one is committed to *at least* all the consequences of possessing the mental state.

The second stage of the argument is what Stoneham calls the ‘Containment Claim’. This assumes a ‘mild functionalism’:

- v. ‘If a thinker is in a state which meets all the conditions of conceptual grasp and involvement or activation for believing that  $p$ , [nothing] more could be required for it to be the case that she believes that  $p$ ’ (*ibid.*)
- vi. Having a higher-order thought to the effect that ‘I believe that  $p$ ’ is just such a state
- vii. Therefore, having the higher-order thought that ‘I believe that  $p$ ’ ‘involves or contains the state of believing that  $p$ ’ (*ibid.*)—i.e. it entails that I do believe that  $p$

So, this is good—this looks like it gets me what I want: necessarily, if I believe that I believe that  $p$ , I believe that  $p$ . This is the Incorrigeability claim.

However, one might worry that there is a difference between believing that not- $p$  and not believing that  $p$ :  $A\neg p$  and  $\neg Ap$ . Incorrigeability requires that I not only not have a contradictory belief, I must also not lack any belief at all about  $p$  in those cases where I believe that I believe that  $p$ . The difficulty with Stoneham’s argument from this point of view lies in premises two and three of the subsidiary argument for stage one, which read:



- ii. So, if one believes that  $p$  one may infer anything that follows from  $p$  and act in any way that would help achieve one's goals if  $p$  were true, and one may not infer anything that is inconsistent with  $p$
- iii. If one believed that one believed that  $p$  a rational believer would be equally prohibited from inferring anything inconsistent with  $p$  or 'acting in any way that will only achieve her goals if not- $p$ ', and 'similarly for all other constraints' (*ibid.*)

What if one were to act and infer in ways that were not inconsistent with  $p$ , but were also consistent with having no opinion about  $p$ ? One would not act as though one believed that not- $p$  or infer in ways that supposed the falsity of  $p$ . But premise ii only states that one *may* infer anything that follows from  $p$ , not that one *must*. If one should fail to infer things that one ought to if one really did believe that  $p$  this is presumably OK. Recall that I supposed that the change from 'should' to 'may' here was perhaps in consequence of the clarification that was added to the effect that it must be expected that we will sometimes fail to conform to the requirements on thought and action, as we are not perfectly rational beings. From this point of view it may have seemed that saying that we should always reason in line with the inferences we ought to draw if we did believe that  $p$  was too restrictive. But switching to 'may' seems too permissive. Stoneham does also state that 'talk of the commitments of specific beliefs does not commit one to the thought that these commitments are specifiable in advance, that they are not codifiable' (*ibid.*)—i.e. we should only speak in generalities—but I worry that this leaves too much room for the sceptic.

Plausibly, a reply to this point might be made that although some failures to reason in conformity with believing that  $p$  would be permissible, if I were completely indifferent to the truth or falsity of  $p$  then overall I would generally not be reasoning in conformity with believing that  $p$ , and would no longer be able to honestly assent to the higher-order thought that I believe that  $p$ . But what if I so happen to never be in situations where my lack of belief is revealed?

### 3.10 The Superstitious Man

Consider the following example: a man in 1762 might have heard that the Mayans predicted the end of the world in 2012. He may be of a superstitious bent, loving to collect old sayings

and mythology and claim them as beliefs, though he in actual fact believes very little of it, as may be noted by the fact that he behaves in an ordinary fashion most of the time. Plausibly, he will never have to engage in reasoning or actions based on the truth of a belief that the world is ending in 250 years' time. The date is too far off to have any significance to his present actions and thoughts. In truth, he neither believes, nor disbelieves it, but if you asked him about it he would claim it as a belief and he might even reflect on it in his own mind, wondering: 'Do I really believe that the world will end in 2012?' and mentally responding: 'Oh yes, of course I do!' with no thought of the consequences of that remark.

Now, it might be that if someone were to engage him in thorough questioning upon the matter, demanding that he examine the consequences of his beliefs, he would discover that he was really unsure what conclusions he should draw, and this would reveal his lack of any real belief. But until such a time it seems that he would retain his belief that he believed the world would end in 2012, and a sceptic could point to an example such as this as evidence that incorrigibility has not been established.

We should be suspicious of this example, however, and of any example that stems from similar motivations. For it looks very much as though what the Superstitious Man calls 'beliefs' are not standardly what we would call so. It is not simply that he does not act on his belief about the Mayan end of the world; he has no intention of doing so. He may call it a 'belief', but he seems to be using the word very differently to most people. Beliefs are not collected for their novelty, but possessed for the attribution of their truth. We might call instead this attitude an 'affectation', understood as something that causes one to expound its truth (perhaps with a certain amount of theatricality), but not be prompted to any further substantial action by it. His goal in claiming it as a belief is not to act upon any goals relating to goals one might have if one possessed such a belief; his goals relating to the proposition are concerned with things like gaining attention, feeling knowledgeable, seeming mystical. His goals in relation to espousing it as a belief are not the sort of goals one would have if one believed that one held that proposition true *not* because he genuinely believes that he holds a belief he does not actually hold, but because in saying he believes that he believes it he is not aiming at a truth that would guide his action in ways that relate to the content of the proposition, rather, he is aiming at evoking an emotional response from his listeners (himself included).

Recall that one of the requirements for attitude attribution was a proper understanding of what was being attributed. The Superstitious Man is either knowingly not using the word ‘belief’ as other people do, or he has such a fuzzy understanding of what counts as a belief that he cannot be said to successfully attribute it, for he does not know what it is he attributes. No case along these lines can be used to undermine the argument.

### **3.11 Capgras Syndrome**

A more concrete counter-example can be drawn from patients suffering from Capgras syndrome. These are people who maintain that someone close to them is an imposter. In the case of patient DS, interviewed by Hirstein and Ramachandran (1997), it was his parents. DS did not otherwise act as one might expect from a person who believed their parents to be imposters. He did not report the ‘imposters’ to the police or file a missing persons report. He did not attempt to flee, or react violently towards these imposters. Gregory Currie and Jon Jureidini have looked to explain Capgras syndrome on the basis that it, and similar delusions, are cases where one misidentifies an imagining as a belief: ‘The deluded subject fails to monitor the self-generatedness of her imagining that P’ so treats it ‘as if it were a belief, when in fact it is not one’ (Currie and Jureidini 2001: 160).

Two worries arise from this: 1) DS attributes to himself the belief that his parents are imposters, but he does not act as though he believes that his parents are imposters; 2) if Currie and Jureidini are right, we can mistake imaginings for beliefs, and attribute a belief to ourselves when what we really have is an imagining.

With regard to point 1, it is important to note that although the characterisation of Capgras syndrome given above is typical of how the case is presented when discussed by philosophers, it is not typical of Capgras syndrome itself. Thus, Rose, Buckwalter, and Turri generalise DS’s passivity to the matter and note, ‘Capgras patients typically continue to share their lives with the purported impostors... [DS] does not run away from these impostors; he does not call the police; he does not seek out lost loved ones’ (2014: 684). The generalisation from DS does not really reflect the evidence, though. Not all those diagnosed with Capgras syndrome show this apparent divide between professed belief and other action. Thus, Passer and Warnock report the case of Mrs D, who believed her husband to be an imposter, and who took specific and intelligible actions to protect herself from the intruder in her life:

She refused to sleep with the impostor, locked her bedroom and door at night, asked her son for a gun, and finally fought with the police when attempts were made to hospitalise her. (Passer and Warnock 1991: 446)

Moreover, Hirtsein writes in a more recent paper: ‘The majority of Capgras’ patients are suspicious of the “impostor” at the very least. Many are paranoid about the “impostor”, and attribute intent to harm to him’ (2010: 237).

Nevertheless, it may be thought that even one such case lends plausibility to what might otherwise have seemed implausible: that one might have a belief, but fail to act in accordance with possession of that belief. So then the question becomes: is it a good explanation of *DS* that he has the higher-order belief that he believes his parents to be imposters, but lacks the actual belief that he does?

Currie and Jureidini’s account of delusions as imaginings that have been misidentified as beliefs is based on the supposition that imaginings are both relevantly like beliefs, and that they are different in ways that account for the oddities of cases like Capgras syndrome. On their account, imaginings are like beliefs in that they mirror the ‘inferential characteristics’ of beliefs (Currie and Jureidini 2001: 159); thus, a reader may extrapolate causes and effects within the narrative, based on both the imaginings and actual beliefs. Thus, in a murder mystery, I work out ‘who done it’ based on what I have imagined about various characters, their access to weapons, their movements about an imagined place, as well as my knowledge of how real humans behave, what can kill them, and so forth. Imaginings differ from beliefs, they argue, in that they are ‘much more easily triggered by perception’, and that it is common for a rational agent to imagine all sorts of ‘wild hypotheses’ in response to odd experiences, but they are far less likely to adopt wild beliefs on such prompting. Imaginings are also not apt to be revised in the light of new evidence or contrasting beliefs, rather ‘beliefs inconsistent with the imagining move temporarily into the background’ (Currie and Jureidini 2001: 160). Thus, they claim, imaginings provide a good explanation of delusions because it is in their nature to ‘coexist’ with contradictory beliefs and be resistant to revision on the basis of evidence.

Currie and Jureidini then appeal to Stone and Young’s (1997) analysis that what is happening in Capgras cases is that the aspects of vision for overt recognition and for affective significance come apart. This means (the thought goes) that the sufferer may recognise that a

person *looks* the same without the accompanying ‘cognitive significance’—the sense of a person’s identity. Currie and Jureidini (2001: 161) interpret this as a reason to suppose that the sense of conviction a person feels when they have a belief can occur without the usual guide to action that accompanies a belief. However, it is not clear how the coming apart of the sense of conviction that *I have* a belief from that belief’s guide to action follows from the thought that overt recognition and affective significance come apart. Moreover, this is not how Stone and Young see it themselves. They conclude that there is no compelling argument that deluded individuals do not believe their delusions (Stone and Young 1997: 360). The effect of the affective part of face recognition being damaged in changing emotional response, they argue, is because the person they see *looks* like their loved one, but without the accompanying sense of recognition. From which, they conclude, it is a perfectly rational response to the anomalous experience to conclude that although the person looks like their loved one, they are not that person.

Besides this point, there is good reason to think that Stone and Young’s account is not the best explanation for Capgras. Firstly, over-reliance on a biological account of visual systems means that this account cannot explain the existence of Capgras in blind patients, such as that detailed in Rojo, Caballero, Iruela, and Baca, (1991). Hirstein, who originally interviewed the philosophically problematic patient, DS, has provided a different analysis in his more recent work, which better fits the symptoms of Capgras and other related disorders more generally. He argues that the damage that precipitates Capgras syndrome is to the part of the brain that understands and predicts the behaviour of others. It does so by constructing ‘individualised egocentric representations of her mind’ (Hirstein 2010: 243). These are constructed for specific, well-known people, based strongly off their actual behaviours, as opposed to the generic representations that allow us to get by with people we do not know particularly well. ‘Capgras’ syndrome occurs when egocentric representations of a particular person are damaged or inaccessible and are replaced by other, incorrect representations’ (Hirstein 2010: 243). This explains why Capgras delusions affect our (apparent) beliefs about loved ones, and not casual acquaintances, and are distinguished from general paranoid beliefs about all humans one encounters. However, Capgras syndrome does not affect our ability to recognise the person in other ways, ‘This creates the appearance of a familiar body and face, but without the familiar personality, beliefs, and thoughts’ (Hirstein 2010: 233). The key is: Capgras patients only fail to know their loved ones in certain ways.

A Capgras patient is able to function in limited ways as though they knew the person in question—they recognise the face/body, and may interact with them in some ways much as they would before—but in others (including asserting that they know them) they do not. This is because the belief that you know someone, or that ‘He is my father’, is complex, and made up of other beliefs—beliefs about what your father is like, how he behaves, what he looks like, etc. It is not the case that DS believes his father to be an imposter but acts as though this belief is false. DS believes his father to be in an imposter because the man no longer matches the individualised egocentric representation DS associates with his father—he no longer recognises his father *in the specific sense by which he identifies his father*. And he acts just as though he no longer believes the man to be his father *in that sense* by saying the man does not feel familiar and is not his father. But any other beliefs he might have had about the man that might cause him to behave as though he were his father *in other ways* could well persist.

What we have here are competing explanations of a phenomenon that *might* provide a counter-example to the argument for incorrigibility *if* the explanation for certain observed behaviours is that delusions are not beliefs, but can be misidentified as such. It is only a counter-example *on that explanation*. As we have seen, there are other explanations, ones that fit better with empirical evidence at this time. That might change in the future—the correct explanation for Capgras syndrome is not something that can be settled a priori. What is clear is that the sceptic cannot help herself to the assumption that Capgras syndrome *is* to be explained in this way and *therefore* constitutes a counter-example. The burden of proof is upon her.

Moreover, if ‘My parents are imposters’ is an imagining I have incorrectly identified as a belief, as evidenced by my lack of action in accordance with that belief, then one can ask whether I truly believe that I believe it. If the only way in which I act as though I believe that I believe it is to assent to the assertion that ‘My parents are imposters’, then I am failing to fulfil the rational requirements of someone who believes that she believes something. It looks like I am either failing to understand what it is to believe something, or my rationality has been sufficiently affected by my mental illness to prevent me from drawing inferences where I should. However, if my very inferential capacities are brought into doubt, then I must also question the inferences drawn from the apparent possibility of sceptical scenarios (that I might be a victim of something like Capgras syndrome, that I could be imagining all this, etc. etc.). This is not a path my sceptic can intelligibly go down.

For these reasons, I do not think my sceptic has reason to doubt incorrigibility based on the case of Capgras syndrome.

### **3.12 Conclusion**

I have shown that, although Cartesian infallibility cannot be claimed, this is not reason to suppose all our beliefs about mental states to be as dubiously justified as our perceptual beliefs. Where the mental state is vaguely delimited, not very intense, and where the words we use to label it are broadly defined, I concede that it can be difficult to know whether I am in that state or not. However, intense, clearly defined experiences, such as being in agony, do not seem possible to doubt. The phenomenal experience of bright, foveal colour also remains untouched by introspective sceptical arguments. And where I believe that I believe something, so long as I do properly understand the proposition I claim to believe, it is necessarily the case that I do believe it. I can therefore know that I have the beliefs I have, where I believe myself to have them.

Knowing solely that I have beliefs is enough to justify the first premise of the Davidsonian triangulation transcendental argument: that I have beliefs. In addition to knowing that I have beliefs, there seem to be at least some cases in which I know that the content of those beliefs is true (bright, foveal colour; pain). The above is not intended to be a complete catalogue of such beliefs, but should be enough to show that at least some aspects of the phenomenology of our experience can be certainly known.

## **Chapter 4: Mind in World—Strong Transcendental Arguments Without A Chasm of Necessity**

In chapter two I outlined a number of different ways in which transcendental arguments may be defined, and how these arguments have been used to either rebut or dismiss the sceptic. I noted that a defining attribute of transcendental arguments is that they start from a premise concerning thought or experience and move towards a conclusion that some proposition must be true because it is a necessary condition of that premise. That the first premise should concern something mental—thought or experience—is important because we commonly take ourselves to have better justification for our beliefs concerning the mental than we do beliefs concerning the extra-mental. This is particularly important in seeking to combat scepticism because such a premise ought to be something that the sceptic accepts, and by showing the truth of the conclusion to be a necessary condition of the truth of the first premise, the sceptic may thereby be forced to accept something she initially held in doubt. In chapter three I defended the assertion that at least some propositions concerning the mental are immune from doubt by the type of sceptic that interests me, and are therefore available for use in the first premise of a transcendental argument. The advantage of using transcendental arguments against the sceptic, then, is that they start from a premise the sceptic accepts.

Where transcendental arguments diverge in type is in the type of proposition they claim to be a necessary condition of the truth of the first, mental premise. Since Stroud's 1968 article, 'Transcendental Arguments', and the development of his criticisms since then, most transcendental theorists have focused their attention on weak transcendental arguments—arguments that aim to show us that it is a necessary condition of some the truth of some mental proposition that some other mental proposition be true. I have argued that these are ineffectual in dealing with my sceptic, and even if they could be made to work, it would be in such a way whereby the situation would be even worse than the sceptic feared.

I therefore advocate for the use of strong transcendental arguments in offering the kind of robust response to the sceptic I have argued we have reason to want to give. Yet, strong transcendental arguments have fallen out of favour due to criticisms raised by Stroud. It will be my task in this chapter to defend the plausibility of strong transcendental arguments against these criticisms. I argue that these criticisms rest illicitly on the assumption of a metaphysical



picture that the sceptic explicitly does not assume: her position is one of doubt as regards the nature of reality. I then move on to make a positive argument for the plausibility of a metaphysical picture upon which a strong transcendental argument could be formed. I stress that I do not argue *directly* for such a metaphysical picture, as if this could be achieved it would negate the point of giving a transcendental argument. Rather, I substantiate the claim that Stroud's assumption of a different metaphysical picture is implausible, as a plausible picture exists upon which a successful strong transcendental argument could be found.

#### **4.1 Stroud, Transcendental Arguments, and the Verification Principle**

In his 1968 paper, Stroud (2000a: 13–20) analyses a number of transcendental arguments and diagnoses them as either containing a suppressed verification principle, or requiring one for validity. I.e. what they need, but (at least on the surface) lack is a premise that asserts the meaningfulness of a non-analytic sentence be dependent on its empirical verifiability. Thus, for Strawson's arguments in *Individuals*, Stroud claims that the missing premise is: 'if the notion of objective particulars makes sense to us, then we can sometimes know certain conditions to be fulfilled, the fulfilment of which logically implies either that objects continue to exist unperceived or that they do not' (Stroud 2000a: 15–16). As Stroud points out, the arguments would thus be dealing with the conditions of *meaningfulness*, not knowledge. As such, they leave the sceptic free to simply say that talk of that which is in question simply is not meaningful: 'it is always open to the sceptic to accept the arguments and conclude that talk about, say, the continued existence of unperceived objects really doesn't make sense to us' (Stroud 2000a: 20). This, Stroud says, would only make scepticism stronger—not only can we not know whether these propositions are true, we would not even understand them. On this basis, Stroud argues that a successful anti-sceptical argument would have to be completely general, and deal with the necessary conditions of anything's making sense.

Thus, Stroud suggests that transcendental arguments must be such that they aim to show that the truth of their conclusion is a necessary condition of there being any experience or thought at all. They would thus be especially powerful against the sceptic: 'If the conclusion were true, there could be no experience to falsify it' (Stroud 2000a: 21). He therefore supposes that transcendental arguments will need to make use of a premise that asserts that the proposition which the sceptic doubts is a member of a privileged class of

propositions. These propositions are such that to be mentioned at all (as they are when doubted) guarantees their truth.

Stroud is doubtful that we shall actually find that that any of the propositions which are members of this 'privileged class' actually are the very same propositions that the sceptic doubts. He concedes that it seems difficult to demonstrate that there is no such proposition, but adds in favour of its unlikelihood the thought that the sceptic can always just say that it is enough to make talk of such things possible if we *believe* the proposition to be true (*ibid.*). He further asserts that any attempt to refute the sceptic on this point would involve a retreat to the verification principle, again, and if we can establish the verification principle, then we do not need the apparatus of transcendental argument to combat the sceptic anyway.

This does seem to be a necessary consequence. True, one might retreat to a verification principle at this point, and it may be that we can say of a number of transcendental arguments that this is what they (covertly) have done. Surely, though, on Stroud's definition, these are not really transcendental arguments after all. They aimed to show that  $x$  was the case because it is a necessary condition of  $y$ , but if they actually retreat to a verification principle, they are no longer offering a transcendental argument. As Stroud says, the verification principle could then do the job without the more complicated transcendental apparatus at its disposal. However, if one really were committed to arguing for  $x$  on a transcendental basis, one's reply to a sceptic who maintained that it was sufficient for  $y$  that one believe that  $x$ , would not be that a belief that  $x$  is only meaningful if it can be shown whether or not  $x$  is true, but rather that the belief that  $x$  *requires* that  $x$  be true.

Moreover, Stroud has artificially limited the scope of transcendental arguments to what I have called 'very strong' transcendental arguments—those where the very proposition doubted is shown to have as a necessary condition something which the sceptic must assume in justifying her doubt (i.e. something very general, such as that I have any beliefs at all). True, the first proposition of a transcendental argument will need to be something the sceptic cannot weasel out of accepting. However, I have shown that the sceptic can be motivated to agree that certain claims about the mental must be true without necessitating that the first premise be completely general.

Here, Stroud will step in with what, if successful, will be a more substantial criticism: that no strong transcendental argument can succeed, because it is not possible to cross the

‘bridge of necessity’ (2000b: 159) from psychological premises to non-psychological conclusions.

## 4.2 The Bridge of Necessity

I find this somewhat puzzling, as a criticism. It is difficult to see what the basis for the thought that we cannot move from psychological premises to non-psychological conclusions is, beyond mere incredulity. The case is not helped by the emotive terms Stroud uses when he discusses it. The phrase ‘bridge of necessity’ implies a yawning gap between ‘psychological’ and ‘non-psychological’ facts, and he says that its achievement would be a ‘truly remarkable feat’ (*ibid.*), calling such a feat a ‘truly remarkable bridge of necessity’ (2000b: 162). However, although these phrases express quite clearly Stroud’s incredulity, they do nothing to explain *why* it is thought to be so very implausible. It is tempting to think that there is nothing more to it than a strong intuition on Stroud’s part.

Stroud (2000b: 156–7) is quick to stress that he is not meaning to challenge that there might be a *causal* connection between our thought/experience/concepts etc. and the world—the physical activities of the brain and so forth. However, this is not the sort of connection we are looking for, and which he is dubious could exist; this appears to be a contingent connection, and we are looking for necessity. However, saying what will not be a solution is not the same as saying why no solution will be possible.

Stroud asks:

Even if we allow that we can come to see how our thinking in certain ways necessarily requires that we also think in certain other ways... how can truths about the world which appear to say or imply nothing about human thought or experience be shown to be genuinely necessary conditions of such psychological facts as that we think and experience things in certain ways. (Stroud 2000b: 158–9)

But again, this is not a *reason*. Stroud asks a rhetorical question, in response to which we are presumed to share his incredulity, but something more substantial is needed than this.

He indicates that these truths about the world are such as to appear to imply nothing about human thought or experience, but it is hardly a novel philosophical trait for it to not be obvious at first that one thing implies another. This alone cannot make it ‘truly remarkable’

that a truth that seems on the surface to imply nothing about human thought or experience should do so anyway.

### 4.3 True of the Mind vs True of the World

In ‘The Goal of Transcendental Arguments’, Stroud (2000d: 210) notes that we *can* move from ‘psychological’ statements to ‘non-psychological’ ones under special circumstances. For instance, if ‘Jed knows that the cat is on the mat’ then it is true that ‘The cat is on the mat’. However, as Stroud points out, ‘if would-be transcendental reflection is allowed to start from the fact that we know certain things, there would be no need for further enquiry’. Stroud does not go into details as to why, but we can surmise that, even though he concedes that the verb ‘to know’ is a psychological term, it is impurely so. If you know something you do not just believe it, you are correct in that belief. We have *smuggled in* the truth of some proposition about the world in such cases.

A question arises here: does Stroud forbid *all* uses of the word ‘know’ in transcendental arguments? What about the purely psychological premises along the lines of: ‘I have experiences’, ‘I have *x* experience’, ‘I have concepts’, ‘I have *x* concept’, ‘I have beliefs’ and so on. Presumably Stroud is willing to concede that the reason transcendental arguments start from such premises is that they are ones about which we do find ourselves able to speak with some authority. The truth of such premises cannot pass on to any conclusion if we cannot concede that they are true to begin with. Stroud does not want to forbid *all* arguments from psychological premises, just those that move from psychological premises to non-psychological conclusions. Does he mean to say that we can only have conditional premises of a psychological form, e.g. ‘*If* I have experiences, *then* I must believe that *y* (although it is by no means certain that I have experiences)’?

Stroud would not seem to be saying this. Recall from section 2.7.4 that in proposing a weak interpretation of Davidson’s argument from triangulation, Stroud wrote:

The possibility from which the sceptical line of thinking begins is therefore not one which anyone could consistently find to be actual. It involves the presence of certain specific beliefs, but attribution of those beliefs requires our finding them to be largely true. Believing a certain set of propositions to be all or mostly false precludes our

assigning them as contents of the beliefs of any people we find to have beliefs. (Stroud 2000c: 198)

We cannot get to the conclusion that we must believe that they and we are largely correct if there is not an additional premise to the effect that we *do* ascribe belief. This cannot be hypothetical, if it is to bite, it must be a truth or knowledge claim.

Here we reveal the problem with Stroud's somewhat nebulous categories of 'psychological' and 'non-psychological'. He seems to be drawing the distinction between 'true of the world' and 'about the mind'; but if he wants to motivate any argument, he must concede that his so-called 'psychological' premises are not just 'about the mind', they are '*true of the mind*'. To say that anything is *true* is to say that it is a fact, and facts concern the world. Here is how we motivate the thought that the mind is a part of the world without relying on some contingent causal claim: to say that the mind is a part of the world is not to assume that it is physical, that it supervenes on the physical, that it reduces to the physical, that it is caused by the physical, or that its experiences or thoughts are caused by the physical after some fashion—none of this is assumed. The world is that which exists, that which things are true of.

Stroud must concede this for his weak transcendental argument to go through. If he does so, though, his criticism of strong transcendental arguments collapses. There is nothing 'truly remarkable' in saying that one truth about the world entails another. The fact that the one truth is psychological and the other not can have no impact if this is no longer seen as a dichotomy, as an either/or situation. Just as Jed's knowing that the cat is on the mat entails something about the world because it is already saying something about the world, my knowing that I *believe* or *doubt* something about the world entails something about the world also, namely, that I believe or doubt that thing.

To summarize: Strong transcendental arguments aim to refute the sceptic by showing that an extra-mental proposition is a necessary condition of thought or experience: some aspect of thought or experience that the sceptic does not doubt. Weak transcendental arguments aim not to refute the sceptic, but to show that the doubted proposition is invulnerable in the sense that we could never find it to be false. Stroud urges that weak transcendental arguments are to be preferred because strong transcendental arguments are always open to the criticism that it is enough that we *believe* them to be true (thus requiring a verification principle to move from the thought that it is meaningful to assert this to the

conclusion that it is true). There is no direct evidence that all strong transcendental arguments must be open to this criticism, but Stroud argues that we have reason to think that they are because it is simply so implausible that one can move from psychological premises to non-psychological premises about the world. Not only is Stroud's incredulity insufficient reason to reject strong transcendental arguments, but his distinction between psychological and non-psychological propositions is, in fact, artificial. Any assertion of a psychological proposition *is* an assertion of a truth about the world.

#### **4.4 An Objection: the Psychological and Non-Psychological are Different Categories of Fact**

Keith Allen has raised an objection to this point that even if we concede that the world is the totality of that which things are true of, this does not entail that there are not different categories of facts<sup>48</sup>. It could be argued that a truth of economics just *does not* seem to be a truth of physics. Similarly, although truths about the mind may be truths about the world in some sense, but this does not mean that we can move from truths about the mind to other, non-psychological truths about the world.

However, I think this misunderstands the relationship between different categories of fact. In the case of the relationship between the mental and non-mental, conditions are imposed *on* the mind *by* the world, and it is in virtue of *this* that we can deduce that, as the mind is *x*, the world must be *y*. Furthermore, I do not think that categories of facts can be as exclusive as the objection suggests. Rather, what we have are overlapping filters. Physical facts *can* influence economic facts. Facts about the location of a country, its size, the molecular makeup of the different products available for sale there—these facts, even if they are not economic facts themselves, undoubtedly have an impact upon economic facts. So, even if we grant that mental and non-mental facts are of different categories, this does not mean that the one cannot be deduced from the other.

---

<sup>48</sup> Raised in conversation, 2009.

## 4.5 Mind in World

What does it mean to say that the mind is a part of the world, and that, therefore, there is no chasm of necessity to be bridged? In a trivial sense, even Descartes would accept that the mind is a part of the world. If the world is just the totality of existents, then if even a single, immaterial mind is instantiated, then that mind is a part of the world.

I do think there are some pictures of mind and world on which reasoning from premises concerning the mind to premises concerning the extra-mental would be problematic. However, I shall argue that these pictures of the relationship between the mind and the rest of the world are not the only ones available, nor the most palatable. Stroud is not free to assume that these are the case.

I will argue that it must be a picture in which the mind and the extra-mental world are not seen as conceptually and essentially distinct in the manner of Cartesian dualism. I will also argue that idealism is neither desirable for the strong transcendental theorist, nor strong transcendental arguments desirable for the idealist. I will conclude from this that what I need to defend is some non-idealist, monist picture of mind and world. I will argue that such a picture is physicalism, and furthermore, that supervenience physicalism can explain the picture of entailments from different categories of fact suggested above. If such a picture can be shown to be plausible, I maintain, the possibility that we may find some strong transcendental argument becomes equally plausible. Hence, I will therefore defend supervenience physicalism as a picture of the relationship between mind and world that is both coherent and plausible. This is not a proof of physicalism, which would negate the need for a strong transcendental argument. Rather, in defending the plausibility of a picture of the world that would allow for a strong transcendental argument to be constructed, I show that Stroud cannot assume a picture of the world upon which reasoning from premises about the mental to a conclusion about the extra-mental is *implausible*.

## 4.6 Substance Dualism

If substance dualism is true, I concede that we shall be hard pressed to claim knowledge of the physical substance by transcendental means.

A substance, as characterised canonically by Descartes, is something that can be understood as completely and essentially distinct from other substances. For Descartes, we know that mind and body are distinct substances because they can be clearly and distinctly conceived of as separate, and this separateness is founded in the essence of each entirely excluding the other. When I consider mental substance, he argues, I can see that all that belongs to it is thought, for I can conceive of it completely without reference to such material concepts as body or extension. Thus, he writes:

[O]n the one hand, I have a clear and distinct idea of myself, in as far as I am only a thinking and unextended thing, and as, on the other hand, I possess a distinct idea of body, in so far as this is simply an extended, non-thinking thing. (Descartes *Meditations*: VI.9)

The clear distinctness of these conceptions is taken to be demonstrated by Descartes's ability to conceive of himself (as a thinking thing) in entirety, without conceiving of his body; and his ability to conceive of his body entirely without conceiving of it as thinking. The ability to conceive of himself entirely as a thinking thing, and nothing else, is supposed to show that he is essentially a thinking thing, and nothing else. This is taken to be demonstrated by Descartes's ability to conceive of himself as existing when all he knows of himself is that he thinks : 'simply by knowing that I exist, and seeing at the same time that absolutely nothing else belongs to my nature or essence except that I am a thinking thing' (ibid.). Similarly, when one considers material substance, Descartes argues, one can understand it completely as 'an extended, non-thinking thing' (ibid.).

Of course, if this is accepted, we shall not be able to argue from facts about the mind to facts about the material world; the two will be mutually exclusive, as they will be known to be different substances based on their fundamental conceptual exclusivity. We would have to, as Stroud suggests, employ some verificationist principle, such as that God guarantees the truth of our clear and distinct ideas, that: 'he is no deceiver, and thence... all which I clearly and distinctly perceive is of necessity true' (Descartes, *Meditations*: V.15). Further, even supposing this, we would need to have some clear and distinct ideas about the physical world. It is not at all clear that we have such ideas, nor that an argument can be made in favour of a guarantor, like God, without begging the question against the sceptic.



So be it. If substance dualism is true, then we will not be able to construct a strong transcendental argument that does not rely upon a suppressed verification principle. However, the mere fact that there is some picture of the relationship between mind and non-mental world upon the truth of which it would not be possible to construct a strong transcendental argument does not show that no such argument is possible. Stroud's argument is that it is implausible that a strong transcendental argument might reason from a mental premise to an extra-mental conclusion. But substance dualism itself is held to be implausible in many quarters. If some other account of mind and world can be offered on which it is possible to reason from mental premises to a non-mental conclusion, this will be enough.

#### **4.7 Monism and Idealism**

From the discussion above, it seems to me that what follows from Stroud's objection is not that strong transcendental arguments are not possible, but rather that for a strong transcendental argument to be true, some form of monism must hold, at least in the sense that the entities which exist must be of a sort such that they are not conceptually exclusive in the manner of Cartesian substances. Such a monism could, theoretically, be idealism—arguing that the mind's being a certain way is dependent on the world's being a certain way is trivial, once it is granted that everything is mental. Yet, this is precisely why monist idealism and strong transcendental arguments can have no use for each other.

An immaterialist idealist, such as Berkeley, dismisses the sceptic on the basis that the existence of non-minded things simply consists in their being perceived—there is no question of whether they resemble real things existing independently of the mind, so no transcendental argument is needed. It is monist, in that 'there is not any other substance than spirit, or that which perceives' (PHK: I.7<sup>49</sup>), but this is argued for on the basis that ideas cannot exist apart from the thing which has them: 'for an idea to exist in an unperceiving thing, is a manifest contradiction' (ibid.). For 'sensible ideas' ('colour, figure, motion, smell, taste, and such like' (ibid.)) to exist is to be perceived, and he argues that nothing can be like an idea except an

---

<sup>49</sup> I refer to the 1998 edition of *A Treatise Concerning the Principles of Human Knowledge*, edited by Jonathan Dancy, using the abbreviation 'PHK'.

idea. How can colour be like something invisible? How can hardness be like something intangible? (PHK: 1.8).

An idealist in Berkeley's sense feels no need for a strong transcendental argument because he maintains that the picture of mind and world which generates the sort of scepticism which concerns me is nonsensical. He does not seek to move from premises about the mental to premises about the non-mental, as it is his contention that all entities are mental entities.

It may then seem to be a difficulty for me, if this idealist conception of mind and world is true. We could not argue from the mental to the non-mental, because there will be no non-mental (no non-spiritual). However, I do not see this as a very significant problem. That I might be a Berkeleyan solipsistic consciousness is precisely the kind of thing the sceptic worries might be true, but she cannot *assume* this. Her point is that we do not *know* whether this is the case. No case has been made for the *truth* of this situation and it is not widely held as a particularly plausible thesis.

#### **4.8 Transcendental Idealism**

Something similar goes for the transcendental idealist. The transcendental idealist may accept that the noumenal, at least, is extra-mental, but he will assert that the chasm between the mental and the noumenal is too wide to ever be breached. It cannot be known, as what we can know is limited to the way in which we, as limited beings, must organise the manifold of intuitions so as to make it intelligible.

For the transcendental idealist this is not a worrying prospect, for everything they are interested in establishing is of the phenomenal realm. We can reason between the mental and the phenomenal because the structure of the phenomenal is imposed by the mental. For our sceptic, it is rather more of a concern that we can never know about things as they are in themselves, but by the same token, this is not something the sceptic can assume, and transcendental idealism is hardly so popular a position that all others are deemed implausible. If another suitably plausible position it will not be enough for Stroud to point to transcendental idealism and say that 'Well, *if* transcendental idealism is true, we can never reason from premises about the mental to a conclusion about the noumenal'.

## 4.9 Physicalism

What follows from the above is that if I wish to claim that it is plausible that some STA can be found, I should argue that some picture of mind and world is plausible which is neither idealist nor dualist. It should defend the plausibility of a claim that the mind is made of the same kind of ‘stuff’ as the world, without resorting to the claim that all stuff is mind stuff. For the rest of this chapter, therefore, I shall defend the plausibility of a (at least a certain kind of) physicalism.

Note that it is not my purpose to argue that physicalism must be known to be true in order for some strong transcendental argument to be successful. Rather, I merely wish to make the case that if physicalism is true, it could well be possible that some strong transcendental argument can be made. I do not need to prove physicalism, for my purposes (that would negate the necessity for a strong transcendental argument), I merely need to show that the position is intelligible, and clarify how it would make a strong transcendental argument feasible.

### 4.9.1 Defining Physicalism

However, it has been argued by Crane and Mellor that physicalism is highly problematic. They characterise physicalism as the belief that ‘everything is physical’ (Crane and Mellor 1990: 185), or that there are no entities that are not physical entities, and, in particular, that all mental entities are really physical. Their thesis is that this position cannot be defined in any way that is clear, credible, and non-vacuous (*ibid.*). They point out that although physicalism is, historically, a descendent of materialism, we cannot make any appeal to matter. This is because the original advocates of materialism defined matter in an a priori way that has turned out not to reflect the facts of modern physics. Crane and Mellor take the seventeenth century definition of matter to be: ‘solid, inert, impenetrable and conserved, and to interact deterministically and only on contact’ (p.186). They note that: ‘the "matter" of modern physics is not at all solid, or inert, or impenetrable, or conserved; and it interacts indeterministically and arguably sometimes at a distance’ (*ibid.*). So, if this were the definition of what it is to be physical, it would be simply false that all objects are physical objects.

I think one may still say that the physicalist thesis contains an assertion that non-mental objects exist, and that the mind is made of the same basic ‘stuff’ as these. However, it

may be thought that to take this thought as a rebuttal is to miss the point of Crane and Mellor's paper. The title of the paper is 'There is No Question of Physicalism', and we can see them as not so much rejecting physicalism as arguing that it is a theory in response to a question that is no longer relevant. If materialism has been shown to be false, then we shall lose the motivations materialists held, and we are faced with a challenge as to why we should want to argue that all is physical. Crane and Mellor do not simply doubt that it can be done, they think the attempt cannot be coherently made, and see no need to try. That which is studied by physics may be such that we cannot offer an a priori definition for it. If this is the case, how could it make sense to pick the 'physical' as somehow fundamental.

From my perspective, the motivation stems from the thought that it should be coherent to assert that the non-mental both exists and that the mental is not of a significantly different sort of stuff such that one could say that a mental fact could not entail a physical fact, for the reason that in some sense they are not different kinds of facts. In order to support this, I need to be able to define the sense in which they are not different kinds of facts. I believe that supervenience physicalism intelligibly captures this sense.

The objection from section 4.4 was that we could at least understand the mental and the physical as different categories of facts in the same way that facts of economics might be said to be of a different category to facts of physics, and that we might consider mental and physical facts to be exclusive in this weaker sense. However, I wish to argue that in this weaker form, categories of facts can overlay one another—like coloured filters on the world. Later in this chapter I will go on to clarify how supervenience physicalism cashes out this metaphor so as to rebut Allen's criticism. However, first I must defend supervenience physicalism against Crane and Mellor's attack.

#### **4.9.2 Defending Supervenience Physicalism: Differing Mentally Whilst Not Differing Physically**

Crane and Mellor attribute to physicalists who have given up on the idea of an a priori definition the claim that the empirical world contains all and only what a 'true complete physical science' would say it contains (Crane and Mellor 1990: 186). The problem then, is: how do we provide a definition of what it is for something to be a physical science in a way that avoids making the physicalist thesis trivially true, without restricting it such that it is

obviously false? We cannot simply state that the subject matter of current physics will be the subject matter of a true, complete physics: this was the mistake of seventeenth century thinkers, and is obviously false, as much is still unknown, and we have good reason, from precedent, to suppose that some aspects of the doctrines of current physical sciences will turn out to be false. However, we also cannot simply include anything empirical. If psychology is included amongst the physical sciences, then physicalism will be vacuous. Or so the thought goes.

Crane and Mellor then examine and reject a number of possible answers to this challenge. As I am concerned primarily with supervenience physicalism, I shall focus on their objections to that. They characterise supervenience physicalism as the thesis that: ‘Two things will never change or differ in any way without also changing or differing in some non-mental way’ (Crane and Mellor 1990: 203); thus, they say the physical excludes the mental by being that upon which everything supervenes. Supervenience does not face the problems of reductionism in saying what the ‘real’ physical things are that everything can be in principle reduced to. However, they still find supervenience to be doubtful.

Crane and Mellor’s argument against supervenience is made on the principle that two thinkers could be alike in non-mental properties, and still have different beliefs. They argue for this based on a Putnam-like Twin Earth example (see Putnam 1973). My twin and I are ‘intrinsically’ identical, yet when she thinks ‘Water is wet’ she has a belief concerning XYZ, and when I think ‘Water is wet’ I have a belief concerning H<sub>2</sub>O. Crane and Mellor argue that she and I have the same intrinsic properties in this case in as much as we have the same ‘properties, same relations, same properties of the thing thought about’ (Crane and Mellor 1990: 204). However, this is simply false. Quite clearly, my relations have all been to water, up until now, and my twin, prior to this point, only had experience of twater (twin-water/XYZ).

Crane and Mellor (*ibid.*) anticipate that the supervenience theorist may say that thoughts supervene not just on the thinker’s intrinsic properties, but on her relations to other things. However, they do not think this is an effective response, based on the ways in which speakers can make mistakes. They suppose that my twin and I might look at the same tree, and she might think ‘That’s an oak’ where I would think ‘That’s an elm’—we would be relating to the same object, but having different thoughts. Again, though, this seems to be simply false. In one relation, the relation to the tree before us, we may be exactly the same (or as near to as

109

should make no significant difference), but we have had countless relations in the past with other trees that have been different. In learning the names of trees she must have been shown different trees to the ones I was when we were introduced to oaks, just as my twin was relating to XYZ and not H<sub>2</sub>O when she was taught the word 'water'.

Crane and Mellor (1990: 205) anticipate this response: 'past experiences do have present effects'. However, again they import the language of intrinsic properties into the supervenience theorist's defence, which seems unnecessary to me. They suppose that the supervenience theorist will say that, as these past experiences are at a temporal distance, they cannot have their effects immediately, and hence must be mediated by some present intrinsic properties of myself and my twin. Presumably what they mean by this is that there must be significant changes in my brain states that act as causal surrogates to the change in meaning caused by the way I was introduced to trees. However, this seems to be a very odd lesson to learn from Putnam. Surely the point was that not all meanings are 'in the head' (see Putnam 1973: 704).

Burge argues for this point in 'Individualism and Psychology':

Without the slightest conceptual discomfort we may individuate mental events so as to allow distinct events (types or tokens) with indistinguishable chemistries, or even physiologies, for the subject's body. Information from and about the environment is transmitted only through proximal stimulations, but the information is individuated partly by reference to the nature of normal distal stimuli. (Burge 1986: 17)

The causes for myself and my twin, locally, may be identical, but it does not follow from this that the mental events will be the same, because the way in which the mental events are individuated may rely on things in our environment, including events from our past. As Burge would say, Crane and Mellor have conflated causation with individuation.

The supervenience theorist is committed to saying that mental facts supervene on physical facts, not solely on 'intrinsic' physical facts about my brain or body. She is free to differentiate between facts about what is interacting causally with me locally and the facts that make my mental items the sort of mental items they are, without denying that the latter are facts about physical things. It seems perfectly comprehensible for me to say that it is a fact about me right now that when I was introduced to trees I was introduced to trees that looked such and such a way and told they were elms. (In fact, it seems plausible to me that the

connection of a different sound to the identifying features of what I call elms might intelligibly be recorded in minutely different ways in my brain than in the brain of someone who connected those images with the noise for ‘oak’, but even if it is insisted that my body has the exact same physical positionings of its particles, it seems to me that it will still relate to different events in the past to those of my twin.) My ‘intrinsic’ properties may not have changed, but there are still facts about me—physical facts—which have: precisely those facts which determine what sort of mental state I am in.

#### **4.9.3 Defending Supervenience Physicalism: Indeterminism**

Crane and Mellor’s next move is to say that mediating properties between past events and their mental effects need not differ just because their mental effects do. This argument stems from the thought that causation need not be deterministic, and, they say, ‘modern physics tells us that often it is not’ (Crane and Mellor 1990: 205). When causation is indeterministic, causes and effects will not supervene on one another. Crane and Mellor anticipate that the supervenience theorist will say that supervenience, unlike causation, is simultaneous, but they do not regard this as important. They offer the following argument to motivate this.

Suppose a non-mental property, *P*, causes a mental property, *M*, indeterministically. Suppose, further, that at *t*<sub>1</sub> many people share all their non-mental properties, including *P*. As the causation is indeterministic, many, but not all those who possessed *P* at *t*<sub>1</sub>, will possess *M* at *t*<sub>2</sub>. They then suppose that *a* and *b* are two people such that at *t*<sub>2</sub> *a* is *M* and *b* is not. Further, they state that their non-mental properties at *t*<sub>2</sub> may all be determined by their non-mental properties at *t*<sub>1</sub>. This is presumably because, for some people, it might so happen that whilst the non-mental property has, in this instance, been determined according to some probability outlined by the laws of physics, the mental property has not been so determined, as causation is fundamentally indeterministic. If this is the case, then it seems that we have a case of mental properties differing whilst there is no difference in the non-mental properties.

However, there are a number of things wrong with this argument. Firstly, this point about causation seems to have missed something fundamental about the supervenience thesis. The supervenience physicalist does not say that such and such mental events will always be caused when such and such physical events occur—supervenience relations are not causation relations. This is not, as Crane and Mellor seem to take it, a point about supervenience being

some special case of causation in which causes are simultaneous with their effects. Perhaps some supervenience theorists will want to say this, but it is clear that many will not. As Robinson (1991: 135–6) points out, Davidsonian anomalous monism will want to say that a physical event can cause a mental event, but only in so far as the mental event is itself a physical event, caused by other physical events (Davidson 1980: 223–4). Hence, we will not get cases where indeterminism in causation means that *a* and *b* are physically identical at *t*<sub>2</sub>, but differ mentally, because the physical events have followed the normal causal course, but for one of them the mental events have not. If there is indeterminism mentally for one, there will be indeterminism mentally for both, because for there to be indeterminism mentally will be for there to be indeterminism physically.

We can see this also in Papineau's 1990 paper where he expresses the doctrine of supervenience as being that 'mental particulars have physical features as well as mental features' (Papineau 1990: 66). He goes on to add that physical features determine those mental features (but not vice versa—this is the asymmetrical element of supervenience), but this is clearly not a causal determination, as the thesis so characterised is that the particular that has mental features is the very same thing that has physical features (see Papineau 1990: 67). The doctrine is consistent with token identity. It would also be consistent for a supervenience physicalist to posit that the mental is a property of a physical object.

The other thing that seems odd about this is the appeal to indeterminism as an explanation for our mistakes. It is true that modern physics tells us that, at the sub-atomic level, there is a certain amount of indeterminacy, but to acknowledge this is not the same as to acknowledge the rampant indeterminacy suggested by Crane and Mellor (1990: 205) where they write: 'Causation need not, after all, be deterministic, and modern physics tells us that it often is not'. Although causation would seem to often be non-deterministic at the sub-atomic level, it does seem fairly reliable for many larger objects—my interactions with everyday, medium-sized objects seem very much reliable. Moreover, indeterminacy seems entirely irrelevant as an explanation of what goes wrong when we make mistakes. When my twin mistakes an elm for an oak it does not seem that the correct explanation for this will make an appeal to quantum indeterminacy. My twin thinks elms are oaks because she has a different history to me, one which has determined that she thinks elms are oaks.



Accepting indeterminacy of the non-mental does not entail accepting indeterminacy between the physical and the mental, because the way in which the physical determines the mental need not be a causal determinacy.

#### **4.10 Supervenience, the Completeness of Physics, and Coloured Filters**

For these reasons, I argue that Crane and Mellor's argument against supervenience physicalism is ineffective. Furthermore, I think supervenience physicalism is, indeed, a very good way of understanding how different categories of fact can interact and entail things of one another. If supervenience physicalism were the case, then this would explain how it is that one might reason from the mental to the physical because, in some sense, they are the same kind of stuff, even though we can understand that stuff in different ways via different categories of fact. This is illustrated by the argument which Papineau (1990) gives for physicalism from the completeness of physics.

Papineau's argument goes like this: if physics is complete, then all physical effects are entirely determined by the physical events that precede them (Papineau 1990: 67). What follows from this is what Papineau calls the 'principle of no independent causal powers' (NICP): 'whenever any mental event causes another event, it does so entirely in virtue of its physical features, in the sense that the context of the effect is fixed entirely by the physical features of the mental cause'. If this is the case, then any mental event that has a physical effect must do so entirely by virtue of its physical features. Thus, although Eric Bristow's desire to win a game of darts, or for his dart to hit a particular number on the board, may be a good way of explaining his dart's hitting a double sixteen, the part that Eric Bristow's desire plays in determining the dart's trajectory is done by virtue of the physical features of his desire, the brain state that triggers his arm's movements perhaps. This must be true, or the completeness of physics will be false.

If it is true, we shall have a good way of understanding the differences between physical and mental facts. The mental facts may offer a good explanation of what happens. Indeed, in many ways they will be more useful. It is certainly more handy to know that Eric Bristow's having such-and-such desires tends to produce such-and-such actions. It is more predictively useful than knowing that such and such a brain state is likely to bring about those arm movements. This is in part because we do not have ready access to the map of his brain states; he is unlikely to play darts whilst inside an fMRI scanner. It is also useful because it seems

plausible that such desires are multiply realisable, both in Eric, and in other individuals, and we may take our predictive knowledge of the causal productions of desires into broader situations.

However, as Papineau (1990: 68) notes, the mental, unlike the physical is not complete. Even supposing psychological laws of some kind, it must still be true that not all mental events are determined by previous mental events. The pain caused by my stubbing my toe does not follow from some infinite mental train, it follows from my toe being stubbed: a physical event. Thus, we have reason to believe that although both mental events and physical events form explanatory categories of facts, it does not make sense to say that the mental is distinct from the physical. The mental ‘overlays’ the physical in the manner of a coloured filter. I.e. the same event can be differently presented, differently ‘coloured’, by speaking of its ‘physical’ or ‘mental’ features; but just as the shapes of things seen through a filter are determined by the way the world is, mental effects are determined by the underlying physical features of the world.

This also works, as I suggested in section 4.4, for other categories of facts. It is often more useful to see certain events in terms of economic features—buyings, sellings, recessions—but the causes and effects of economic events must be completely determined by the physical features of those events, if physics is complete. It is more explanatorily helpful, to talk of buyings and sellings and lendings and their effects on the recession, but these events all have their effects by virtue of the physical features on which they supervene—the movements of money, or the electrical signals which we also use to represent money when symbolised on computers.

We can thus see how the different features (economic, mental, physical) are all exhibited by virtue of the same thing, but they can also influence each other, just as coloured filters interact to change the appearance of something, whilst still being an appearance of the same thing. The economic event of paying bills causes the mental event of my worrying, and the broader economic facts, whilst not being facts about my state of mind, can colour an account of my state of mind so as to be more effective as an explanation of it. Thus, if it is a fact about me that I care about financial security, we can explain my worrying by reference to some facts about my financial situation. However, all these facts merely supervene on facts about the physical world. The completeness of physics will entail that a complete description of the physical features of my current situation will entail all the mental and economic effects.

Economic events can have mental effects because they both supervene on the same basic stuff; they both have the features they do by virtue of their physical features.

It may seem as though the above is missing an obvious problem, however. Namely: by what right do we assume the completeness of physics? We have not said what physics is, let alone what it will be for it to be complete. Will we not, then, fall into the same difficulties in defining physicalism by reference to some future physics, whilst also excluding psychology?

#### **4.11 What is ‘Physics’ and Why is it Complete?**

This is the thought that Tim Crane (1991) raises in his response to Papineau. Papineau, he notes, defines ‘physics’ purely in terms of completeness. It will be the science of ‘whatever properties are needed for a complete set of laws covering such effects as stones falling, darts hitting boards etc.’ (Papineau 1990: 70), and Crane labels this ‘PHYSICS’ so as to distinguish it from ‘physics’, the science currently engaged in by modern university departments which operate under that name. Crane (1991: 35) worries that if Papineau’s argument rests on a claim about the completeness of PHYSICS, it cannot rule out the mental in the sense in which it needs to do to have any bite. PHYSICS will incorporate whatever laws and properties are necessary for explaining archetypal physical occurrences, like the flying of darts, but that may include mental features of desires, beliefs, and intentions that are not had solely in virtue of the physical properties. If this is the case, however, the supervenience claim will be trivial, and would not seem to capture what the physicalist is after at all.

However, this criticism entirely misses Papineau’s point, as he explains in his (1991) response to Crane. Of course it will be trivial if mental categories turn out to be a part of PHYSICS, and difference in the mental will be difference in the physical, but Papineau does not assume this, and, indeed, he thinks it unlikely. He concedes that it may be that PHYSICS includes facts about the operation of telekinesis, or some analogue of Descartes’s theory that thought is involved in the directional determination of momentum in a way not accounted for by the objects of physics (see Papineau 2004). If so, Papineau will concede the claim as trivial. However, it is not trivial now, and, he thinks, based on empirical evidence (of for instance the way theories such as Descartes’s have been discounted by modern physics), that the inclusion of such things this will be unlikely.

Do I not need something stronger than Papineau's empirical claim, though, if I am in the business of defeating the sceptic? No. What I need is for physicalism to be an intelligible position, and it will be such regardless of whether the claims of physicalism are trivially true in the unlikely sense that Papineau dismisses, or if it is true in the more substantial sense in which the mental is not required to account for paradigmatically physical effects. If the mental supervenes on the physical because it is the PHYSICAL there will be no problem with the assertion that the mind is a part of the world in such a way that one can reason from facts about the mind to facts about the world. It will have been established that the mental is not distinct from the physical in the limiting way that substance dualism specified. If it supervenes on the physical because it is determined by the PHYSICAL, but not PHYSICAL itself, it will still be possible to argue from facts about the mind to facts about the extra-mental world.

However, to see this more clearly, we will need to get a bit clearer on in exactly what sense are the mental is determined by the physical. If the physical (or the PHYSICAL) is ontologically basic, and everything else supervenes on it, in what sense are these other things 'real' in a way that is neither illusory, as the eliminativist would have it, nor too substantial, as the dualist would have it. In this, I take Frank Jackson's line of thought, and argue for entry by entailment.

#### **4.12 Supervenience via Entailment**

Jackson (2000) looks to tackle what he calls the 'location problem' via the concept of entailment. The problem is that, if one wants to say that one category of things is ontologically basic (e.g. the physical), yet avoid eliminativist conclusions, there is then a problem as to where we 'locate' those things which supervene on that ontologically basic category. This is a question not just for the mental. Jackson (2000: 3) gives the example of solidity. Science tells a story about pens and chairs and rocks in terms of molecules in a 'lattice-like array' held together by inter-molecular forces, and this story does not mention solidity. However, it does not follow from this that pens and chairs and rocks are not solid. Rather, a proper description of the molecules and the forces that hold between them will include some account of their excluding encroachment, and this story about excluding encroachment will, in fact, be sufficient to entail those aspects of solidity not explicitly mentioned by the more basic

characterisation. It will entail that toes are stubbed and cups supported by tables—in short: it will entail all that we need from a concept like solidity.

This, Jackson says, is exactly how we should account for the physicalist's assertions about the mental, too. What the physicalist's assertion about the mental amounts to is the claim that a complete account of the physical will entail a complete account of the psychological, and this is the sense in which the physical determines the mental. Jackson cashes out this thesis in terms of possible worlds: 'Any world which is a minimal physical duplicate of our world is a duplicate simpliciter of our world' (Jackson 2000: 12). By 'minimal physical duplicate' he means a world that duplicates the physical aspects of our world, and no more, so that this definition expresses the thought that if you account for all and only the physical items of our world, you will have accounted for everything in our world, on the physicalist thesis; for another world that duplicated all and only the physical aspects of our world would be a duplicate of everything in our world—i.e. by duplicating only the physical aspects, that world would have nothing 'left out'.

Let us call this thesis 'M'. We can see that M follows from physicalism because if M were false, physicalism would be false. If something that were a minimal physical duplicate of our world failed to contain something that our world contained (e.g. the mental), then it would not be the case that accounting for the physical accounted for everything. Given this, Jackson argues, it is easy to show that, if physicalism is true, then 'the psychological account of our world is entailed by the physical account of our world' (Jackson 2000: 24). The argument goes like this: P is the (long and complex) proposition that provides a complete description of the world in purely physical terms, and which is true of the actual world, and of minimal physical duplicates, but false at all other worlds. Let Q be any proposition about the mental that is true at the actual world, and false at any world that differs in mental terms from our own. If M is true, any world at which P is true is a duplicate simpliciter of the actual world, and it follows from this that it will be a psychological duplicate, also. However, if this is the case, then every world at which P is true, Q is true. Hence: P entails Q.

It is worth noting that it does not entail that there is nothing over and above P. Dualism might be true at some of these worlds. One may wonder again: do not I need something stronger, otherwise, how will I know I am not in a dualistic world? Again: I do not know this. This is not an argument that we are not at a dualistic world, it is merely intended to clarify how we can understand the way in which the mental supervenes on the physical—it does so in

the way that solidity supervenes on the lattice-like array of inter-molecular forces. Ultimately, though, must I not argue that dualism must be false in order for an STA to be possible? Yes, at least in as much as for dualisms understood so that one could not reason from the mental to the physical.

Dualisms that allow for entailments between the mental and the physical will be fine, but it seems to me that they will be the sort of dualism that will sit comfortably with the 'substance' of mind being studied by 'PHYSICS'. Possibly not all of the mind substance would be available to study via PHYSICS, and it might be that not all aspects of mind would entail something about the extra-mental world. However, I only need for there to be some of the right kind in order for an STA to go through. In any case, even if STAs are not possible via dualisms that are included in this entry via entailment, all I need is that it be plausible that there could be worlds at which the mental is entailed by the physical in this way.

### **4.13 Multiple Realisability**

If we accept the above, then it is quite clear how entailment relations may operate between categories of facts, such that they may entail truths about other categories of facts, whilst still remaining distinct. However, one might be worried that the entire point of a supervenience relation is that it is one way: the physical situation entails the mental situation, but not vice versa. The mental is multiply realisable. This seems highly plausible to me. However, although I am willing to concede that the same thoughts I have may be realised by an individual with a slightly different neural structure, or even by some arrangement of computer chips, I do not think that my thoughts could possibly be realised by a shoe, or a cup of tea. There are clearly some constraints. One must not forget that the point of a non-idealist transcendental argument is not that the world must be a certain way because the mind is a certain way, but rather that the mind must be a certain way because the world is a certain way, and thus we may be able to reverse engineer the necessity constraints to see how the world must be to apply such constraints on the mental. It may well be the case that, although the same mental facts could be entailed by different physical facts, there is still a limitation on the physical facts that could entail these mental facts. There is a reason why physical fact  $x$  entails mental fact  $y$ , and not mental fact  $z$ . If the way the world is determines the way the mind is, then there is an explanation for why our mind is the way it is that would depend for its

explanatory power on the world being a certain way. If we could determine what the constraints are for the sorts of world that could produce the minds we in fact have, then we can know, on the basis of how our minds are, how the world is.

Perhaps a successful transcendental argument will deliver a range of different things that might entail the actual mental facts, but it will be a limited range. A range determined by the way the non-mental world must be, if the mental aspects of the world are such and such. If the sort of supervenience physicalism I have described is plausible, then it is plausible to suppose that the range of ways the world must be will include worlds where the extra-mental exists. And if we are in such a world, it seems unlikely that the same sort of mind could be produced by a world so different that it had no extra-mental entities. So it seems likely that the sort of explanation for why our minds are as they are will be such as to exclude the non-existence of the extra-mental, if we do exist in the sort of world the supervenience physicalist describes. Hence, it is reasonable that people in such a world might be able to give some strong transcendental argument for thinking that they are in the sort of world they are in fact in. And, given that supervenience physicalism is plausible, it is plausible that we may be able to find just such an argument ourselves.

#### **4.14 Conclusion**

This, then, is how I cash out the relationship between mind and world such that it is not implausible that there could be entailment relations between premises concerning the mind and a conclusion concerning the extra-mental world. It is a monist, physicalist picture. The physicalist picture so defined is not one rigidly specified a priori, but I hope to have shown that physicalism does not need any such problematic conception to be intelligible. It is a supervenience physicalist view in which the way in which the physical determines the mental is via entry by entailment. Mental facts are entailed by physical facts such that a full description of the physical facts would entail the mental facts. In a sense, there would be nothing over and above the physical, but we can explain the ways in which we categorise facts in relation to their explanatory efficiency and predictive effectiveness. Supervenience physicalism is plausible. The sceptic is not in a position to insist upon an alternate picture of the relation between the mind and the extra-mental world that would make reasoning in this

manner implausible. Given this, it is reasonable to suppose that some strong transcendental argument will be possible.



## Chapter 5: Metaphysical Necessity and the Laws of Nature

I have argued that a strong transcendental argument is needed to answer the sceptical doubt that worries me. I have answered the criticisms from those who do not think such an argument is even possible, and I have shown that there are some candidates to supply the first premise of such an argument: mental premises not open to doubt by my sceptic.

Stroud's complaint was that strong transcendental arguments need to, but cannot, provide a bridge from a premise concerning the mind to a conclusion concerning the world. I have argued that no mind-world bridge is required, as there is no chasm—there is only one world, sorted by interacting categories. Nevertheless, I must show how a premise concerning the mental can give us a conclusion concerning the extra-mental. My sceptic will not be satisfied with an answer that merely shows that such a picture of the world is *coherent* or *consistent* with the evidence available to me from the first-personal perspective. I must show how it could be demonstrated that such first-personal evidence would not be possible were the extra-mental world not a certain way.

It will be my purpose in this chapter to argue that what is needed for an STA to succeed against this sceptic is a metaphysically necessary psychophysical conditional that posits a natural law. As any such conditional will be highly contentious, I shall need to not only defend the possibility of giving such a conditional, but also how it could be that my sceptic could ever acquiesce to such a premise. I do so by arguing that all natural laws are metaphysically necessary, and that therefore whatever law governs the causation of the mental states that sceptic prizes as known is the only possible law there could be. Therefore, it is not possible that (as is the sceptic's contention) a sceptical hypothesis could equally well explain these items of thought or experience. In which case, what the sceptic's claim amounts to is not sufficient to justify the belief that such a proposition cannot be known. It merely reflects the fact that science is not yet complete—a far less worrying prospect.

This is a substantive claim in support of scientific realism, but not, I shall show, an unreasonable one. I start by arguing for the nature of this premise.

## 5.1 The Second Premise

The second premise will need to be known independently of empirical evidence. What does this mean? Primarily, it means it must be something that the sceptic would agree to, given her concerns that I might have all the mental life I now have in the absence of anything extra-mental. In this sense, ‘empirical’ is understood not in the vague sense of being any kind of evidence delivered via the senses—after all, it is my hope to show that many of our qualitative experiences are ‘delivered’ to us via a causal chain that starts with something extra-mental and ends with a qualitative mental experience, via the extra-mental object’s interaction with the senses. The point is that I must avoid assuming such a causal chain, for if such a chain can be assumed, then the purpose of giving a transcendental argument will be void, just as Stroud argued.

Beyond that, it must be a conditional which states that *if* what the sceptic concedes about the mental is true, *then* something concerning the extra mental must be true. In short, what we are looking for is a psychophysical law—one that posits a law-like relation between the mental and the extra-mental. My arguments of chapter four show that it is plausible that there should be such law-like relations. Given that the premise is not supposed to contain anything the sceptic does not concede to be directly knowable, this *prima facie* seems to be something of a tall order. The sceptic will say that no such law-like relation can be known a priori, and so no such premise is admissible to any counter-argument against her sceptical doubts. It was for precisely this reason that Stroud supposed that some further premise would be needed to explain how one could move from the mental to the extra mental.

It shall be the work of the remainder of this chapter to support the inclusion of such a premise in a strong transcendental argument, showing that such a move can be made consistent with the restrictions placed upon what can be directly knowable by the sceptic. Showing that the content of the argument can all be known from the sceptic’s position will demonstrate that no naturalistic premise is needed to act as a bridge, as there will be no gap *to* bridge. This defence will need to show that a psychophysical law can be metaphysically necessary, such that there will be no worlds at which the antecedent (which the sceptic concedes) is true but the consequent (which the sceptic doubts) is false. As such, I will need to provide an account of why a law of this kind should be regarded as metaphysically necessary, where it is often supposed that all natural laws are contingent.

## 5.2 Laws of Nature and the Humean Dictum

One's position on necessity and its relation to the laws of nature will depend on one's position as regards the Humean thesis that there is 'There is no object, which implies the existence of any other if we consider these objects in themselves' (Hume 1739: 1.3.6.1) and 'no objects have any discoverable connexion together' (Hume 1748: 1.10). This thought is often expressed as there being: 'no necessary connection between distinct existents'<sup>50</sup>. Hume is traditionally taken to be arguing for this in both *A Treatise of Human Nature* and in *An Enquiry Concerning Human Understanding*. The object of his critique is causation: that we can find no necessary connection between a cause and its effect, whether considered as events or as the objects that are commonly thought to be related such that the presence of one brings about the other or changes the other. Hume makes the case that neither relations of ideas nor the evidence of sensory experience can reveal a *necessary* connection between cause and effect, merely a 'constant and regular conjunction' (*ibid.*) of the one event with the other.

It is in relation to this that he defines the laws of nature as being those for which our assurance is of the 'highest certainty' in virtue of the cause and effect 'in all countries and all ages, to have been constantly conjoined', and as being 'a firm and unalterable experience... a uniform experience [that] amounts to a proof' (Hume 1748: 10.12). Hume's thinking is that we may place weighty trust in the laws of nature not because we have a proof in the modern sense (of a deductively valid argument) that the laws of nature *must* hold necessarily, but because all experience has shown that event A has always been followed by event B, and that this is the most complete evidence we can hope for on the matter.

This is not a level of 'certainty' that will satisfy my sceptic. Hume's contentment to define the laws of nature with reference merely to constant and regular conjunction, in the absence of necessary connection, offers no guarantee that cause will follow effect in the future

---

<sup>50</sup> It should be noted that the interpretation of Hume as sceptical of causation involving a necessary connection has been under debate. See, *The New Hume Debate: Revised Edition*, Read and Richman (2007). Some philosophers, such as Galen Strawson, argue that Hume 'believes in causal power, or "natural necessity", or "Causation"' (Strawson 2007: 31). Others, such as Read (2007) and Jacobson (2007), take Hume to adopt such positions for the sake of dialectical analysis without necessarily endorsing them. I take no bone in this particular debate. For me, the important thing is the impact that the traditional reading of Hume has had upon our general discussion of necessity, causation, and the laws of nature. The interpretation of Hume given here may thus be read as 'the traditional Hume'.

of *this* world, let alone for any other possible world, or for any possible world that may (unbeknownst to us) be the actual world. It would therefore seem that any attempt to run a strong transcendental argument incorporating metaphysical necessity between distinct existents (as I have argued that any successful transcendental argument must do) will have to be anti-Humean with regard to the laws of nature.

This is not that surprising. After all, Kant attributed his ‘Copernican turn’ to being inspired by, and reacting to, Hume’s work on causation. Hume’s argument was that necessary connections could only be established via a priori reasoning, for:

When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. (Hume, 1748: 7.6)

Hume’s point is that there is nothing in the content of experience that could count as evidence for a necessary connection—all that is present in experience is the experience of one event, followed by another:

In reality, there is no part of matter, that does ever, by its sensible qualities, discover any power or energy, or give us ground to imagine, that it could produce any thing, or be followed by any other object, which we could denominate its effect... [T]he power or force, which actuates the whole machine, is entirely concealed from us, and never discovers itself in any of the sensible qualities of body. (Hume 1748: 7.8)

So, necessary connection is not to be found in those qualities of an object, or the conjunction of objects, that can be sensed. It is equally, Hume argues, to not be found in the relations of ideas (the analytic a priori). These are ‘either intuitively or demonstratively certain’ (Hume 1748: 4.1). The intuitively certain are known because their contrary cannot be conceived (in contrast to matters of fact/empirical knowledge). Hume argues (1748: 4.9–13) that because we cannot predict what an entirely novel object (or event) will do on first encountering it—both some possible action, and another that completely contradicts it, are conceivable—any truth about how the object will react in conjunction with another must be due to its being a matter of fact, and not a relation of ideas: ‘For the effect is totally different from the cause, and consequently can never be discovered in it’ (Hume 1748: 4.10). Given that no evidence of

necessary connection is available in the empirical evidence that we have for matters of fact, Hume can only conclude that:

[A]s we can have no idea of any thing, which never appeared to our outward sense or inward sentiment, the necessary conclusion *seems* to be, that we have no idea of connexion or power at all, and that these words are absolutely without any meaning. (Hume 1748: 4.7)

Which is not strictly to say that no necessary connection exists, but rather that if it does, it is not discoverable to our limited faculties of ‘outward sense or inward sentiment’ (Hume 1748: 7.26). This, at first glance anyway, is all the sceptic needs. It is this *epistemic* constraint—that any necessary connection must be discoverable on *first encounter*—that I reject.

Kant’s transcendentalism (as outlined in chapter two) was designed to offer a third alternative to the failed analytic a priori attempts and synthetic a posteriori attempts of Hume: the synthetic a priori. It may, therefore, seem unsurprising that a strong transcendental argument would also be underpinned by a rejection of Hume’s dictum concerning necessary connections between distinct existents (or the lack thereof). However, there are key differences. Kant’s answer to how there could be a synthetic a priori was to propose a metaphysics in which the mind is directly involved in shaping the phenomenal realm, in which space and time operate not as independently existing dimensions, but rather as forms that must be imposed upon the manifold of intuitions in order for our human minds to make sense of them. Kant does not extend the a priori knowledge he thinks we may have of ‘synthetic’ matters to things as they are in themselves.

Our sceptic is unhappy with Kant on two fronts. Firstly, she finds the idea that we cannot know anything about the existence of things as they are in themselves worrying. To simply tell her that she should stop worrying because they definitively cannot be known will be read as merely conceding her worst case scenario. Secondly, the idea that space and time and the objects located within them are ‘merely’ phenomenal objects, shaped themselves by our minds, she finds deeply disturbing. It is her worry that we cannot know whether such is the case, as opposed to the common belief in mind-independent objects, that motivates the compulsion I have argued that many of us find to desire an answer opposed to this. The strong transcendental theorist, in turn, finds these thoughts similarly disquieting, but asserts that we can fend off the sceptic without resorting to transcendental idealism.

As we discussed in chapter two, some philosophers, such as Strawson, have tried to use Kantian arguments for the synthetic a priori in a way that avoids idealism. However, I have argued that these arguments fall short of any goal that would satisfy my sceptic, being placed to offer a weak transcendental argument at best. Which, given our concerns, is rather more worrying than less.

Therefore, if there is a proposition known synthetic a priori suited to play a role in a strong transcendental argument against my sceptic, it cannot be the synthetic a priori of Kant. Either some new form of synthetic a priori must be proposed, or the strong transcendental theorist must draw on something else all together.

### 5.3 Necessity, Analyticity, and the A Posteriori

Hume supposed that only analytic truths could be known a priori and only synthetic truths a posteriori. Kant proposed a third option: the synthetic a priori. Saul Kripke moved further on and has argued for a separation of necessity not only from analyticity, but from the a priori as well (see Kripke 1981, especially 122–6); whilst Hilary Putnam has provided reason to think that there might be an analytic a posteriori (see Putnam 1975 and 1981).

Kripke's thought was that for natural kinds, such as gold, there are a certain set of properties for which, were they not possessed by an object, the object would not be of that specific kind: 'Any world in which we imagine a substance which does not have these properties is a world in which we imagine a substance which is not gold' (Kripke 1981: 125). Therefore, in terms of necessity:

[T]o the extent that such properties follow from the atomic structure of gold, they are necessary properties of it, even though they unquestioningly are not part of the *meaning* of 'gold' and therefore not known with *a priori* certainty. (*ibid.*)

Kripke assumes that the analytic must be *a priori*, 'such statements are not known *a priori*, and hence not analytic' (Kripke 1981: 122) although he concedes: 'If statements whose *a priori* truth is known via fixing the reference are counted as analytic, then some analytic truths are contingent'. Kripke argues that natural kind terms have their reference fixed by rigid designation, rather than definite description. As a community, we have agreed that, *whatever that kind of stuff is*, all things that share the essential properties of that stuff will be gold; as

opposed to a definite description, such as ‘gold is a yellow metal’ (being yellow has turned out not to be an essential property of gold, there are other metals that are yellow). Kripke is hesitant to commit to contingent analyticity on this basis as he is distinguishing between reference and sense (see Kripke 1981: 122–3, n. 63).

Putnam does not share Kripke’s reticence. Rather, he argues for a ‘semantic externalist’ theory of meaning, which contends that (at least some) meaning is not ‘in the head’. Putnam proposes a thought experiment in which there is a twin of our Earth where everything is exactly the same, except that the clear liquid that runs out of taps and down rivers and fills the oceans and freezes at zero degrees centigrade at sea level and which is called ‘water’ by the people of Twin Earth is not H<sub>2</sub>O, but rather, XYZ. This shows that meaning relies on external factors in that it seems intuitive to say that if I were transported to Twin Earth and were introduced to the stuff they call ‘water’ and *I* called it water, believing that it was the same stuff as in my world (the stuff with the chemical composition H<sub>2</sub>O), I would be mistaken. Because my use of the word is determined in relation to the sample with which my reference to water was fixed and relative to the community of language users with whom I grew up, learning to mean by ‘water’ whatever it was that we, collectively, designated water.

This observation is interesting for our discussion because it means that the proposition ‘Water is H<sub>2</sub>O’ *is* analytic, given Putnam’s semantic externalism about meaning. It is part of the meaning of the word ‘water’ that it is ‘H<sub>2</sub>O’, but, crucially, we could not have known this a priori. We had to encounter water and conduct empirical investigations of it in order to discover its chemical structure. Once we know that water is H<sub>2</sub>O we can know that water is not XYZ, simply because XYZ is not H<sub>2</sub>O—it is analytic. But we would still need to conduct an empirical investigation to discover that the stuff that the Twin Earthers drink is *not* the same as what we call ‘water’, because we simply cannot know the full meaning of the term a priori.

Putnam does offer a transcendental argument against a certain form of scepticism: brain-in-a-vat scepticism. This is sadly not of use for my purposes for a number of reasons: a) Putnam’s argument is targeted at a form of scepticism that does not doubt the existence of an extra-mental world, it presupposes a physical brain and external influences upon one’s

perceptions; b) it is only effective against the recently envatted anyway<sup>51</sup>; c) it only rules out a scepticism of the kind that we can talk about, and not a state of reality like that of being a brain-in-a-vat that we are unable to talk about, which in many ways is a more worrying possibility<sup>52</sup> (this is similar to the criticism of weak transcendental arguments offered in chapter two); d) this reflects a general problem with transcendental arguments based on semantic externalism, which limits them to weak transcendental arguments, i.e. they can only entail what we are *able* to think or talk about, not what exists independently of us.

I therefore do not find fruit in Putnum's own style of transcendental argument. However, the discussion of natural kinds in Kripke and Putnam opens other avenues of thought.

#### **5.4 Dispositional Essentialism and Scientific Realism**

Reasoning along Kripkean and Putnumian lines has led some philosophers to think of natural kinds as *dispositionally essentialist*. Brian Ellis, for instance, derives his 'physical realism' from what he terms the 'new essentialism' (see Ellis 2006), by which he means the kind of essentialism exhibited by Kripke, Putnam, and the like—which is to say that natural kinds have an 'essence' that remains constant across possible worlds. Ellis believes that incorporation of new essentialism into scientific realism will allow us to avoid certain problems he finds in the programme of scientific realism explored by philosophers such as Stathis Psillos (see Psillos 1999).

In this section I will briefly articulate the scientific realist thesis, differentiate the programme of scientific realism that Ellis is reacting to, articulate Ellis's physical realism and the dispositional essentialism it incorporates, and explore how far this can help us towards our purpose.

Finding an agreed definition of scientific realism is fraught, as noted by Anjan Chakravartty: 'It is perhaps only a slight exaggeration to say that scientific realism is

---

<sup>51</sup> If someone wishes to argue that it applies to other brain-in-a-vat scenarios I wish them luck, but I take this claim as relatively uncontroversial and it is not necessary for my thesis to defend this.

<sup>52</sup> See Crispin Wright's arguments in Wright (1992).



characterized differently by every author who discusses it' (Chakravartty 2011: 1.1). Psillos (2006) identifies three theses:

*The Metaphysical Thesis:* The world has a definite and mind-independent structure.  
*The Semantic Thesis:* Scientific theories should be taken at face value.  
*The Epistemic Thesis:* Mature and predictably successful scientific theories are (approximately) true of the world. The entities posited by them, or, at any rate, entities very similar to those posited, inhabit the world. (Psillos 2006: 14)

The 'epistemic' thesis reads rather confusingly metaphysical—surely an epistemic thesis should concern itself with what we should believe there is or what we can know that there is, rather than what entities inhabit the world. We can interpret it as epistemic in the sense that it provides criteria for assessing truth—the success of predictions based on scientific theories.

The 'metaphysical' thesis, on the other hand, seems much too vague. Surely many theses about the structure of reality can be said to posit a mind-independent structure without thereby being *scientific* in any special sense. Yet, Psillos is keen to distance his understanding of scientific realism from any '*deep* metaphysical commitments'—particularly physicalism and non-Humean metaphysics (*ibid.*), so the minimal nature of what *he* takes to be the metaphysical commitment of scientific realism should not surprise us.

Ellis, by contrast, takes the metaphysical aspect of scientific realism as central—'It is... a primarily metaphysical thesis' (Ellis 2006: 1). What, then, is the primarily metaphysical thesis Ellis takes himself to share with Psillos and the other scientific realists with whom he differs on methodology? I take it to be the 'reality of the things and processes' described by the 'dominant theories of science'—which is what Ellis takes to be the conclusion of the Psillosian argument, though he holds the argument itself to be flawed (*ibid.*). This also fits roughly with a combination of the theses Psillos identifies as the metaphysical and epistemic aspects of scientific realism.

The programme of scientific realism that Ellis is reacting to, then, is one that privileges the semantic thesis that scientific theories should be taken at face value. This programme is understood as one that asserts that scientific theories are propositions made true by entities matching a literal reading of those theories (see Psillos 2006: 30). Ellis's concern is that asserting the truth of a literal reading of empirically successful scientific theories is insufficiently discriminating with regard to the type of entities asserted to exist. The Psillosian

programme of scientific realism argues from empirical success to the truth of the dominant theories of science—the reality of those theories—and Ellis finds this flawed as there is no ‘adequate theory of truth to carry the metaphysical burden’ (Ellis 2006: 1). Ellis’s problem is that the empirical success of individual theories cannot distinguish between real objects, events, and processes, and useful fictions— theoretical entities such as the Platonic objects of maths and geometry, idealised objects of model theories, Newtonian extrinsic forces. Nor can it distinguish between fundamental laws and those which can be eliminated.

Instead, Ellis argues in favour of what he calls ‘physical realism’. He posits a global approach in which physical things are the only real things, but such things are categorised via a hierarchical structure of natural kinds of objects/substances and events/processes. These kinds ‘all have intrinsic physical properties in virtue of which they are things of the kinds they are’ (Ellis 2006: 5). This is the sense in which physical realism is dispositionally essentialist— natural kinds are distinguished by virtue of intrinsic (essential) physical properties. The properties essential to these kinds must include causal powers, rejecting the Humean picture of logically independent events (see Ellis 2006: 5–6). It is the interconnectedness of theories, kinds defined by their essential causal powers, that allows for an explanation of the ontology of the scientific image. Theories are not addressed one by one, but compared with reference to their cohesiveness and usefulness to the whole.

For example, Carnot’s model of the heat engine is an ideal that could not exist in reality, as it abstracts away from causes of inefficiency, such as friction; yet Ellis notes that ‘Carnot’s theory is still the fundamental one in the area’ (Ellis 2006: 7–8). Carnot’s model of a heat engine is empirically successful, but its abstraction away from the limitations of reality provides good motivation for saying that it is not real—there are no actual Carnot engines. The ontological distinction between theoretical and physical entities is not one that Psillosian scientific realism is equipped to deal with. Yet it is a distinction worth preserving. The scientific image *as a whole*—as described by multiple, interlocked, successful theories— should be able to account for both the usefulness of the model *and* the fact that the real world (which reflects the existence of friction and other inefficiencies) does not include Carnot engines.

Similarly, Ellis notes that ‘Newtonian forces are always eliminable from physics’ in that, theoretically, ‘We need only combine the laws of their production with those of their combination and action to obtain laws of distribution of their effects’ (*ibid.*). Yet the

postulation of Newtonian forces is empirically successful, as in practice it will not always be possible to calculate effects based on constituent laws (due to the complexity of the factors involved).

This is the advantage presented by Ellis's global approach. Physical things are the only real things—the things whose intrinsic properties and essential causal relations can explain all real things—but an approach that incorporates a hierarchical structure of natural kinds can explain theoretical objects/events/processes in terms of supervenience. In this sense, Newtonian forces are less real because they are eliminable—the laws that incorporate them supervene on other laws which concern physical objects directly—but we may understand their role in successful scientific theories as a shorthand for laws whose application may be too complex for us to account for in practice. Similarly, although there may be, in reality, no Carnot machines, we can again account for the usefulness of their supposition by seeing facts concerning these machines as an extension of facts concerning less ideal machines in the world, and use that supposition to discern facts about the real world machines as an extension of principles concerning the ideal. The derivative nature of their reality—their supervenience on and extension from physical facts—explains ontological distinctions that are valuable, and which could not be accounted for by Psillosian scientific realism<sup>53</sup>.

Ellis's hierarchical picture of categorically distinct kinds with interacting causal powers, each a part of a global metaphysical picture, is very appealing. It closely fits my supervenience physicalist picture in which categories overlap and affect one another, but where categories can still be sorted hierarchically, with some (physical properties) being more 'basic' than others, as they are those upon which other properties supervene.

Ellis even offers something that could be developed as a transcendental argument:

The emergence of this scientific image of the world really has only one plausible explanation, viz. that the world is, in reality, structured more or less as it appears to be, and, consequently, that the kinds distinguished in it... are natural kinds of one sort or another, and that the causal powers they appear to have are genuine... The image is clearly a human construct. But it is a stable and revealing image... that seems to be indispensable. (Ellis, 2006: 12-13)

---

<sup>53</sup> Psillos argues that they can be accounted for, of course, and the debate between Psillos and Ellis is interesting. However, for my purposes it is sufficient merely to understand the motivation for Ellis's position, not to defend all the arguments in its favour.

Construed transcendently, the argument goes like this. The scientific image is indispensable to our understanding of the world. We do act successfully upon this image of the world. The only plausible explanation of this is that the world is really structured more or less as it appears to be, incorporating natural kinds with the genuine causal powers they seem to have. Therefore, an adequate account of the scientific image we do in fact have is only possible if the world is much as it appears to be.

Unfortunately for my purposes, this simply is not enough. It will not satisfy my sceptic. The whole structure is built on empirical knowledge. Perhaps the scientific image of the world is indispensable once we begin to conduct empirical investigations of the world, but my sceptic does not trust us to carry out those empirical investigations for any meaningful result.

Moreover, even if one could argue for the indispensability of the picture, one should pause over the fact that Ellis calls the image a ‘human construct’. I find his terms ambiguous; the scientific image is ‘a stable and revealing image... that seems to be indispensable’. ‘Revealing’ might be intended to indicate revealing of reality, but it is not obvious that this follows. The image ‘seems indispensable’, but if it is a human construct, how do we know it is not merely indispensable to humans? It might be merely that we must believe in such a construct, but then this would be a weak transcendental argument, and open to the criticisms we made before—that it is rather more disturbing than less if some beliefs might be ‘indispensable’ to us, regardless of whether they are actually true.

Here, we run up against the differences of aim between Ellis’s work and mine. Ellis states clearly: ‘I shall propose no concept of truth other than that of epistemic rightness’ understood as: ‘[in assessing] whether a proposition is true or not, we have only to consider whether the grounds for claiming it to be true are sufficient to justify the claim’ (Ellis 2006: 1–2). Epistemic rightness is ‘metaphysically neutral’, and justification is relative to the type of proposition in question. An empirical proposition, he maintains, can be decided on the basis of empirical evidence. But our sceptic is explicitly unhappy with that. The very validity of empirical evidence is what is in dispute.

This, in the end, will always be the problem of any appeal to the necessary or analytic a posteriori: it will always be knowable only *after* empirical investigation, and therefore cannot be of use to the strong transcendental theorist in her bid to defy Hume’s dictum.

However, I think there is another way.

## 5.5 Laws and properties

In his book, *Nature's Metaphysics: Laws and Properties*, Alexander Bird defends a position in which natural laws are metaphysically necessary by virtue of the fact that they supervene on 'potencies'—dispositionally essentialist properties. In this section I will outline the relationship between the nature of laws and the nature of properties articulated by Bird. I consider some reasons to think that Bird has oversimplified this relationship by examining the relationship between properties and laws articulated by Ellis. Nevertheless, I support the weaker claim that where at least some properties are dispositionally essentialist there will also be at least some metaphysically necessary natural laws.

Bird (2007: 1–2) outlines three broad camps in the literature regarding natural laws:

1. the regularity view<sup>54</sup>
2. the nomic necessitation view<sup>55</sup>
3. the dispositional essentialist view<sup>56</sup>

The regularity view takes a (traditional) Humean position in which laws are merely regularities in the relations of things and their properties: 'laws are regularities that fit into or may be derived from the optimal systematization of the facts concerning individual things' (Bird 2007: 1). In this sense, the regularity view represents a kind of supervenience: 'laws depend on the pattern of their instances and other matters of particular fact (things possessing properties)' (*ibid.*); however, this *Humean* supervenience is of no use to me. There is nothing about the supervenience of these laws on the facts of the world that guarantees the continued operation of these regularities, or any necessity relation that allows us to deduce the presence of B from the presence of A. On this view, a law does not even tell us that any regularity which can be found between a mental truth and the truth of an extra-mental proposition must *continue* to hold, merely that it has done in the past. The idea that one might reason from the plausibility of some world where this happens to any other world, let alone all worlds in which

---

<sup>54</sup> Attributed to Lewis (1973), although Bird notes an origin in Mill and Ramsay.

<sup>55</sup> Attributed to Armstrong (1983), Tooley (1977), and Dretske (1977).

<sup>56</sup> Origins attributed to Shoemaker (1984), developed by Swoyer (1982), Ellis (2001, 2002).

the mental truth holds, would simply not be possible to substantiate. Moreover, such regularities can only be known of via empirical investigation—investigation that relies on sensory information about the extra-mental world, which is precisely what is in doubt. Clearly, this understanding of natural laws is of no use to me.

The nomic necessitation view ascribes to natural laws a kind of law-like nature that is more than mere regularity, but which acts only within a possible world. This lack of cross-world, robust, metaphysical necessity will again not be satisfactory to our purposes. If natural laws do not hold across possible worlds, and if I do not know whether I am in a world in which an extra-mental fact, E, is nomically necessary if a mental fact, M, holds, then I cannot defend any premise to the effect of ‘If M, then E’ without first knowing precisely that which the sceptic doubts: that I am not in a possible world for which there are no extra-mental facts, E.

The third option, dispositional essentialism, holds that laws are robust metaphysical relations such that, at any possible world, ‘If A, then B’<sup>57</sup> holds by virtue of the nature of the properties A and B. Rather than laws imposing an order on properties, laws arise *from* the properties. This is exactly what we want—a theory that can explain why possession of mental property M necessitates the instantiation of property E, and which entails that it is true at all possible worlds that where M, necessarily E.

What theory of natural laws one ascribes to is linked to one’s view of properties. Bird distinguishes three options here, also:

- a) categoricalism
- b) the mixed view
- c) dispositional essentialism

Categoricalism is the view that properties are not distinguishable by any features that belong to them; rather they are distinguished by how laws relate them to other properties. Property A is thus only different to property B in virtue of the different ways it relates to other properties. Dispositional essentialism is the view that properties are distinguished by virtue of their potencies—i.e. qualities of the property such that it affects and responds to other properties in specific ways. As this is definitional for what it is to be that property, the relations for that

---

<sup>57</sup> It should be noted that this is a rough, provisional statement of the relations. Counter-examples from the literature (known as ‘finks’ and ‘antidotes’) will force a clarification.

property will hold at all possible worlds. Between these two poles is the ‘mixed view’, which holds that some properties are ‘potencies’, distinguished via innate relations, whilst other properties are ‘bare’—distinguished only by their relations to other properties.

The Humean ‘regularity’ view of laws is contingent in its account of laws and categorical in its view of properties. Bird’s dispositional essentialism incorporates both a metaphysical necessity view of laws and a potency, non-categorical, view of properties. Bird aligns the mixed view to contingent (nomic) necessitation.

It is not clear from the literature that all positions divide up this neatly—regularity view of laws with categoricalism and contingency, dispositional essentialism with potencies and metaphysically necessary laws, nomic necessitation with the mixed view. To gain clarity, it helps to distinguish the metaphysical from the epistemic issues.

Ellis’s view, grounded in the ‘new essentialism’, is surely dispositionally essentialist—as Bird himself notes. Ellis’s talk of the ‘causal powers’ of ‘natural kinds’ seems very much to be what Bird means by ‘potencies’ (see, e.g. Ellis 2006, 12–13, quoted above), and Ellis aims at a ‘metaphysical thesis’ (Ellis 2006: 1). However, Psillos (2006: 16) is surely right to point out that Ellis’s arguments turn on the explanatory power of physical realism, and that in presupposing ‘no concept of truth other than *epistemic* rightness’ (Ellis 2006: 1, emphasis mine) it is difficult for Ellis to get out the metaphysical conclusions that he desires. He describes the scientific (physicalist) image as ‘stable and revealing’, and therefore ‘indispensible’. I suspect an equivocation over the term ‘revealing’, between metaphysical and epistemic senses. One might interpret ‘revealing’ as ‘revealing of the nature of the world’, *or* one may see Ellis’s arguments merely supporting that the stability of the image is revealing about *ourselves*. It may be, as Psillos suggests, a purely epistemic enterprise, one which reveals something about what we must *believe*, but not about what *is*. This might be seen to fit better with Ellis’s concept of truth being epistemic rightness—it is what we need to believe, not necessarily what is.

Given this, I am not sure that Ellis’s philosophy supports a metaphysically necessary view of the laws of nature. *Epistemic nomic* necessity seems a better fit—the laws of the scientific image are indispensable to our understanding of this world we do in fact experience, but no argument has been given to extend this to the world as it really is, let alone that it is indispensable to how all possible worlds are.

The key point to take away from this, however, is not that Bird is wrong to suppose that there is a link between the dispositional essentialist view of properties and the view that the laws of nature are metaphysically necessary. Rather, some further argument is required to establish this; either dispositional essentialism must be augmented with some other claim, or further argument must be made to show that a dispositionally essentialist position entails metaphysical necessity for the laws of nature. Ellis does not need to make this claim. We may read him as *epistemically* dispositional essentialist, in the sense that our best beliefs tested against best fit with what can be said to be known support a belief in dispositional essentialism. This could be seen as falling short of the *metaphysical* claim that natural properties are non-categorical potencies—it is reasonable to *believe* that or say that it *suffices for knowledge* in ordinary contexts, but the defeasibility of his position cannot substantiate *metaphysical necessity*<sup>58</sup>.

What makes Bird's move interesting is his tying of epistemic dispositional essentialism to the metaphysical dispositionally essentialist view that properties are potencies. It is this, and his account of potencies, that he argues entails metaphysical necessity for the laws of nature. He therefore must defend the (metaphysically dispositionally essentialist) potency view against categoricism, and the metaphysically necessary dispositional essentialist view of laws against the regularity view and the nomic necessitation view.

*I also* need to argue for metaphysically necessary laws, and against categoricism, contingent laws, and contingent necessitation. I am not *convinced* that Bird's arguments against the mixed view are as powerful as he needs them to be. However, it should be noted that I do not need to defend what he calls 'dispositional monism' (the thesis that all properties are dispositionally essentialist potencies) over the mixed view. All I require is that we be able to know that *some* properties are potencies whose relationships to one another are such that where one is a mental fact that the sceptic accepts and one is an extra-mental fact that seems

---

<sup>58</sup> Given that Ellis sees his thesis as a metaphysical one, it seems likely he would resist this interpretation. It is not my purpose to give the 'correct' understanding of Ellis's position. I merely want to show that we could do with a more fine-grained account of dispositional essentialism and the relationship between one's view of properties and one's view of laws. Given the plausibility of the interpretation outlined above, one can see that an *epistemic dispositionally essentialist* view of properties is not necessarily committed to a metaphysically necessary dispositional essentialist view of natural laws. What we need is a *metaphysical* dispositionally essentialist view of properties.



prima facie to be something the sceptic can reject, elucidation of such a relationship would reveal that the presence of the mental property entails the presence of the extra-mental one.

With this in mind I will set aside Bird's discussion of the mixed view and focus on defending his dispositional essentialist view of properties (for at least *some* properties) and natural laws (for all fundamental natural laws). I shall first outline this position and examine the criticisms against it. I will defend it against these. I will then examine the nomic necessitarian view of laws and show that this is not appealing, and, further, that it is not available to the sceptic. Lastly, I will consider categoricism about properties and whether the sceptic can use this, in combination with the regularity view of laws, as an alternate picture of natural laws. I will conclude that she cannot.

## 5.6 Bird's essentialism

Bird's argument for a dispositionally monist essentialism focuses on establishing essentialist potencies for 'fundamental natural properties ... those with non-redundant causal powers' (Bird 2007: 13). In this section I will elucidate this definition.

I find some of these terms value-laden, but poorly defined in his text. 'Natural' properties are defined largely by example. They are natural in the sense that it is the task of scientists to discover them:

For example, a scientist may discover or synthesize a hitherto unknown molecule. It would be natural to say that her next task is 'to discover its properties'. In that sense, its properties do not include 'the property of being first synthesized on a Wednesday' or 'the property of being  $\Phi$ ' (where something is  $\Phi$  iff it is a member of the set {molecule *m*, the Eiffel tower, the power set of the natural numbers}). (Bird 2007: 9)

He also writes:

we may distinguish the liberal from the restricted use of 'property' by referring to 'abundant' and 'sparse' properties. Or we may talk of 'natural' properties to capture the restricted sense. I shall use 'sparse' and 'natural' interchangeably. (*ibid.*)

Which is as much as to say that restricted properties are sparse properties and sparse properties are natural properties, and liberal properties are abundant and non-natural. He also contrasts

simplicity (associated with sparse-natural-restricted properties) with complexity (liberal-abundant-non-natural). But defining these terms by reference to each other and to examples does very little to explain what it is that makes these terms apply to some things and not others.

However, I think this game of shifting definitions is best clarified when Bird turns to talk of *fundamental* natural properties—which he defines as ‘those with non-redundant causal powers’ (2007:13)—and the role these have in *explanations*. He gives the example of the property of being charged, and the property of being charged\*. The property of being charged\* has the extension ‘all charged things except some particular future proton’. It is a complex/liberal/abundant/non-natural property in that it combines another property (being charged) with something else (the exception of some future proton). To be charged is in this sense fundamental: it is not made from the combination of other properties. ‘Charged’ has non-redundant causal powers in that its causal powers cannot be explained by anything else. ‘Charged\*’ does not have that advantage—its causal powers are derived from the causal powers of ‘charged’ in conjunction with the exception of ‘some future proton’. Therefore, an explanation of any event that called upon ‘charged\*’ would really be supervening on more fundamental explanations drawn from the properties of which ‘charged\*’ is derived. Bird argues that if ‘charged’ and ‘charged\*’ were on an ontological par we could not explain why one has explanatory value and the other does not. We should therefore accept that there is a genuine ontological difference between natural properties with fundamental dispositions and non-natural properties, with no genuine dispositions of their own.

Bird’s purpose in this is to argue for the ontological robustness of natural/fundamental properties. The ontological robustness of natural properties lies in their explanatory power, and their explanatory power is founded upon their causal powers. As the explanatory power is dependent upon the causal power, the causal powers cannot change from world to world, and can be subject to metaphysically necessary laws.

I am not entirely convinced that the metaphysically necessary laws should not be extended to cover derived laws. It seems to me that if the causal powers of non-natural properties are derived from/supervene upon more fundamental causal powers they are not less causally or explanatorily efficacious. It is true that they will be dependent for their existence upon natural properties, but it will still be true that, due to their identification via causal powers, they should hold from world to world. I am inclined to think that *any* property instantiated at some

world, if it is to be explanatorily useful, will need to have consistent causal powers across worlds if it is to be explanatorily effective.

However, I need not wrangle over this point. For my purposes it will be sufficient if merely *some* properties have this sort of ontological robustness.

## 5.7 Dispositional Essentialism and Realism about Universals

Bird's argument in favour of dispositional essentialism has several stages. He starts by arguing for realism about universals, which in this case is the thesis that *universals are ontologically distinct entities over and above the properties of individual particulars*. This is opposed to the nominalist thesis that resemblance alone suffices, without the supposition of anything over and above particulars. The postulation of universals (or tropes) is not needed as the nominalist asserts that some resemblances may be explained by reference to other resemblances, and others are brute and fundamental and require no explanation. In this section I will examine the debate between Bird and Rodriguez-Pereryra, who has been the most prominent proponent of nominalism in recent times.

Bird's argument against resemblance nominalism is that, firstly, brute resemblance is unsatisfactory, and secondly, resemblance itself, on this account, looks like a universal itself. The first point focuses on the meaning of the term 'resemblance'. Bird resists the idea that brute resemblance can provide a grounding for explanations of resemblance that may avoid regress. He writes:

it is resemblance that needs explaining. When there is resemblance it is natural to want to explain it, and to explain it in terms of something held in common. Explanations may have to end somewhere, but ending in brute resemblance seems unsatisfactory. (Bird 2007: 15)

I think we can understand this as saying that the challenge is not to explain individual resemblances by reference to each other, and prevent this from becoming a regress by postulating that some are brute. Rather, it is resemblance itself that requires explaining—what it is for one thing to resemble another. Typically, this may be done by saying that for one thing to resemble another is for the two things to have something in common, but then one may ask: what is it to have something in common (if, that is, they do not share in a universal—an entity

over and above particular properties of an object)? It is easy to see how this may generate a regress of explanation (or a circular explanation—what it is for two objects to resemble one another is for them to have something in common, and what it is for them to have something in common is for them to resemble each other), and this regress is not one that may be solved by supposing that some resemblances are brute. This is because the question is not why *this* thing resembles another, but rather *what it is* for one thing to resemble another.

Rodriguez-Pereyra has responded to the criticism that ending in brute resemblance is unsatisfactory, but his response misinterprets the nature of the criticism. He takes it as being simply an intuition that “*a* and *b* resemble each other” cannot be true merely in virtue of the existence of *a* and *b*’ (Rodriguez-Pereyra 2003: 229), but I do not think this gives sufficient credit to the point Bird is making. Brute resemblance is not unsatisfactory for realists simply because they ‘intuit’ that it is wrong. Rather, it is unsatisfactory in that it is explanatorily insufficient. Granted, the assertion that some resemblances are ‘brute’ is meant to justify acceptance that no explanation is required. However, where intuitions conflict, as where brute resemblance seems plausible to some but not to others, other factors are considered for or against the positions. *Being explanatorily effective* is a point in favour of universals. Moreover, Rodriguez-Pereyra’s response to this criticism is precisely to reject an intuitions analysis in favour of a pros and cons comparison. He writes that he is not required to:

argue directly against the opposition, but to develop Resemblance Nominalism to the point that all objections against it were properly met and answered (2002: 6). Once Resemblance Nominalism had been developed to this point, I devoted the whole of chapter 12 to explaining why Resemblance Nominalism was superior to theories postulating universals and tropes. Basically Resemblance Nominalism’s superiority lies in that, unlike theories postulating universals or tropes, it does not postulate ad hoc entities, i.e. entities the only or main reason to believe in which is that they solve a certain theoretical problem, namely the so-called Problem of Universals. (Rodriguez-Pereyra 2003: 231–2)<sup>59</sup>

---

<sup>59</sup> This paper is in response to Bird’s 2003 paper, which employs the same arguments as Bird’s 2007. Bird’s original paper was in response to Rodriguez-Pereyra’s 2002 book, *Resemblance Nominalism*, referenced within this quotation—the book is his 2002a in my bibliography.

The advantage Rodriguez-Pereyra sees in resemblance nominalism is that it postulates no ‘ad hoc’ entities, but if the alternative explanation is simply that *there is no explanation*, this seems remarkably more ad hoc than to postulate entities that *would* provide an explanation.

Bird’s second argument is drawn from Russell’s regress of resemblance argument (see Russell 2008: 65). This argument runs that if one takes three white objects, *a*, *b*, and *c*, the nominalist will say that what they share is not the instantiation of the universal, whiteness, but a resemblance. *a* resembles *b*, and *b* resembles *c*. Russell then notes that if it is the same *resemblance* that is shared between *a* and *b* as that between *b* and *c*, then that shared resemblance is a universal.

This argument has been disputed, in particular, by Rodriguez-Pereyra (2001: 397). Rodriguez-Pereyra rightly points out that it is not obvious that what follows from Russell’s argument (as given at 2008: 65) is a regress, vicious or otherwise. Instead, he identifies two points that can be drawn from Russell: 1. if Russell has shown that resemblance is a universal, then it seems that resemblance has not helped us to avoid universals; and 2. it seems pointless to deny other universals, like whiteness, if we have admitted to resemblance as a universal. The regress comes in if we insist that the resemblance between *a* and *b*, and the resemblance between *b* and *c*, are each, themselves, particulars. In this case, there would then need to be a resemblance between these two resemblances, and a resemblance between the higher-level resemblance and the original resemblances, and so on, generating the regress. (Russell explores this in 1956: pp. 346–7)

Rodriguez-Pereyra responds to this by pointing out that the resemblance nominalist is under no obligation to reify resemblances. They need ‘only suppose that particulars resemble each other, not that there are any resemblances’ (Rodriguez-Pereyra 2001: p. 401). If there are no resemblances, there can be no regress of resemblances, merely finite particulars that resemble each other.

This raises the problem of the ‘Many over One’, or, as Rodriguez-Pereyra puts it, “‘how can the same thing be of different types?’” (Rodriguez-Pereyra 2002: 39), that is: what does it mean to say that two particulars resemble each other? In order to resist the appeal to universals, Rodriguez-Pereyra proposes that what accounts for this is not *resemblances*, but *resemblance facts*, i.e. what makes ‘sentences like “*a* and *b* resemble each other” true’ (Rodriguez-Pereyra 2001: 401). The question then becomes: what makes it true that *a* and *b* resemble each other? And his answer is simply ‘*a* and *b* together’:

If *a* and *b* resemble each other then they do so because of their being the entities they are, and so *a* and *b* are the sole truthmakers of ‘*a* and *b* resemble each other’. There is then no need to postulate extra entities to account for facts of resemblance: the resembling entities suffice to account for them. (Rodriguez-Pereyra 2001: 403)

This is a deeply unsatisfying answer for the realist—it seems to explain nothing at all. *How* can ‘*a* and *b* together’ be the truthmakers of ‘*a* and *b* resemble each other’? How can their mere existence entail resemblance?

To complete his answer and provide a full explanation, Rodriguez-Pereyra (see 2001: 403–5) must insist on realism about possible worlds, and Lewis’s counterpart theory. On this view the mere existence of *a* and *b*, being the particulars they are, entails ‘*a* and *b* resemble each other’ for the world in which *a* and *b* exist, because on this view of possible worlds, no individual exists in more than one world—they merely have *counterparts* which differ from them to varying degrees. Therefore, if *a* and *b* are resembling particulars, this is entailed by them being the particulars they are in the world in which they exist. The truth of the one proposition makes for the truth of the other.

This is an oddly unexplanatory sort of truthmaking. It reduces entailment, as Bird (2001: 17) points out, to material implication. No true relationship is revealed as holding between ‘*a* and *b* exist’ and ‘*a* and *b* resemble each other’, as is illustrated by the fact that *for each fact in any possible world*, this theory will ensure that all the other true facts at that world are entailed. To use Bird’s example: “‘David Beckham plays football’ entails “‘Snow is white’” (*ibid.*) in the same way that can ‘*a* and *b* exist’ entails ‘*a* and *b* resemble each other’, and there is clearly no explanatory relation between David Beckham playing football and snow being white.

For these reasons, Bird argues that at best this renders the concept of truthmaking rather trivial, giving the necessary, but not sufficient, principle (T):

(T) If  $E_1, \dots, E_n$ , are joint truthmakers of ‘S’ then ‘ $E_1$  exists & ... &  $E_n$ , exists’ entails ‘S’. (Bird 2003: 224)<sup>60</sup>

---

<sup>60</sup> Rodriguez-Pereyra gives (T) as ‘(T\*\*)', and seems rather put out that Bird made this change in nomenclature. I hope he will forgive me for following Bird, I must confess that I find Bird’s nomenclature easier to digest. See Rodriguez-Pereyra (2003: 230, n. 1) for more on the fraught issue of changing nomenclature.

(T) is not enough to support that  $a$  and  $b$  are truthmakers of ‘ $a$  resembles  $b$ ’. It would only show that if they exist they are truthmakers for ‘ $a$  resembles  $b$ ’ in the sense that everything else that exists at the world where ‘ $a$  resembles  $b$ ’ is true must exist in order for it to be true. Counterpart theory is doing all the work, here. The more significant claim for the truthmaking theory is what Bird calls ‘(Y)’<sup>61</sup>:

(Y) Entity  $E$  is a truthmaker of ‘ $S$ ’ if and only if  $E$  is an entity in virtue of which ‘ $S$ ’ is true. (*ibid.*)

This, Bird criticises by saying that ‘in virtue of’ is a ‘thick’ concept, one which it is hard to understand in a way that is more than the trivial entailment claim of (T) without cashing out in terms of ‘a property that is shared by  $a$  and  $b$ , or their resembling one another’ (Bird 2003: 225). If this is the case, it looks like we have come back to exactly what Russell was criticising in the first place. Either  $E$  is a truthmaker of  $S$  because it instantiates some universal, or because it involves ‘resemblance’ in some sense that is circular, or leads to a regress, or that renders ‘resemblance’ a universal in itself.

Rodriguez-Pereyra does offer a reply to this argument (see Rodriguez-Pereyra 2003), of which Bird’s only comment is that he finds it ‘perplexing’ (Bird 2007: 18, n. 13). The challenge to fill out the thick concept of ‘in virtue of’ is not met. Rodriguez-Pereyra maintains that his defence of the entailment relation between ‘ $a$  and  $b$  exist’ and ‘ $a$  and  $b$  resemble each other’ was not intended as an argument for  $a$  and  $b$  being as truthmakers of ‘ $a$  and  $b$  resemble each other’. He thinks it is enough to show that resemblance nominalism is internally consistent and superior to realism in terms of answering the problem of universals. Given this, he asks what candidates there could be for the truthmakers of ‘ $a$  and  $b$  resemble each other’ except for  $a$  and  $b$ . ‘[T]he ontology of Resemblance Nominalism consists of particulars and classes (sets)’, but classes are merely sets of particulars, and the only particulars in virtue of which  $a$  and  $b$  might resemble each other are  $a$  and  $b$ .

As an answer, this completely fails to address the issue of what ‘in virtue of’ means beyond the triviality of (T). No wonder Bird is perplexed. Overall, one rather feels that Bird and Rodriguez-Pereyra perplex each other. I think what this comes down to is the pros and

---

<sup>61</sup> What Rodriguez-Pereyra calls ‘(T)’ (see his 2003: 230, n. 1).

cons analysis Rodriguez-Pereyra describes in his 2003 (231–2). For Bird, the explanatory value of the account is prized; for Rodriguez-Pereyra, avoiding what he regards of the ontological extravagance of invoking universals in order to offer such an explanation. Rodriguez-Pereyra thinks no deeper explanation is required; Bird does not think universals are an ontological extravagance if they provide a deeper explanation. As such, they are at an impasse.

I, like Bird, favour an account that allows for the deeper explanation, but if one is not convinced by that, one might consider that *if* resemblance nominalism is required in order to have an alternative to dispositional essentialism, and dispositional essentialism allows us to construct a working strong transcendental argument, then it must count in the ‘cons’ column that resemblance nominalism opens the door to scepticism. By a similar token, whilst nominalism might offer a way out for the sceptic who wishes to avoid dispositional essentialism and a possible strong transcendental argument, the sceptic is thereby forced to adopt and defend a specific philosophical position, which is a step away from her place as one who merely prompts us to doubt. As a figure who tests the success of our theories, she would count against the resemblance nominalist. As the intermittent doubt in one’s own mind, she would thereby only gain admittance if one accepted resemblance nominalism, and this serves to greatly narrow her threat.

## 5.8 Dispositions

Having rejected resemblance nominalism, Bird then defends the realist analysis of properties in terms of dispositions. This analysis is complicated by the fact that some property names resist clear analysis in these terms. In this section I will consider the difference between ‘covert’ property names and ‘overt’ dispositional property descriptions. I will argue that the existence of covert property names does not undermine the dispositions analysis overall; that their covert nature is a reflection of the imprecision of natural language and not any deep truth about the nature of properties. I will then give the traditional Conditional Analysis of properties from which the account of natural laws will be developed.

‘Covert’ property names are names such as ‘fragility’, ‘brittleness’, and ‘elasticity’. These covert property names align to covert dispositional predicates, such as: ‘fragile’, ‘brittle’, ‘elastic’. These are ‘covert’ in the sense that they do not explicitly give their stimuli



and manifestations. ‘Overt’ dispositional property descriptions overtly take the form ‘the disposition to M when S’ (where M is the manifestation and S the stimulus). Overt disposition property descriptions then align to overt dispositional predicates of the form ‘is disposed to M when S’. Thus, fragility cannot be simplistically cashed out in a description because different stimuli and manifestations may describe fragility in relation to context. The glass is fragile because it is disposed to break when dropped, but the ancient document is fragile because it may fragment on touch (see Bird 2007: 19–20).

One might be concerned that a lack of neat correspondence between covert disposition names and overt disposition property descriptions (and their dispositional predicates) creates a problem for dispositional analysis of properties. This is too quick a move. Fragility *can* be cashed out relative to context—to what it is that makes the glass fragile or the ancient document fragile. Language is not neat. Some terms may be applied to a number of similar, but distinct things. Some may align more definitively—especially if the term was devised to describe a specific phenomenon, as opposed to evolving through centuries of use. Bird (2007: 20) gives as an example: ‘inertial mass may perhaps be characterized as the disposition to respond to the stimulus of a force by accelerating in proportion to that mass’. What is important, for my purposes, is that some dispositional properties can be characterised by overt dispositional property descriptions and overt dispositional predicates such that they can be expressed as:

$x$  possesses the disposition to M when S iff  $x$  is disposed to M when S. (Bird 2007: 19)

Under the Conditional Analysis (CA) of dispositions, this can be rendered as:

$x$  is disposed to manifest M in response to stimulus S iff were  $x$  to undergo S  $x$  would yield manifestation M. (Bird 2007: 24)

I will now turn to the more serious challenge to CA, which are the proposed counter-examples of ‘finks’, ‘antidotes’, and mimics.

## 5.9 Finks, Antidotes, and Mimics

It has been pointed out that where dispositions take time to manifest following a stimulus, the disposition may have time to go out of existence. Where this happens, the disposition is called ‘finkish’<sup>62</sup>:

an object’s disposition is finkish when the object loses the disposition after the occurrence of the stimulus but before the manifestation can occur and in such a way that consequently that manifestation does not occur. (Bird 2007: 25)

Bird considers an example which is a variation of Charlie Martin’s ‘*electro-fink*’<sup>63</sup> (Martin 1994: 2–3). Suppose a wire has the property of being live, defined by the dispositional property description ‘disposed to conduct current when in contact with a conductor’. Suppose that the wire is attached to a machine called an ‘electro-fink’, and this machine has a way of detecting when the wire is in contact with a conductor and changing its disposition to be ‘dead’ (i.e. disposed to not conduct current when in contact with a conductor) by cutting its connection to a generator. The wire is live. It is disposed to conduct current when in contact with a conductor. However, when it receives that stimulus (being in contact with a conductor) it does not evince the manifestation CA suggests is a necessary part of having that disposition. Surely, the objection goes, CA is false.

An antidote to a disposition is something that breaks the causal chain that leads from the stimulus to the response. The obvious example of this would be an antidote to a poison:

Let us understand ‘fatally poisonous’ to mean ‘disposed to kill if ingested’. It is possible to ingest a dose of a fatal poison yet survive if one has also taken an antidote. (Bird 2007: 27)

---

<sup>62</sup> The origin of the term ‘fink’ is attributed to David Lewis, although the concept it describes seems to have been first raised by Mark Johnston, in ‘Objectivity Refigured: Pragmatism Without Verificationism’, Appendix 2. See Blackburn (1993: 278, n. 11) concerning origins of the term.

<sup>63</sup> In Martin’s original example the concern is to undermine the conditional in the other direction (i.e. were  $x$  to undergo  $S$ ,  $x$  would yield manifestation  $M$  iff  $x$  is disposed to manifest  $M$  in response to stimulus  $S$ ). In that example, the electro-fink ensures that the wire is not live unless touched by a conductor, and so it would seem that the conduction of a current when touched by a conductor was not a disposition of the wire, even though the wire underwent the requisite stimulus and manifested the requisite manifestation.

If I ingest a poison, but then take an antidote that changes my body in some way such that it does not react fatally, then the poison's disposition remains intact, but the manifestation (killing) does not follow necessarily upon the ingesting.

Mimics attack the biconditional in the other direction. Suppose an object lacks the disposition to M when S, and yet it *does* M when S. How might this be? Bird gives the example of a sturdy cooking pot—not fragile in the least—being attached to an explosive device with a sensitive detonator. The cooking pot will be broken if knocked even lightly, just as if it were fragile, and yet it is not fragile.<sup>64</sup>

### 5.10 Repairing the Conditional Analysis: *Ceteris Paribus* Laws

Bird considers a number of possible repairs to CA to avoid these difficulties (see Bird 2007: 37–42), drawing on the work of Lewis, Mumford, and Choi, but finds these proposals wanting. There is no adequate space to discuss these, here, so I will assume for the sake of argument that he is correct, and instead move on to consider Bird's response. I think an even more powerful response is at his disposal, but as it builds on Bird's work, it helps to outline this first.

Bird reasons that the respects in which CA is false can be changed from a negative, to a positive outcome. For it is precisely the areas in which CA fails that allow it to account for *ceteris paribus* laws. He therefore proposes the following modification to CA:

(CA→\*) if  $x$  has the disposition  $D(S,M)$  then, if  $x$  were subjected to  $S$  and finks and antidotes to  $D(S,M)$  are absent,  $x$  would manifest  $M$ . (Bird 2007: 60)

This might seem a rather arbitrary and convenient addition, exactly the kind of thing the sceptic can easily doubt. To support it, one needs to examine what exactly *ceteris paribus* laws are, and why it would be a good thing if one's account of laws were to accommodate them.

*Ceteris paribus* laws are generally taken to arise for the special sciences, in contrast to the laws of physics. A law of physicals, such as Newton's second law 'F=ma', admits of no exceptions. Whereas a law of biology, such as Mendell's law of segregations ('In a parent, the

---

<sup>64</sup> The example is from Bird (2007: 29), but the objection is attributed to Johnston (see 1992: 232), and this style of counter-example to A D Smith (see 1977: 440).

alleles for each character separate in the production of gametes, so that only one is transmitted to each individual in the next generation' (Rosenberg and McShea 2008: 36)), may be subject to interference, and therefore exceptions. Because of this, some philosophers of science argue that such laws should be treated as though subject to a *ceteris paribus* clause, as, other things being equal, they still hold.

This is disputed.

On one understanding, a law is a *ceteris paribus* law if it does not entail strict regularities (see Earman, Roberts, and Smith 2002: esp. 288). It is suggested that to be a true *ceteris paribus* law, 1) the law should fail to be specific about what regularities would have to obtain for the law to hold, or 2) the law should be vague and impossible to be made explicit due to the conditions upon its regularity composing 'an indefinitely large set' (Earman, Roberts, and Smith 2002: 284). This looks like a significant problem—if *ceteris paribus* laws do not entail strict regularities it sounds as though there is room for the sceptic to cast doubt on their metaphysical necessity.

However, Cartwright (2002: 425) disputes this definition. She insists that *ceteris paribus* laws are laws that: 'can be stated in precise and closed form or they entail strict or statistical regularities in the course of events or both'. For our purposes, *ceteris paribus* laws fit Cartwright's definition as they can be stated precisely, in that the only exception to their holding would be finks and antidotes. I do not think it should worry us that only one definition of *ceteris paribus* laws fits our requirements. The point of a definition of a type of law is to attempt to codify a lawlike relation. Both definitions assert the existence of something in reality that requires explanation. It is of advantage to Cartwright if  $(CA \rightarrow *)$  explains the presence of *ceteris paribus* laws as she describes them, just as it is an advantage to Bird if the modification he proposes would account for such laws.

## **5.11 Fundamental Dispositions cannot be Finkish**

In addition to finding a place for explaining *ceteris paribus* laws in terms of finkish dispositions, antidotes, and mimics, Bird argues that *fundamental* dispositions cannot be finkish in the standard sense. He writes:

The standard cases of finkish dispositions are those where the causal basis for a

disposition is removed before it can complete its causal work that would otherwise lead to the manifestation of the disposition. In the case of a fundamental disposition there is no distinct basis. Hence there can be no cases of finks of this sort. (Bird 2007: 60)

This is confusing. What is meant by ‘distinct causal basis’? What is being distinguished from what? In this section I will clarify Bird’s thought, and in doing so, lead us to an even stronger argument in favour of the impossibility of finkish<sup>65</sup> dispositions at the fundamental level.

The confusion arises from a little carelessness of phrasing. The manifestation of a finkish disposition *is* for it to fail to manifest something or other. What it is manifesting is the failure to manifest of a *different* disposition, one which contributes to the make-up of the finkish disposition. Take the example of the electro fink. The wire is disposed to be live up until the point at which it is put in touch with a conductor, at which point the electro fink causes it to lose this disposition. This is a disposition that never manifests—the live wire is never touched by a conductor and therefore never conducts a current. It is not the live wire whose disposition is finkish. The finkish disposition is a composite of the properties of the wire and the electro fink; it is the disposition for the wire to lose the disposition to conduct electricity (to be live) at the moment it is touched by a conductor. Obviously a fundamental disposition can never be finkish in this sense. Its simplicity is such that it can never be explained by appeal to a ‘deeper level’ of dispositions, just as the finkish disposition is explained away by appeal to the dispositions of the live wire and the electro fink combined.

Note that this point is made without clarifying whether the ‘distinct causal basis’ is intended to refer to the distinct dispositions, the distinct objects, or both. This becomes relevant when Bird goes on to discuss whether:

objects can be made to acquire or lose dispositions not only by causing them to gain or lose their distinct causal bases. *It must be possible for an entity to gain a fundamental potency directly.* For example, electrons can be spin-up or spin-down and they can be caused to change from one to the other. *If these are fundamental properties* then the interactions that result in a change of spin direction will act by bringing about this change directly, rather than indirectly by causing a change at some even deeper level of property. (*ibid.*, emphasis mine)

---

<sup>65</sup> And antedotish and mimichish.

Here an object is theorised to be able to gain a fundamental potency without losing or gaining a ‘distinct causal basis’. Bird (2007: 60–2) worries that this looks finkish, concluding that whether or not there can be finkish properties depends on one’s understanding of time. However, I think he’s missed something that, if properly understood, gives us a more powerful reason for supposing that finkish dispositions cannot occur on the fundamental level at all.

In the example, the electrons are still caused to change from spinning up or down by an external object, and this is the gain/loss of a fundamental potency. The changes happen ‘directly’ in the sense that they are not brought about by causing a change at a deeper level of the *property*. But then, this does not look finkish at all. The object simply gains or loses a property. I think Bird has become confused in slipping between talk of fundamental properties (simple dispositions with no deeper levels) and fundamental potencies (dispositional essences that make an object the kind of object it is). The electrons are changed from being spin up to spin down electrons. This change happens by a direct causal connection from the stimulus to the spin of the electron. The object is still the same object (possessed of other fundamental properties that make it an electron, like being negatively charged), but it has changed potencies (the causal dispositions determined by it spinning one way, rather than the other). It is still an electron, but now it is an electron of the kind with spin down causes and effects. The illusion of finkishness must come from supposing that the electron staying the same object is somehow contradictory with its loss of potency, but this is not so. It is just an electron that has changed how it spins. This will have causal consequences for other things, but it is not finkish.

I therefore argue that Bird’s position is even more persuasive than he takes it to be: there cannot be finkish dispositions at the fundamental level. Finkish dispositions are dispositions that are composites of more complex dispositions—ones that change over time. The simpler dispositions upon which they are based are not finkish—they simply change to a different disposition before they can manifest.

Similarly for antidotes. If I have taken an antidote before consuming a poison, then I have changed the disposition of my body to react to those chemicals. It is no longer the kind of thing that can be poisoned by such chemicals, even though prior to consumption of the antidote it was. The antidote sensitive disposition is a combination of my (complex) body’s states over a period of time. But at a time when I have taken an antidote, I simply do not have the disposition to manifest ‘being poisoned’. This makes the disposition to poison humans complicated and difficult to express otherwise than via a *ceteris paribus* law, but that is only

150

due to complexity of the objects in question and their changing dispositions. At the fundamental level the root causes covered by the law will not be *ceteris paribus*. The disposition of arsenic to be an allosteric inhibitor of pyruvate dehydrogenase<sup>66</sup> does not change simply because one applies a DMSA monoester<sup>67</sup>, preventing death or reducing symptoms.

As regards mimics, it should be obvious that the kind of object that can be a mimic is a complex entity. It is the combination of the properties of the pot *with* the properties of the explosive device that make it manifest in a way that suggests fragility in response to stimulus. This is a distinguishing feature of covert dispositions: they are the result of combination of several properties, to which more properties could be combined to change outcomes. So adding something to a non-fragile pot that makes it behave in a fragile way can be explained in that it is not the pot by itself that is fragile, but the combination of the pot with something else (the explosive). The very act of adding something to a fundamental potency will mean that *the two properties combined* are not fundamental. And any mimic will involve some such addition. This explains why *ceteris paribus* laws are needed to account for mimics, and why mimics are impossible at the fundamental level.

Having established this, I will now turn to the matter of giving a precise statement of the laws of nature in terms of potencies.

## 5.12 The Metaphysical Necessity of Laws of Nature

Bird makes the following argument from his characterisation of dispositional essentialism to the necessity of the laws of nature.

He proposes that we can understand the individuation of dispositionally essentialist properties (potencies) by their powers. Even that we might identify them with their powers<sup>68</sup>. We may then characterise dispositionally essentialist potencies as follows:

---

<sup>66</sup> The property that makes arsenic normally poisonous to humans.

<sup>67</sup> An antidote to arsenic.

<sup>68</sup> He says that this is not required for his argument; I do not think it is required for mine either, but it does seem to me correct—proposing that potencies are something over and above the powers by which we individuate them seems ontologically extravagant.

(DE<sub>P</sub>)  $\Box(Px \rightarrow D(S,M) x)$ . (Bird 2007: 45)

Where P is a potency, D is a disposition, M is a manifestation, and S is a stimulus. He also suggests that we consider CA as a necessary equivalence:

(CA $\Box$ )  $\Box(D(S,M) x \leftrightarrow Sx \Box \rightarrow Mx)$ . (Bird 2007: 43)

This seems reasonable: (excepting finks, antidotes, and mimics) necessarily, if and only if  $x$  has the disposition such that it is disposed to manifest M when S, then if  $x$  has stimulus S, it necessarily follows that  $x$  will manifest M.

Combining these two principles generates:

(I)  $\Box(Px \rightarrow (Sx \Box \rightarrow Mx))$ . (Bird 2007: 46)

I.e. necessarily, if  $x$  has a potency P, then  $x$  is disposed to manifest M when stimulated by S; and given that the disposition for  $x$  to manifest M when stimulated by S is equivalent to it being necessary that if  $x$  is stimulated by S, then  $x$  will manifest M, it follows that it is necessarily true that if  $x$  is a potency, then, necessarily, if  $x$  is stimulated by S, then  $x$  will manifest M.

For any world,  $w$ , where some  $x$  has potency P and receives stimulus S:

(II)  $Px \ \& \ Sx$  (*ibid.*)

It will then follow from (I) and (II) that:

(III)  $Mx$  (*ibid.*)

So (I), (II), and (III) deliver:

(IV)  $(Px \ \& \ Sx) \rightarrow Mx$ . (*ibid.*)

Which, given that  $x$  is arbitrary, can then be generalised to:

(V)  $\forall x((Px \ \& \ Sx) \rightarrow Mx)$ . (*ibid.*)



Bird suggests two ways via which (V) can account for the laws of nature.

One view might take (V) to state a law of nature in itself. Laws would be regularities, as with the Lewisian view, but to be a regularity would not be sufficient for being a law. What *makes* one regularity a law, when another might be happenstance, would be whether the regularity is guaranteed by the essentially dispositional nature of the property or properties involved.

Alternatively, one might take laws to be a kind of relation between universals. A law might then be characterised as necessitation ‘N’ for relating P, S, and M, such that N(P&S,M). This would be similar to the Armstrongian position that laws are nomic relations described as ‘N(F,G)’, where:

N(F,G) iff  $\Box(\forall x(Fx \rightarrow Gx))$  (where ‘F’ and ‘G’ are names of universals). (Bird 2007: 47)<sup>69</sup>

However, on the dispositional essentialist account, ‘N’ is restricted to instances that arise necessarily from the essence of a property, such that it is a relation between a potency, stimulus, and manifestation as described in (V). Bird also stresses that the dispositional essentialist’s ‘N’, unlike Armstrong’s, need not be a universal itself; it may be supervenient on the dispositional nature of properties. Moreover, it differs from Armstrong’s in that it is metaphysically necessary, as opposed to Armstrong’s ‘nomic’ necessity, which has been critiqued earlier.

As (V) holds for any arbitrary world, it follows that it is necessary:

$(V \Box) \Box \forall x((Px \ \& \ Sx) \rightarrow Mx)$ . (Bird 2007: 48)

---

<sup>69</sup> Bird introduces an ‘r’ to accompany the ‘x’ for ‘ $\forall rx$ ’ here which I have omitted. He does so without defining r, and continues from this point onwards to use ‘r’ in variations of (V). It’s not quite clear what the role is. It might be a way of specifying that x is a relation, although as ‘N(F,G)’ and ‘N(P&S,M)’ are supposed to be the relations, this seems unlikely. Additionally, as (V) is supposed to be ambivalent between laws-as-regularities and laws-as-relations-between-universals, this would seem peculiar. I therefore think it is clearer to remove the ‘r’, both here, and in the variations of ‘V’ below.

If we then built in the *ceteris paribus* requirement, with ‘finks, antidotes, and mimics to D are absent’ as the *ceteris paribus* clause,  $V_{\square}$  would become:

$$(V^*) \forall x (\text{finks and antidotes [and mimics] to D are absent} \rightarrow ((Dx \ \& \ Sx) \rightarrow Mx)). \text{ (Bird 2007: 60)}$$

Given what we have previously said about finks, antidotes, and mimics, the absence of finks, antidotes, and mimics is a given, and we could therefore give the following as a fundamental law of nature:

$$(V^{**}) \forall x (P^*x \rightarrow ((Dx \ \& \ Sx) \rightarrow Mx))$$

Where ‘ $P^*$ ’ is a fundamental property.

It is now time to turn our attention back to the strong transcendental response to the sceptic. Will we be able to incorporate a metaphysically necessary law of the kind Bird has argued for into our argument?

### 5.13 Finkish Dispositions, Psychophysical Laws, and the Sceptic

Recall that finkish dispositions were dispositions such that the disposition is lost after the stimulus but before the manifestation. What we need to examine is how a psychophysical law would operate such that it would follow from dispositional essentialism in the manner of  $(V^*)$  or  $(V^{**})$ .

If we assume that D is a mental disposition and M is a physical manifestation, then we seem to be headed for trouble. Something could interfere with the causal process such that mental dispositions disposed to have physical consequences might have no connection to a physical world at all. However, this is not the set-up. A transcendental argument works on a premise such that something extra-mental is a necessary condition of something mental, so it is not concerned with necessary physical consequences of mental dispositions.

Let us try it the other way around, then. Let D be a physical disposition to give rise to a mental consequence, M, in response to stimulus S. Again, this does not give us what we need. The necessary connection in that set-up is one way, so knowing M will tell us nothing about the causal story that led up to M. That  $Mx$  necessarily follows from  $Dx + Sx$  does not tell us

that M can have no other cause, and we do not get the result we were looking for from metaphysically necessary laws of nature, as this cannot be known to be a law from knowledge of mental propositions alone.

The problem is that our supposition to get around there being a ‘gap of necessity’ that needed to be bridged was that mental properties necessarily supervened upon extra-mental properties (hence, no gap); but the metaphysical necessity drawn for the laws of nature by dispositional essentialism looks to be a causal relation, not a relation of supervenience. This is how the time factor creates room for the fink problem.

The mistake, here, is in misunderstanding the nature of a psychophysical law. The supervenience relationship is not one of a mental disposition that has a physical consequence; nor is it one of a physical disposition that has a mental consequence. The mental and physical dispositions that supervene upon one another are not causally linked to each other. Rather, they have the same causes and effects, although the one may rely upon the other to offer a more satisfactory explanation of why that is the case. Supervenience physicalism is motivated by explanatory efficacy.

How can this fit into Bird’s analysis of dispositional essentialism? I propose the following analysis for psychophysical laws (PP):

$$(PP) \forall x ((\text{finks and antidotes to B are absent} \rightarrow ((Bx \ \& \ Sx) \rightarrow Mx)) \leftrightarrow (\text{finks and antidotes to A are absent} \rightarrow ((Ax \ \& \ Sx) \rightarrow Mx)))$$

Where A is a physical disposition and B is a mental disposition. This expresses the supervenience of any law that applies to a mental disposition, A, on any law that applies to a physical disposition, B. Once it is clarified that the link between A and B is not causal, it becomes obvious that finks and antidotes cannot disrupt the necessary connection between A and B. This is because any case in which the connection between  $(Bx \ \& \ Sx)$  and  $Mx$  is disrupted will also be a case in which the connection between  $(Ax \ \& \ Sx)$  and  $Mx$  is disrupted. In other words, should disposition B be at all finkish, it will only be finkish in the exact same ways as disposition A. Similarly, if something is an antidote to disposition B, it will also be an antidote to disposition A. For example, if I take ibuprofen I do it because it is an antidote to the mental state ‘pain’, but it is only an antidote to the mental state ‘pain’ because it is an

antidote to the conversion of arachidonic acid to prostaglandin H<sub>2</sub> by cyclooxygenase<sup>70</sup>. This conversion would be the normal disposition in response to various physical stimuli associated with physical injury or damage, but the antidote, in preventing this, also serves to prevent the mental state of 'pain' that supervenes on such conversion; in the absence of ibuprofen, the disposition of bodies such as ours to respond by producing the chemical reactions that are the physical manifestation of pain would necessarily be accompanied by the mental manifestation of pain.

Now, if physicalism is correct, then there will be some mental states that can only be adequately explained by their supervenience upon physical states. It will not be possible to fully articulate their stimulus and manifestation conditions solely in terms of relations between mental properties. *If* however, a theory can be provided that can explain the existence of the mental state in terms of something extra mental, then it will take the form of (PP), for such an explanation would involve articulation of stimulus and manifestation conditions in terms of something other than the mental.

*Given* that the metaphysical necessity of the laws of nature follows from dispositional essentialism, then there can only be one set of stimulus and manifestation conditions that governs the mental property A. So if an explanation (an account of the stimulus and manifestation conditions) for A can be provided in terms of extra-mental stimulus and manifestation conditions only *at some world*, then, because that explanation will involve a law of nature, those will be the only possible stimulus and manifestation conditions *at all worlds*.

In chapter three I established that there are at least some mental propositions that the sceptic must accept as true. Her doubt rests on the thought that these could be equally well explained by some sceptical scenario in which the world is radically different to how it seems to me and there is no extra-mental world. Her doubt is not one that supposes that the world might work in some completely random manner: there are regularities in our experience and this demands explanation. She doubts the existence of the extra-mental world because she worries that truths about the mental content she does not doubt could be equally well explained in purely mental terms (such as that I am a deluded solipsistic purely mental consciousness). However, attempts to explain the regularities in our experience based solely on propositions relating to other mental states fail. If physicalism is correct then a complete

---

<sup>70</sup> One way in which painkillers act upon the body to prevent pain.

science would be able to explain our mental states in terms of physical states, including an explanation of mental states (such as certain pains) that would otherwise come seemingly out of nowhere. If dispositional essentialism is correct, then, the explanation offered by a completed science in a physicalist world will take the form of a psychophysical law structured along the lines of PP and, due to the metaphysical necessity of the laws of nature, it will follow that this that the relationships between properties so described could not have been otherwise.

The sceptic might then further worry that it is not humanly possible that a complete science can be known. Yet, this is not sufficient to undermine the possibility of providing the explanation we need. All a strong transcendental theorist needs is for *some* part of science to be revealed that is not current known. Namely: that part of science that gives the stimulus and manifestation conditions for mental state prized by the sceptic—a much smaller hurdle.

The sceptic might further object that articulating the totality of any one thing's stimulus and manifestation conditions may be beyond human ability. As discussed earlier, covert property names resist overt dispositional property descriptions. What if all mental states the sceptic takes herself to know she possesses have only properties that can be given covert property names, and not overt dispositional property descriptions? Two points can be made against thinking that this will be the case.

Firstly, as indicated earlier, dispositional property descriptions for covert property names can be cashed out once context is clarified. 'Fragility' is difficult to pin down when we do not know if we are talking about a vase or an ancient document: ancient documents are fragile because they easily tear or fragment when touched; vases are fragile because they easily break upon being dropped or hit. Fragility as a whole is difficult to define; the fragility of a specific object is not.

Secondly, the difficulty of providing overt dispositional property descriptions vanishes at the foundational level. Part of the reason the difficulty of providing such descriptions arises for covert property names is because they arise through complex evolution of language. By contrast, descriptions of foundational properties are simple, and thus precise and well-defined. If it can be shown that a mental property supervenes upon a foundational physical property or properties, then it will be explicable in terms of overt dispositional property descriptions.

To return to our painkiller example: if I sustain a muscular injury I will usually experience pain, but the application of ibuprofen will usually reduce or entirely prevent the

pain. This conditional is clearly open to antidotes. However, if supervenience physicalism is correct, wherever there is pain, there will be some kind of physical event upon which the experience of pain supervenes, and the prevention of the mental event (say, by the application of painkillers) will be explained by the absence of that physical event. At the broad, complex level, the law does not hold, but at the fundamental level it does. As the pain comes out of nowhere on the mental level, we would then have reason to think that the physical world exists, supplying the cause. The pain itself can be certainly known to be experienced, giving us some reason to think that it supervenes on something that is a fundamental potency. Nothing can come between my thinking that I experience pain and my actually experiencing it. This in itself is not open to finks or antidotes, although clearly there are broad characterisations of when I experience pain that are open to antidotes. Spraining a muscle (or thinking that I do, or having an experience as of spraining a muscle, if all muscles are imagined, purely mental constructs) is broadly a stimulus for pain, but taking ibuprofen can prevent the experience of that pain. However, if we can identify some foundational property (some property of C-fibres when fired, perhaps<sup>71</sup>) that, whenever appropriately stimulated, pain manifests, it will be explained how we could not possibly be mistaken about pain, because *that* property is not possibly finkish or open to antidotes.

#### **5.14 Necessitarianism about Laws**

To return to the argument for the necessity of natural laws: as  $V \square$  follows from dispositional essentialism, partial necessitarianism about laws follows:

(PNL) At least some of the laws of nature are metaphysically necessary. (Bird 2007: 50)

On this view at least some laws would be metaphysically necessary on the basis of dispositional essentialism. Other laws might be accounted for in a more Lewisian or Armstrongian way. We would thus have a mixed view of laws, as well as a mixed view of

---

<sup>71</sup> This is the oft cited event upon which pain is thought to supervene. As I am not claiming to make a successful transcendental argument here, nothing turns on the accuracy of this description.

properties. This is opposed to full necessitarianism, on which theory all laws would be metaphysically necessary.

The transcendental theorist is not forced to argue for full necessitarianism. It will suffice for my purposes if at least *some* potencies exist and therefore some metaphysically necessary laws of nature exist. What I require, after all, is that metaphysically necessary laws of nature entail a metaphysically necessary connection between certain mental propositions that the sceptic accepts and the truth of some extra-mental propositions. So long as it can be shown either that the truth of some mental proposition depends on a mental manifestation which only manifests if some extra-mental potency exists, the metaphysical architecture will support a strong transcendental argument. However, it would be beneficial to the strong transcendental theorist if full-necessitarianism were true, in as much as simply showing that a law of nature so related some mental property and some physical one, without having to show in addition that the law of nature in question happened to be a metaphysically necessary one.<sup>72</sup> In any case, although Bird does not offer a smack down argument in favour of full necessitarianism (i.e. not one the sceptic would obviously have to agree to), most scholars seem to prefer a unified approach to accounting for the laws of nature, and if partial necessitarianism is true, and a unified approach should prove to be necessary, full necessitarianism will be true by default.

Bird also identifies two views one might take on the nature of the metaphysical necessity in question. If the law supervenes upon the presence of the universals it concerns, then it would seem that some laws do not hold at all worlds because the potencies they operate over do not. This is what Bird (2007: 49) calls a ‘conservative’ view of necessitarianism. The thought is that on this view the necessary unity would reside in ‘nomological consistency’ of *facts*: no fact at any world would provide a counter-example to any law, existent at another, even though some laws would not hold for some worlds simply because the necessary potencies were not present to operate over. This is contrasted with a ‘radical’ or ‘strong’

---

<sup>72</sup> To an extent, the ‘labour-saving’ point is moot. If all laws of nature are metaphysically necessary because they supervene on the operations of potencies, then identifying a law of nature will still be done by demonstrating its operation over potencies. This is the same operation one would employ in identifying those laws of nature which are metaphysically necessary were partial necessitarianism true. However, the simplicity of appeal to a unified approach would be a point in favour of necessitarianism as a whole, so it seems that if a strong necessitarian account can be given, it would be beneficial to do so.

necessitarianism for which the laws of nature are necessary in the sense that they hold over all possible worlds, and thus the universals they govern exist at all possible worlds. Although Bird clarifies that: ‘the proposition can be true of a world where the fact does not exist’ (Bird 2007: 50); just as with Kripke’s notion of the necessity of identity, where ‘Geroge Orwell = Eric Blair’ will be true at all worlds, regardless of whether the fact of Orwell’s existence is true at that world (see Bird 2007: 49–50)<sup>73</sup>. On this view, worlds may be nominologically consistent as regards *laws*, but not *facts*.

Bird sees strong necessitarianism as problematic for those, like Armstrong, who accept the Aristotelian notion that to exist at a world, a universal must be instantiated at that world. The thought being that the law is true by virtue of the universals it involves. If the universals only exist when instantiated, then no law can be applied, as there will be no universals to apply it to, or to express a relation between. Bird sees a reliance upon particulars by universals as a ‘failure of nerve as regards realism about universals’ (Bird 2007: 55) and ‘diagnoses’ the failure as being because Armstrong, and all who favour the instantiation condition, are tacitly treating universals as categorical—defined by relations in the world, and not by any essence that they possess.

For our purposes, it does not matter if the laws needed by a strong transcendental argument only hold where the universals are instantiated. The sceptic is only concerned with worlds at which those universals are instantiated anyway. Worlds where a mental property that is a part of one of the sceptic’s prized mental propositions exists will, by the nature of that property, have whatever relations are necessary to it *being* that property. Therefore any psychophysical law that showed that if such a property were to exist, some physical property would have to exist, will hold. It does not matter if it should be the case that this law will not hold at worlds where these properties do not exist.

---

<sup>73</sup> Note that this would not be accepted by philosophers such as Timothy Williamson (2002) and David Eford (2010), who argue that any necessary truth about a particular entails the necessary existence of that particular. Hence, they would say that it is not true that ‘Geroge Orwell = Eric Blair’ is necessarily true even if Orwell/Blair does not exist at that world. There is no space to adequately address their arguments here, but I do not think it is necessary to do so. If all particulars for which there are necessary truths exist and do so necessarily, so much the better! We will have all the more reason to think that the potencies that make them of one kind or another will continue to ensure laws of nature. However, if they are wrong and Bird is right, my conclusions will still go through on a partial necessitarian view.



## 5.15 Regularity and Dispositional Essentialism

It may be wondered whether dispositional essentialism about properties might not be compatible with a mere regularity view of the laws of nature, rather than entailing metaphysical necessity. Bird concedes: ‘Strictly a regularity view of laws need not require categorical properties nor be inconsistent with dispositional essentialism’ (Bird 2007: 83). The sense in which dispositional essentialism is compatible with the regularity view of laws stems from the fact that it is not the laws that impose order upon the properties. The laws are a pattern of regular relations that supervene on the essential dispositions of potencies. They *describe* these relations, rather than governing them. In this sense they are merely ‘regularities which hold as a consequence of the essentially dispositional nature of fundamental properties’ (*ibid.*).

However, unlike the regularity view of laws associated with categoricism about properties, the sense in which the regularity view of laws that flows from dispositional essentialism supervenes upon those properties, whose powers remain the same transworld, entails that the laws are metaphysically necessary. Therefore, the only way in which a regularity view of laws can work in the absence of categorical properties is for that regularity to be metaphysically necessary. It is a derived metaphysical necessity—owed to the nature of the properties, rather than anything specific about being a law—but it is metaphysical necessity nonetheless.

## 5.16 Conceivability and the Illusion of Nomic Contingency

Having given the reasons the sceptic should accept dispositional essentialism and the consequent metaphysical necessity of the laws of nature, it must be explained why the illusion of nomic contingency—that laws are not metaphysically necessary—is so compelling. The contingency of the laws of nature is often simply assumed, suggesting a strong intuition that should be explained.

With Bird, I doubt the authority of this intuition. It does seem plausible that the tradition of regarding the laws of nature as contingent has helped form our intuitions, rather than the other way around. Bird (2007: 170) suggests that the historical origin stems from

imagining that God might have chosen other laws to govern the world. Chalmers (2002: 146) gives this definition for metaphysical possibility: ‘a world that God might have created, if he had so chosen’. Resting such a definition on the possible existence and powers of God is highly problematic, of course. Especially for our sceptic, who professes not to know if anyone but me exists. If the intuition that the laws of nature are not metaphysically necessary stems from this, she must question such an intuition.

Recall that for Hume the motivation for supposing the laws of nature could not be metaphysically necessary was because necessary connections should be either ‘either intuitively or demonstratively certain’ (Hume 1748: 4.1). The intuitively certain are known because their contrary cannot be conceived, and Hume supposed that the contrary of the laws of nature *could* be conceived. That on encountering an entirely novel object one cannot tell how it will act in conjunction with another object. However, as Chalmers (2002: 159–60) has outlined, there is a gap between *prima facie* conceivability and possibility. *Prima facie* conceivability is contrasted to ideal conceivability, where ideal conceivability is what is conceivable on ideal rational reflection. Upon *prima facie* imperfect information and reflection on an object, I may imagine all kinds of actions to be conceivable. For example, the law of gravity means that if I drop a billiard ball on Earth under normal conditions, it will fall; however, Hume would say, I can *prima facie conceive* that it might fly off in any direction, or that it might explode etc. However, this conception is incomplete in two senses: firstly, were I to become better acquainted with the object, I would find that there are many actions that are not possible due to the laws of nature operating at the actual world. The materials of which the ball is made are not volatile or likely to explode under normal conditions, and they are denser than air, thus disposed to fall through air to the ground as an operation of the law of gravity. Secondly, the conception is incomplete in terms of what rules could consistently apply at other worlds. The first sense is a matter of what Chalmers calls ‘primary conceivability’—of what could be conceived to happen at the actual world—whilst the second is a matter of ‘secondary conceivability’—of what might have been true at other possible worlds (see Chalmers 2002: 156–9).

It is not ideal primary conceivable that the ball would fly into the air or explode, and this is relatively uncontroversial. The illusion of nomic contingency comes from the *prima facie* secondary conceivability of the ball doing so at other worlds, where the laws of nature are different. However, such conceivability *must rest* upon a categoricist understanding of

162

the nature of properties. For, as we have argued, if properties are understood as dispositionally essentialist potencies, then they are defined by their powers. These powers determine a network of relations upon which the laws of nature supervene. Wherever the properties of the billiard ball truly are the same, they must necessarily react in the same ways given the same conditions. Otherwise they would not be the properties they are, the ball would not be the kind of ball it is, the conditions (density of air, proximity to a planetary body, etc.) would not be the kinds of things *they* are.

Thus we can explain the illusion of nomic contingency whilst also showing that it *is* an illusion.

## 5.17 Conclusion

There is more to be said about dispositional essentialism and the laws of nature than there is room for here. Nevertheless, I think enough has been said for my purposes. My aim here is to defend the *plausibility* of a strong transcendental response to external world scepticism. Dispositional essentialism is a plausible theory which we have good reason to think entails metaphysical necessity for the laws of nature. Moreover, the way in which this entailment works reveals a plausible mechanism via which a psychophysical law of the kind the strong transcendental theorist needs for her second premise can work. Namely, that some mental state the sceptic prizes may be necessarily possessed of a mental property, A, whose stimulus and manifestation conditions can only be given in terms of the stimulus and manifestation conditions of some extra-mental property, B, in accordance with (PP).

I do not hereby offer a completed strong transcendental argument, but I have offered a plausible account of how such a psychophysical law could be shown to be metaphysically necessary if one accepts dispositional essentialism as an account of at least fundamental properties. I have also shown that there is good reason for the sceptic to accept dispositional essentialism. This is not a complete defence, but it is enough to show that there is a case here to answer. It is not enough to assert that some sceptical hypothesis could equally well explain our mental states. It is not enough to assert that a bridge of necessity is needed to reason from mental premises to physical conclusions, and that such a bridge is impossible. The burden of proof is shifted. I have shown that there is at least one plausible line of approach that could exclude the possibility of an explanatory gap between the physical and the mental; more, there

is the potential to show that a physical explanation is necessary to give an adequate account of the mental.

The sceptic must now find reasons to defend specific metaphysical views in order to maintain her sceptical doubt. She would need to attack dispositional essentialism and the metaphysical necessity of the laws of nature; defend categorical monism about properties, or show that, under a mixed view, we have reason to think that all relevant prized mental properties are categorical—it is not at all obvious how such an argument could be made. These are not views that can be trivially dismissed. A much richer view would need to be provided by the sceptic than merely suggesting that I might be a mad, solipsistic consciousness.

With plausibility restored to the possibility of offering a strong transcendental response to scepticism, the dynamic of the dialectic with the sceptic is changed. We are not left choosing between doubt and insisting that the question itself does not make sense. We can demand greater justification from the sceptic, defences of specific views. She is not free solely to attack and shift position on whimsy. This, in itself, is no small feat.

## Conclusion

I have shown that it is entirely plausible that a strong transcendental argument could be found that offers a robust defence against the sceptic who worries that I might have all the thoughts and experiences I do now without the existence of an extra-mental world.

I have shown that transcendental arguments as a whole are well-suited to addressing the question on the basis that they start from premises the sceptic accepts and can take seriously her standards for justification. As the conclusion of a transcendental argument is shown to be a necessary condition of the truth of a proposition concerning the mental that the sceptic prizes, the justification she values having in such a proposition is transferred across to the conclusion.

I have established that strong transcendental arguments are to be preferred to weak transcendental arguments, and I have defended the plausibility of such arguments. I have shown that it is not necessary for a strong transcendental argument to incorporate a suppressed verification principle, and that the supposition that that it rests on the deeper criticism that it is not possible to reason from premises concerning the mind to a conclusion concerning the world. I have shown that this deeper criticism rests on the mistaken assumption that the mind is not a part of the world. I have shown that we can understand the relationship of mind to the extra-mental world as being one of a category of facts (mental facts) that supervenes on other categories of facts—in particular, physical facts. Any argument that it is not possible to have entailment relations between mental facts and extra-mental facts has been shown to rest on an illicit assumption about the nature of the relationship between mind and the extra-mental world that we have been given no reason to think plausible. I have shown that an alternate picture of the relationship of mind to extra-mental world is available in the form of supervenience physicalism, and I have defended the plausibility of this picture.

Lastly, I have given a plausible account of how the second premise of a transcendental argument could work, based on a dispositional essentialist understanding of properties, which would entail a metaphysically necessary conception of laws of nature, on which it would be plausible to offer a psychophysical law of a form that could be used as the second premise of a strong transcendental argument.

I have not here given a strong transcendental argument. The development and defence of any such argument would be a much larger project than there is room for here. However, I

have shown that the materials are available from which one plausibly might construct such an argument. Moreover, I have shown that any attempt to reject this plausibility will have to rely upon the defence of specific philosophical positions, such as resemblance nominalism concerning the laws of nature or a specific and limiting conception of the relationship between mind and world. This forces the sceptic into a more definitive position than simply insisting that a sceptical scenario could just as easily account for the thoughts and experiences we have as the supposition that the extra-mental world exists. To defend the thought that no strong transcendental argument could be offered against her doubt, she must adopt philosophical positions that we, ourselves, may not agree with. This shifts her from her position of voicing a natural doubt many humans feel from time to time and forces her into adopting positions that many of us would reject.

Whilst scepticism itself has not been refuted, the mere plausibility of offering a strong transcendental argument against this type of sceptic can offer us a certain amount of comfort by revealing something about how the world would have to be in order for a strong transcendental argument to not be possible.

## Bibliography

Al-Ghazali (2002), *Deliverance from Error and Mystical Union with the Almighty*, Muhammad Abulaylah (trans.) and George F. McLean (ed.), <<http://www.ghazali.org/books/md/>>, accessed 7<sup>th</sup> May 2013, Council for Research in Values and Philosophy

Andrews, F M, and Withey, S B (1976), *Social Indicators of Well-Being*, New York: Plenum Press

Anzai, Tetsuo (1999), *Shakespeare in Japan*, Lewiston, NY: Lampeter

Armstrong, David Malet (1983), *What is a Law of Nature?* Cambridge: Cambridge University Press.

Augustine (1991), *Confessions*, Henry Chadwick (int. and trans.), Oxford: Oxford University Press

Augustine (1995), *Against the Academics and the Teacher*, Peter King (int. and trans.), Indianapolis: Hackett Publishing Company

Austin, J L (2003), 'Are There *A Priori* Concepts?' (pp. 32–54), *Philosophical Papers*, J. O. Urmson, and G. J. Warnock (eds), Oxford Scholarship Online, <<http://www.oxfordscholarship.com/view/10.1093/019283021X.001.0001/acprof-9780192830210-chapter-2>>, accessed 5<sup>th</sup> November 2008

Bailey, Sally (2011), 'Play as a Staging Ground for Performance and Life' (pp. 137–154), *Play and Performance: Play and Cultural Studies*, Vol. 11, Carrie Lobman and Barbara E. O'Neil (eds), Lanham, Maryland: University Press of America

Bayne, Tim and Spencer, Maja (2010), 'Introspective Humility' (pp.1-22), *Philosophical Issues* 20, *Philosophy of Mind*

Berkeley, George (1998), *Three Dialogues between Hylas and Philonous*, Jonathan Dancy (ed. and intro.), Oxford: Oxford University Press

Bett, Richard (2010), 'Introduction' (pp. 1–10), *The Cambridge Companion to Ancient Scepticism*, Richard Bett (ed.), Cambridge and New York: Cambridge University Press

Beck, Aaron T, Brad A Alford (2009), *Depression: Causes and Treatment*, Philadelphia: University of Pennsylvania Press

Berkeley, George (1998), *A Treatise Concerning the Principles of Human Knowledge*, Jonathan Dancy (ed.), Oxford and New York: Oxford University Press

Bird, Alexander (2007), *Nature's Metaphysics: Laws and Properties*, Oxford: Oxford University Press

Blackburn, Simon (1993), 'Circles, Finks, Smells and Biconditionals' (pp. 259–279), *Philosophical Perspectives*, Vol. 7

Brecht, Bertolt (1961), 'On Chinese Acting' (pp. 130-36), *The Tulane Drama Review*, Vol. 6, No. 1, Eric Bently (trans.)

Brooks, E. Bruce, and A. Taeko Brooks (1998) *The Original Analects*, New York: Columbia University Press

Burge, Tyler (1986), 'Individualism and Psychology' (pp. 3–45), *The Philosophical Review*, Vol. 95, Durham, NC: Duke University Press



Care Quality Commission (2015), *Monitoring the Mental Health Act in 2013/14*, <<http://www.cqc.org.uk/content/mental-health-act-annual-report-201314>>, accessed 17<sup>th</sup> April 2015

Cartwright, Nancy (2002), 'In Favor of Laws that are not *Ceteris Paribus* After All' (pp. 425-39), *Erkenntnis*, Vol. 57

Chakravartty, Anjan (2014), 'Scientific Realism', *The Stanford Encyclopedia of Philosophy*, (Spring 2014 Edition), Edward N. Zalta (ed.), <<http://plato.stanford.edu/archives/spr2014/entries/scientific-realism/>>, accessed 19<sup>th</sup> March 2015.

Chalmers, David (2002), 'Does Conceivability Entail Possibility?' (pp. 145–200), *Conceivability and Possibility*, Tamar Szabó Gendler and John Hawthorne (eds), Oxford and New York: Oxford University Press

Cicero, (1933), *De natura deorum; Academica*, H. Rackham (trans.), London: William Heinemann

Cinpoedș, Nicoleta (2010), *Shakespeare's Hamlet in Romania, 1778-2008: A Study in Translation, Performance, and Cultural Adaptation*, Lewiston, NY: Lampeter

Clark, Andy (2008), *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford: Oxford University Press

Coleridge, Samuel Taylor (2004), *Biographia Literaria*, Tapio Riikonen and David Widger (eds), Project Gutenberg, accessed 15<sup>th</sup> April 2013, last updated 26<sup>th</sup> January 2013

Crane, Tim and D. H. Mellor (1990): 'There is No Question of Physicalism' (pp. 185-206), *Mind*, Vol. 99, Oxford: Oxford University Press

Crane, Tim (1991), 'Why Indeed? Papineau on Supervenience' (pp. 32-37), *Analysis*, Vol. 61, Oxford: Oxford University Press

Cronenberg, David (1998), *eXistenZ*, UK: Momentum Pictures

Currie, Gregory and Jon Jureidini (2001), 'Delusion, Rationality, Empathy' (159–62), *Philosophy, Psychiatry and Psychology*, Vol. 8.

Davidson, Donald (1980), 'Mental Events', *Essays on Action and Events*, Oxford: Clarendon Press

Davidson, Donald (2002a), 'Knowing One's Own Mind' (pp. 15–38), *Subjective, Intersubjective, Objective*, Oxford University Press: Oxford

Davidson, Donald (2002b), 'A Coherence Theory of Truth and Knowledge' (pp. 137–57), *Subjective, Intersubjective, Objective*, Oxford University Press: Oxford

Davidson, Donald (2002c), 'Empirical Content' (pp. 159–75), *Subjective, Intersubjective, Objective*, Oxford University Press: Oxford

Davidson, Donald (2002b), 'Three Varieties of Knowledge' (pp. 205–20), *Subjective, Intersubjective, Objective*, Oxford University Press: Oxford

Dennett, Daniel C (1978), *Brainstorms*, Cambridge, MA: Bradford Books

Dennett, Daniel C (1991), *Consciousness Explained*, London: Penguin

Descartes, René (1996), *Meditations on First Philosophy: With Selections from the Objections and Replies*, John Cottingham (ed. and trans.), Cambridge: Cambridge University Press

Descartes, René (1932), 'The Principles of Philosophy', *The Philosophical Works of Descartes*, Elizabeth S Haldane and G R T Ross (eds and trans.), Cambridge: Cambridge University Press

Dretske, Fred (1977), 'Laws of Nature' (pp. 248–68), *Philosophy of Science*, Vol. 44

Dretske, Fred (2004), 'Change Blindness' (pp. 1-18), *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 120

Earman, John, John Roberts, and Sheldon Smith (2002), 'Ceteris Paribus Lost' (pp. 281-301), *Erkenntnis*, Vol. 57

Efird, David (2010), 'Is Timothy Williamson a Necessary Existent?' (pp. 97-108), *Modality: Metaphysics, Logic, and Epistemology*, Oxford: Oxford University Press,

Ellis, Brian (2001), *Scientific Essentialism*, Cambridge: Cambridge University Press.

Ellis, Brian (2002), *The Philosophy of Nature*, Chesham: Acumen

Ellis, Brian (2006), 'Physical Realism' (pp. 1–13), *Metaphysics in Science*, Malden, MA and Oxford: Blackwell Publishing

Frances, Bryan (2005), *Scepticism Comes Alive*, Oxford: Oxford and New York

Ghent, Henry of (2002), *The Cambridge Translations of Medieval Philosophical Texts*, Robert Pasnau (trans.), Cambridge: Cambridge University Press

Haybron, Daniel M. (2007), 'Do We Know How Happy We Are? On Some Limits of Affective Introspection and Recall' (pp. 394–428), *Noûs*, Vol. 41, No. 3

Henrich, Dieter (1989), 'Kant's Notion of a Deduction and the Methodological Background of the First *Critique*', *Kant's Transcendental Deductions*, Eckart Förster (ed.), Stanford: Stanford University Press

Hirstein, William and V S Ramachandran (1997), 'Capgras Syndrome: A Novel Probe for Understanding the Neural Representation of the Identity and Familiarity of Persons' (pp. 437–44), *Proceedings: Biological Sciences*, Vol. 264.

Hirstein, William (2010), 'The Misidentification Syndromes as Mindreading Disorders' (pp. 233–60), *Cognitive Neuropsychology*, Vol. 15

Hobbes, Thomas (2007), 'Objection 1' (p. 42), *Objections to the Meditations and Descartes's Replies*, Jonathan Bennet (ed.), <<http://www.earlymoderntexts.com/pdf/descobje.pdf>>, last updated November 2007, accessed 21<sup>st</sup> April 2013

Huang, Alexa (2011), 'Shakespeare and Translation' (pp. 68–105), *The Edinburgh Companion to Shakespeare and the Arts*, Stockport: University of Edinburgh Press.

Hume, David (1748 and 1777), *An Enquiry Concerning Human Understanding*, Peter Millican and Amyas Merivale (eds), <<http://www.davidhume.org/texts/thn.html>>, davidhume.org, last updated May 2012, accessed 18<sup>th</sup> May 2015

Hume, David (1739–40), *A Treatise of Human Nature*, accessed online from <http://www.davidhume.org/texts/thn.html>, Peter Millican and Amyas Merivale (eds), davidhume.org, last updated May 2012, accessed on 5<sup>th</sup> April 2014

Jackson, Frank (2000), 'Serious Metaphysics and Supervenience' (pp. 1-?), *From Metaphysics to Ethics*, Oxford: Oxford University Press

Johnston, Mark (1992), 'How to Speak of the Colours' (pp. 221–63), *Philosophical Studies*, Vol. 68

Johnston, Mark (1993), 'Objectivity Refigured: Pragmatism Without Verificationism' (pp. 85–130, *Reality, Representation, and Projection*, John Haldane and Crispin Wright (eds), Oxford: Oxford University Press

Kant, Immanuel (1962), *Prolegomena to any Future Metaphysics that will be able to Present Itself as a Science*, Peter G. Lucas (trans.), Manchester: Manchester University Press

Kant, Immanuel (1996), *Critique of Pure Reason*, Werner S. Pluhar (trans.) and Patricia Kitcher (intro.), Cambridge, Indianapolis: Hackett

Klima, Gyula, Fritz Allhoff, and Anand Jayprakash (2007), *Medieval Philosophy: Essential Readings with Commentary*, Oxford and Malden, MA: Blackwell Publishing

Koller, John M. (2000), 'Syādvāda as the Epistemological Key to the Jaina Middle Way Metaphysics of *Anekāntavāda*' (pp. 400-7), *Philosophy East and West*, Vol. 50, No. 3

Kripke, Saul (1981), *Naming and Necessity*, Malden, MA and Oxford: Blackwell Publishing

Kumar, Dehdev (2001), *Jain Temples of Rajasthan*, Janpath, New Delhi: Indira Gandhi National Centre for the Arts

Lao-tzu (1891), 'Tao Te Ching', *Sacred Books of the East*, Vol 39, J. Legge (trans.), <<http://www.sacred-texts.com/tao/taote.htm>>, accessed 7<sup>th</sup> May 2013

Lewis, David Kellogg (1973), *Counterfactuals*, Oxford: Blackwell Publishing.

Litvin, Margaret (2011), *Hamlet's Arab Journey : Shakespeare's Prince and Nasser's Ghost*, Princeton, N J: Princeton University Press

Loewe, Michael and Edward L. Shaughnessy (1999), *The Cambridge History of Ancient China: From the Origins of Civilization to 221 BC*, Cambridge: Cambridge University Press

Martin, C B (1994), 'Dispositions and Conditionals' (pp. 1-8), *The Philosophical Quarterly*, Vol. 44

McManus, Sally, Howard Meltzer, Traolach Brugha, Paul Bebbington, and Rachel Jenkins (eds) (2009), 'Key Facts', *Adult Psychiatric Morbidity in England—2007, Results of a Household Survey*, Health and Social Care Information Centre, <<http://www.hscic.gov.uk/pubs/psychiatricmorbidity07>>, accessed 14 April 2015, last updated 27 January 2009

Nolan, Christopher (2010), *Inception*, UK and USA: Warner Bros Pictures

O'Daly, George (2001) 'The response to scepticism and the mechanisms of cognition' (pp. 159-70), *The Cambridge Companion to Augustine*, Eleonore Stump and Norman Kretzmann (eds), Cambridge: Cambridge University Press

Papineau, David (1990), 'Why Supervenience?' (pp. 66-71), *Analysis*, Vol. 50, Oxford: Oxford University Press

Papineau, David (1991), 'The Reason Why: Response to Crane' (pp. 37-40), *Analysis*, Vol. 51, Oxford: Oxford University Press

Papineau, David (2004), 'Supervenience and Identity', *Philosophical Naturalism*, <<http://www.kcl.ac.uk/ip/davidpapineau/Staff/Papineau/PhilNat2nded/ch1%20revised.htm>>, accessed 14<sup>th</sup> February 2010

Passer, Kevin M and Julia K Warnock (1991), 'Pimozide in the Treatment of Capgras' Syndrome: A Case Report' (pp. 446-8), *Psychosomatics*, Vol. 32

Peirce, C S (1902) *Collected Papers of Charles Sanders Peirce*, 8 vols, vols. 1-6, C Hartshorne and P Weiss (eds), Cambridge MA: Harvard University Press

Pellegrin, Pierre (2010), 'Sextus Empiricus' (pp. 120–41), *The Cambridge Companion to Ancient Scepticism*, Richard Bett (ed.), Cambridge and New York: Cambridge University Press

Phillips, D. P. and Wills, J. S. (1987), 'A Drop in Suicides Around Major National Holidays' (pp. 1–11), *Suicide and Life Threatening Behaviour*, Vol.17

Plato (2002), *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo*, G M A Grube (trans.), revised by John M Cooper, Indianapolis and Cambridge: Hackett

Psillos, Stathis (1999), *Scientific Realism: How Science Tracks the Truth*, London and New York: Routledge

Psillos, Stathis (2006), 'Scientific Realism and Metaphysics' (pp. 14–32), *Metaphysics in Science*, Malden, MA and Oxford: Blackwell Publishing

Putnam, Hilary (1973), 'Meaning and Reference' (pp. 699–711), *The Journal of Philosophy*, Vol. 70, New York: Journal of Philosophy

Putnam, Hilary (1975), 'The Meaning of "Meaning"' (pp. 131–93), *Minnesota Studies in the Philosophy of Science*, Vo. 7

Putnam, Hilary (1993), *Reason, Truth, and History*, Cambridge: Cambridge University Press

Robinson, Don (1991), 'On Crane and Mellor's Argument against Physicalism' (pp. 135–6), *Mind*, Vol. 100

Rodriguez-Pereyra, Gonzalo (2001), 'Resemblance Nominalism and Russell's Regress' (pp. 395–408), *Australasian Journal of Philosophy*, Vol. 79

Rodriguez-Pereyra, Gonzalo (2002a), *Resemblance Nominalism*, Oxford: Clarendon Press.

Rodriguez-Pereyra, Gonzalo (2002b), 'The Problem of Universals and the Limits of Conceptual Analysis' (pp. 29–47), *Philosophical Papers*, Vol. 31

Rodriguez-Pereyra, Gonzalo (2003), 'Resemblance Nominalism and Counterparts: Reply to Bird' (pp. 229–37), *Analysis*, Vol. 63

Rose, David, Wesley Buckwalter, and John Turri (2014), 'When Words Speak Louder than Actions: Delusion, Belief, and the Power of Assertion' (pp. 683–700), *Australasian Journal of Philosophy*, Vol. 92

Rosenberg, Alex and McShea, Daniel W. (2008), *Philosophy of Biology: A Contemporary Introduction*, New York and London: Routledge

Russell, Bertrand (1956), *An Inquiry into Meaning and Truth*, London: George Allen and Unwin

Russell, Bertrand (2008), *The Problems of Philosophy*, Rockville, Maryland: Arc Manor

Schwitzgebel, Eric (2008), 'The Unreliability of Naive Introspection', (pp. 245-273), *Philosophical Review*, Vol. 117

Sextus Empiricus (2005), *Outlines of Scepticism*, Julia Annas and Jonathan Barnes (eds and trans.), Cambridge and New York: Cambridge University Press

Shakespeare, William (1997), *The Riverside Shakespeare: Second Edition, The Complete Works*, G Blakemore Evans, J J M Tobin, et al (eds), Boston and New York: Houghton Mifflin Company

Shakespeare, William (Wil'yam Shex'pir) (2000), *The Klingon Hamlet: The Restored Klingon Version*, Nick Nicolas and Andrew Strader (trans.), New York: Pocket Books



Shoemaker, Sydney (1984), *Identity, Cause, and Mind*, Cambridge: Cambridge University Press

Smith, A D (1977), 'Dispositional Properties' (pp. 439-44), *Mind*, New Series, Vol. 86

Sosa, Ernest (2007), *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, Oxford: Oxford University Press

Swoyer, Chris (1982), 'The Nature of Natural Law' (pp. 203–23), *Australasian Journal of Philosophy*, Vol. 60,

Sparhawk, T. G. (1987), 'Traditional Holidays and Suicide' (pp. 245-246), *Psychological Reports*, Vol. 60

Spitzer, Robert L., Janet B. W. Williams, Kurt Kroenke, and colleagues (2010a), 'Patient Health Questionnaire-9 (PHQ-9)', *Patient Health Questionnaire (PHQ) Screeners*, Pfizer, < <http://www.phqscreeners.com/>>, accessed 8<sup>th</sup> May 2015, last updated April 2010.

Spitzer, Robert L., Janet B. W. Williams, Kurt Kroenke, and colleagues (2010b), 'GAD-7', *Patient Health Questionnaire (PHQ) Screeners*, Pfizer, < <http://www.phqscreeners.com/>>, accessed 8th May 2015, last updated April 2010.

Stanislavski, Constantin (1961), *Creating a Role*, Joshua Logan (int.) and Elizabeth Reynolds Hapgood (trans.), New York: Theatre Arts Books

Stern, Robert (2000), *Transcendental Arguments and Scepticism: Answering the Question of Justification*, Oxford: Clarendon Press

Stoneham, Tom (2004), 'Self-Knowledge' (pp. 647-672), *Handbook of Epistemology*, M Sintonen Niiniluoto and J Woleński (eds), Netherlands: Kluwer Academic Publishers

Strawson, P F (1976), *The Bounds of Sense: A Essay on Kant's Critique of Pure Reason*, London: Methuen

Strawson, P F (1985), *Skepticism and Naturalism: Some Varieties*, London: Methuen

Strawson, P F (1987), *Individuals: An Essay in Descriptive Metaphysics*, London and New York: Methuen

Stroud, Barry (1979), 'The Significance of Scepticism' (pp. 277–97), *Transcendental Arguments and Science: Essays in Epistemology*, Peter Bieri, Rolf-Peter Horstmann, and Lorenz Krüger (eds), Dordrecht, Holland and Boston USA: D Reidel Publishing Company

Stroud, Barry (2000a), 'Transcendental Arguments', *Understanding Human Knowledge: Philosophical Essays* (pp. 9-25), Oxford: Oxford University Press

Stroud, Barry (2000b), 'Kantian Argument, Conceptual Capacities, and Invulnerability', *Understanding Human Knowledge: Philosophical Essays* (pp. 155-176), Oxford: Oxford University Press

Stroud, Barry (2000c), 'Radical Interpretation and Philosophical Scepticism', *Understanding Human Knowledge: Philosophical Essays* (pp. 177-202), Oxford: Oxford University Press

Stroud, Barry (2000d), 'The Goal of Transcendental Arguments', *Understanding Human Knowledge: Philosophical Essays* (pp. 203-223), Oxford: Oxford University Press

Svavarsson, Svavar Hrafn (2010), 'Pyrrho and Early Pyrrhonism' (pp. 36–57), *The Cambridge Companion to Ancient Scepticism*, Richard Bett (ed.) Cambridge and New York: Cambridge University Press.

Tooley, Michael (1977), 'The nature of laws' (pp. 667–98), *Canadian Journal of Philosophy*, Vol. 7

Unger, Peter (2003), *Ignorance: A Case for Scepticism*, Oxford Scholarship Online: <<http://www.oxfordscholarship.com/view/10.1093/0198244177.001.0001/acprof-9780198244172>>, accessed 12<sup>th</sup> March 2013; print publication date 1978

Wachoski, Andy and Lana Wachowski (1999), *The Matrix*, Australia and USA: Warner Bros Pictures

Weir, Peter (1998), *The Truman Show*, USA: Paramount Pictures

Weiskrantz, L (1986), *Blindsight: A Case Study and Implications*, Oxford: Clarendon Press

Whiteley, C H (1975), 'Ignorance: A Case for Scepticism By Peter Unger Clarendon Press: Oxford University Press, 1975, vii + 319 £6.50' (pp. 489–90), *Philosophy*, Vol. 50, Issue 194, published online 25<sup>th</sup> February 2009

Williams, B (1978), *Descartes: The Project of Pure Enquiry*, Harmondsworth: Penguin

Williamson, Timothy (2002), 'Necessary Existents' (pp. 233–51), *Royal Institute of Philosophy Supplement*, Vol. 51

Wright, Crispin (1992) 'On Putnam's Proof that we are Not Brains in a Vat' (pp. 67-94), *Proceedings of the Aristotelian Society*, New Series, Vol. 92