

2475/3924

MONASH UNIVERSITY
THESIS ACCEPTED IN SATISFACTION OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ON..... [REDACTED] October 2002

.....
/ Sec. Research Graduate School Committee

Under the copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing for the purposes of research, criticism or review. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

Confabulating Consciousness

Alan Crooke

BBSoc AMusA DipEdPsych MA



Submitted for the degree of
Doctor of Philosophy

April 2002

School of Philosophy and Bioethics
Monash University

This thesis is dedicated in memory of

David K Lewis

1941 – 2001

Abstract	7
Declaration	9
Acknowledgements	11
Conventions and Abbreviations	12

Confabulating Consciousness

Introduction	15
--------------	----

PART 1 The Materialist Background

1.0	INTRODUCTION: THE HISTORICAL PRIMACY OF MATERIALISM	23
1.1	DEFINING MATERIALISM – LEWIS	
1.1.1	<i>Lewis as a principal theorist of contemporary materialism</i>	25
1.1.2	<i>Properties: fundamental and physical</i>	25
1.1.3	<i>Lewis's definition of materialism</i>	31
1.2	DEFINING MATERIALISM – JACKSON	
1.2.1	<i>Basic terminology and assumptions</i>	36
1.2.2	<i>A critique of minimality</i>	39
1.2.3	<i>Entailment by physical facts</i>	42
1.2.4	<i>Possible modifications of the entry by entailment thesis</i>	44
1.3	DEFINING MATERIALISM – OTHER APPROACHES	
1.3.1	<i>Campbell</i>	46
1.3.2	<i>Tye</i>	48
1.3.3	<i>Chalmers</i>	50
1.3.4	<i>Van Fraassen</i>	52
1.4	DEFINING MATERIALISM – A RADICAL PROPOSAL	
1.4.1	<i>Physics before Descartes – a useful ancient myth</i>	55
1.4.2	<i>Descartes's physics and after – a useful modern myth</i>	60
1.4.3	<i>Actual materialism</i>	62
1.4.4	<i>Necessary Materialism</i>	65
1.4.5	<i>Consequences, objections, and replies</i>	66

PART 2 Necessity and Nature

2.0	INTRODUCTION: A NEW LOOK AT MODALITY	75
2.1	WHY MODAL REALISM?	
2.1.1	<i>Explication of modal realism</i>	78
2.1.2	<i>Ersatzism</i>	82
2.1.3	<i>Combinatorialism and other fictionalisms</i>	84
2.1.4	<i>Neo-Spinozism: lip-service possibilism?</i>	84

2.2	DEFENDING MODAL REALISM	
2.2.1	<i>Counterintuitive?</i>	87
2.2.2	<i>A wildly inflated ontology?</i>	88
2.2.3	<i>Too many worlds?</i>	89
2.2.4	<i>Concrete worlds? Duplicates?</i>	92
2.2.5	<i>Arbitrary worldmates? Arbitrary worlds?</i>	96
2.2.6	<i>Occam defied?</i>	100
2.3	RECOMBINATION RECONSIDERED	
2.3.1	<i>Lego™-worlds or logo-worlds?</i>	102
2.3.2	<i>Distinct existences, necessary connexions, human algebra</i>	104
2.3.3	<i>Problems specific to Lewis's worlds</i>	113
2.3.4	<i>Warrant for modal beliefs – the 'antilogical' turn</i>	122
2.4	ANTHROPIC REASONING AND MODALITY	
2.4.1	<i>Transcendental reasoning: the transcendental and the rational</i>	127
2.4.2	<i>Ersatzism and 'equiprobability' of worlds</i>	134
2.4.3	<i>An anthropic line articulated</i>	141
2.4.4	<i>Objections to the anthropic line</i>	146
2.5	FEWER WORLDS THAN ARE DREAMT OF	
2.5.1	<i>Introducing paranthropic reasoning</i>	166
2.5.2	<i>First preliminary: the 'epiphenomenal' objection to MR</i>	167
2.5.3	<i>Second preliminary: the future irregularity objection to MR</i>	171
2.5.4	<i>A pure paranthropic argument</i>	177
2.6	CONCLUSION: MODERATE NECESSITARIANISM	
2.6.1	<i>Some addenda</i>	183
2.6.2	<i>Summing up</i>	188

PART 3 The Dualist Temptation

3.0	INTRODUCTION: UNEASE WITH MATERIALIST MONISM	195
3.1	MODALITY AND NATURE IN CHALMERS' DUALISM	
3.1.1	<i>The taxonomy of modalities</i>	199
3.1.2	<i>Modal ontology and Chalmers</i>	210
3.2	SUPERVENIENCE: WORLD-TALK OR WORD-TALK?	
3.2.1	<i>Some history</i>	217
3.2.2	<i>The varieties and uses of supervenience theses</i>	219
3.2.3	<i>Natural supervenience: a restricted supervenience?</i>	220
3.3	QUALIA AND THE ZOMBIES	
3.3.1	<i>The secret life of zombies</i>	231
3.3.2	<i>Fermi and the dancing zombies</i>	233
3.3.3	<i>In search of Russell's neutral monism</i>	236
3.4	CONCLUSION: CONFABULATING CONSCIOUSNESS	241

Bibliography	243
---------------------	-----

ABSTRACT

I make a case for removing (phenomenal) consciousness from our metaphysical inventory of the world. I do this in three stages:

In Part 1 I adduce evidence that materialism has excellent 'historical' credentials as a default assumption, and examine ancient and modern understandings (Descartes representing a watershed between these) of the monism or dualism of the whole causal domain we call *the world*. On the basis of this history-of-ideas study, and examination of the work of both major and representative minor contemporary theorists, I conclude that materialism is not easily definable because coherent alternatives to it are elusive. By similar investigation I conclude that the best way to construe materialism is as a necessarily true and benignly vacuous thesis. Divisions of the world, and of all of modal space, into physical and non-physical compartments result in infelicities at best, and insoluble theoretical difficulties at worst. The true physics of a world must yield the whole causal nature of that world. This is my *necessary materialism*.

In Part 2 I establish a *moderate necessitarianism*, underpinned by: Lewisian modal realism, which I defend as providing the best set of assumptions to guide modal speculation and to rectify modal intuitions (I also defend it as much more acceptable as true metaphysics than most have granted); rejection of Lewis's principle of recombination (wrongly called *Humean*); and certain powerful *anthropic* inductive inferences and equally powerful *paranthropic* (my term) inductive inferences from the nature of the actual world to that of the aggregate of all worlds. The result is a Lewis-style modal-realist ontology, conjoined with equating the broadly (or 'logically') possible and the 'naturally' possible, so that there are far more restrictions on the kinds of worlds than Lewis would have. Most of the important theoretical innovations are in this part, motivated and abetted by a concern for 'ontological seriousness'.

In Part 3 I deploy the results from the first two parts in a critique of David Chalmers' tempting and brilliantly argued *naturalistic dualism* – the most formidable representative of current dissent from broad monist orthodoxy. I continue from Part 2 my program of abandoning the standard unruly distinctions between some 'natural' possibility (along with other restricted possibilities) and possibility *simpliciter*. I argue also against reliance on notoriously ill-defined Russellian neutral monism, which inspires Chalmers and others. I show how my approach can be used in criticism of Chalmers' famous thought experiments with qualia and qualia-less zombies. I take a Dennettian line on qualia, which I extend to making plausible the elimination of consciousness itself. I adapt the neuropathological notion of confabulation, arguing that it is unexpectedly pervasive in non-pathological mental life. From the short history and patchy cross-cultural spread of notions of consciousness, I sketch a case for phenomenal consciousness itself being an entrenched confabulation, linguistically and socially mandated, as a reason both for the idea's seeming ineradicability and for its intractable philosophical awkwardness. I conclude by re-affirming a necessary 'materialist' monism, acknowledging that this term could be misleading, or even empty – but *benignly empty*.

DECLARATION

This thesis contains no material that has been accepted for the award of any other degree or diploma in any university; and, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due reference is made in the text.



Alan Crooke

ACKNOWLEDGEMENTS

The writing of a thesis can seem simultaneously to be the loneliest and the most socially connecting experience. Many people have helped me in this; and many involvements go back a long way.

In university circles, thanks are due to:

Frank Jackson (a supervisor for much of the time this work was in gestation); Aubrey Townsend (to whom I owe a great deal); the late Harry Stainsby and the late John McGeachie (both laudably heretical and oddly influential, in their different ways); Jamie Dreier (valued and intrepid correspondent, fierce pedant); Lloyd Humberstone (equally fierce logician); Stuart Brock (for saving me from some of the even more outrageous things I would have included concerning the grand Lewisian tradition); Graham Oppy (for very helpful and judicious directional advice); Karen Green; Dirk Baltzly (for erudite assistance with matters Greek, and more); Stephen Barker; John Heil (brief dialogue with whom confirmed me in my convictions regarding ontological seriousness); Steve Curry and Steve Clarke (my very forbearing and closest colleagues as I completed this work); Steve Mathews; Robert Pargetter; Jack Smart (very active and benign elder statesman); Linda Peach and Lesley Hamilton (the world's pre-eminent departmental administrative professionals), and their colleagues; Michael Pendlebury; Alex Pruss; Daniella Forster (for efficient research assistance); Callan Ledsham (for enthusiastic support, and many excellent leads on formative – and some forgotten – philosophers of an earlier time); Bill Webster (very encouraging!); other staff, and students, at Monash University, Charles Sturt University, and University of the Witwatersrand – the three universities at which I have revelled in the arduous delight of teaching philosophy.

And mostly outside the academy, but certainly in my own circle, I thank:

Eva Scheer (I know! it was hard); Julian Crooke (for his filial readiness to help; he also knows something of what it's like); Carmel Bird (great friend); Harry Edhouse (worthy philosophical companion); Karl Williams; Frodo (who should have persisted with those Tractarian opuscles way back in the seventies, which Frank Jackson still recalls in anecdote – for a haven of hospitality when it was needed); all of my friends not so far mentioned, and especially Annette Davey, whose eye for detail is unmatched (almost!).

I thank also:

Vivien Wheeler of *The Bridgeman Art Library* (www.bridgeman.co.uk) and Mark Harden of *The Artchive* (www.artchive.com) for kind permission to use the fine art images I have enigmatically included. (These are, in order: Vermeer, *The Lacemaker*; Holbein, *The Ambassadors*, detail; Vermeer, *The Astronomer*; Vermeer, *The Geographer*; and Gossaert, *Danae*.)

I thank especially and separately John Bigelow, who engaged seriously with this work, was maieutic to some of my best ideas, tolerated my floridity of expression where others just got bemused, and helped me continue to suspect that I might have something rather original and interesting to say here – which, partly thanks to him, I have!

CONVENTIONS AND ABBREVIATIONS

Conventions for quoted material:

All square brackets [] and text enclosed by them within quoted material are my own.

All of my own ellipses in quoted material are in square brackets [...].

All *italics* in quotes are in the original.

All underlining in quoted material is to show my own added emphasis.

Other conventions:

There are no footnotes.

Still less are there any endnotes (the bane of civilised intellectual life).

Parts 1, 2, and 3 are divided into *sections* (labelled like this: 3.2).

Sections are divided into *subsections* (labelled like this: 3.2.2).

<i>Abbreviation:</i>	<i>Explanation:</i>	<i>Page where first used:</i>
(C)	the thesis (C) (Jackson)	38
CTA	coarse-tuning argument	149
FIO	future irregularity objection	171
FTA	fine-tuning argument	142
LR	Lewisian recombination	80
M Γ D	minimum Γ [gamma] duplicate	40
M Φ D	minimum physical duplicate	40
M Ψ D	minimum psychological duplicate	40
MR	modal realism (i.e. <i>concrete</i> modal realism)	78
PNP	perfectly natural property (Lewis)	26
PoI	principle of indifference	137
PP	paranthropic argument	167
SAP	strong anthropic principle	130
w_a	the actual world	43
WAP	weak anthropic principle	130

Confabulating Consciousness



Introduction

Introduction

This is an essay in ontologically serious philosophy. When I set out to write a thesis in philosophy of mind I had a reasonably coloured-in picture of the destination. My philosophical instincts were materialist, or at least monist, and whatever the term might mean I felt it in my bones that I would never object to being called a *naturalist*. I have never been able to see the problem with naturalism: should we naturalise epistemology? How about aesthetics? And surely someone should have a go at ethics! But as for *mind*... well, that seems to many people to be an especially hard case. Somewhere in the winding journey ahead of us I shall quote David Chalmers asserting that 'moral facts are not phenomena that force themselves on us' as the facts of phenomenal experience do. There is something very privileged and unfathomable, it seems to both philosophers and non-philosophers, about experience, which cannot be investigated as any of the *objects* of experience are. Very well. But I could never make sense of the idea that there was something that had to be *freed* from the natural realm here; or indeed that there was something that needed to be *re-corralled*. I cannot remember ever thinking that anything could ever *fail* to be natural – so the term itself faded into meaninglessness, as we shall find Kim Sterelny noting, also. Where do these philosophical predispositions come from? There's a good research program for some PhD student!

Say what you will about the world, I thought, it has to be *one* world. A naturalistic, materialistic, whole – well characterised as a web of causal connectivity, whose empirical investigation could be undertaken by physical sciences, in a very inclusive sense of *that* expression. And all of these terms seemed to me equally vain in the end, because there was nothing to oppose them. They had no opposites: always a bad sign when it comes to picking out anything real! So when I encountered the work of Chalmers, his so-called *naturalistic dualism*, I was intrigued. I had thought I would be writing a thesis about qualia and functionalism, and connectivity and perhaps if I felt daring emergentism, and all that. In short, I thought I would write within the domain of

mind, and not ever have to stray too far outside into the nearby sciences and the nearby areas of philosophical inquiry. How wrong I was!

Reading Chalmers from the point of view of a secret sympathiser with David Lewis's modal realism, I just *knew* (I'm ashamed to say) that something must be wrong here. Chalmers was claiming to be a naturalist, but not a physicalist: nor even a monist (at least not a materialist monist). I couldn't even understand how that could make sense *prima facie*. I now feel confident displaying this disability, because I think I have the arguments to show that my feeling is warranted. I think I fail to understand in just the right places. But I was uneasy for quite a while at the start.

I am still a Lewisian sympathiser, and have developed a modal realist position of my own, which will be exposed to scrutiny about one hundred pages from now. It differs from Lewis's, but I am certain I would never have come to it now without having read *Plurality of Worlds* long ago. A great deal of this present work is inspired by his work, but I'm sure that if Lewis were alive to read it today he would not be impressed by my abandoning the so-called 'Humean' principles that he held so dear and thought were pretty well to be taken as axioms. My naturalism (if this can mean anything) extends well out into modal space; I hold that scientific, empirical investigation of our minimal sample of just one world can tell us surprisingly many things about the worlds. With the greatest respect to Lewis (and this thesis is dedicated in memory of him), I cannot agree with a certain element of Chalmers' assessment of Lewis at the memorial service when he praised and thanked Lewis in such a genuine and touching way. I read that Chalmers believed Lewis had

a beautiful simple system: a fundamental physics for philosophy, one might say,

and that

we can see David as a scientist of the a priori, and his career as one long inference to the best explanation.

No! Or... not quite! With the greatest respect and admiration for Chalmers also, however resolutely scientific Lewis is in physically informed actual world analysis, and however masterly he is at managing the ontology of worlds beyond 'the' world, he leaves something crucial in his scientism behind in his modal forays. His 'fundamental

physics' for philosophy is out of tune with his ontology, and with all that can be derived by an uncompromisingly serious adoption of that ontology.

As this thesis is dedicated to the memory of Lewis, it is in another sense dedicated to the other David, too. For while I think David Lewis is curiously wrong about the 'physics' of the worlds, I think that David Chalmers is even more spectacularly, engagingly, and *usefully* mistaken. And I think I can show how.

Most of Chalmers' critics share his assumptions about possible-world ontology. You could be a modal realist, an ersatzist, an Aristotelian, or a Presbyterian for that matter: but you were entitled to and could make good use of the possible world formalism for your metaphysical investigations. And this is where I disagree with him most fundamentally. While other critics endlessly debate what I shall call the 'logistic' nuances of Chalmers position, I ignore them almost entirely. This is partly practical. It's very hard stuff! But it also reflects my interest in primary ontological matters, and my interest in induction from sparse experiential evidence, like what we seem to have in the case of just *one* world, which I seriously take to be one of many equally *real* worlds. And I shall argue for these worlds.

In Part 1 I set out to analyse various attempts to define materialism, and I find all of them deficient in some way. In the end I account for this by making a case for a benign presumption of materialism as true by definition. Almost everyone holds that materialism is at best a contingent thesis. While in Part 1 I do not examine modal assumptions deeply, I do argue there that if such a claim of contingency is to be made, we really ought to be look at how things stand at other worlds, or *would* stand at other worlds, if you favour that way of talking. It is not much use being a contingent-materialist if you can't say by what principles you would decide which worlds are in the materialist fold, which are in *another* materialist fold, and which are out in a presumed immaterialist modal wilderness. I argue for cutting the Gordian knot; gathering what support I can from Teller, Earman, and few other mavericks I push for a necessary materialism. I genuinely cannot think what I should count as a world for which it makes sense to say that materialism is false at it. Not in any principled sense, anyway. My aim in Part 1 is preliminary. There I seek to make this position plausible, not inevitable.

In Part 2 I get tougher. This part is very long, and a lot of it is given over to explicating Lewis's modal realism in my own way. It seems to me that few take Lewis at his word, despite his long advocacy and lonely persistence in the modal-realist cause. To me it seems that people just don't entertain the hypothesis for long enough to get a feel for it. Perhaps our selection of primitives is really just an appearance of deliberated choice, and most people are constitutionally unable to countenance any reduction of modality to something unmodal through and through. We don't even have the language with which to get comfortable with such a reduction. Or we are so used to our familiar gallery of philosophical types that every new approach must immediately be classified and filed away. It's a reflex, perhaps. Lewis's ontology of 'real', 'concrete', 'possible', 'world' (the language, again!) is something more radical than nearly everyone takes it to be. Many see it as radical; but they do not diagnose the radicalism accurately. It does not importantly violate Occamist parsimony: but it is a *genuine reduction*, and that is a rare thing. I do not believe, however, that the principles of worldmaking that Lewis is committed to are sustainable. I like the broad ontology, but not the ontological excess baggage. I hold that modal realism and what I call Lewisian recombination are quite separate theses, and I support the first but not the second. To argue for this position I adopt styles of argument from standard statistical inference theory – quite straightforward ones, really. I also bring in the anthropic principle (the sensible one, that is!) from cosmology, and show how it can be used to support Lewis's basic ontological thesis. But I devise a competing or complementary principle that I call the *paranthropic principle* (to lend it the gravitas that is already assured for the really rather trivial anthropic principle). The paranthropic principle weighs against the anthropic principle, and the resultant of these opposing forces is what I have chosen to call a *moderate necessitarianism* – Lewisian many-world ontology without the independent commitment to enormous 'free' variety among the worlds. In fact, I argue positively for there being severe restrictions on the types of possible worlds. Chalmers might therefore want to call me a 'type-B materialist', who believes in a distinction between 'metaphysical' and (broader) 'logical' necessity, so that logical considerations alone do not constrain how worlds might be, but physical, natural, or metaphysical considerations do too. I am *not* such a materialist. I believe in one seamless kind of necessity, which I prefer to call *broad necessity*. It is also 'logical' necessity, but I think that this name for

it is very misleading. Part 2 is probably the most radical and philosophically bold of the three parts, and if my thesis were to be judged on the work done there I would be happy enough. It is an important preliminary for Part 3, while Part 1 was a matter of simply making more plausible the final conclusions in Part 3.

In Part 3 I close in on Chalmers' dualism, and apply the work done in the first two parts to his account of modality, which I diagnose as seriously unserious ontologically. I cannot think that any amount of good husbandry in the logical domain will yield ontological fruit for mind, matter, the worlds, or anything else, so I do my best to pin Chalmers down to something more definite at the level of worlds – which he must agree is fair, though he *doesn't* make the commitment himself.

I conclude with a re-affirmation of my necessary-materialist monistic vows.

Part 1



The Materialist Background

PART 1

The Materialist Background

Metaphysics not only succeeded physics and mythology in the manner observed, and became as great a fund of superstition, but they were carried still farther, and corrupted all knowledge, as well as retarded the progress of it. Metaphysicians have not been quite agreed about the nature and object of their supposed science. Those we have last mentioned may be called and distinguished by the title, if they like it, of pneumatic philosophers, since their object is spirit and spiritual substance; how ridiculous soever it be to imagine spirit less an object of natural philosophy than body.

— Henry St John, Viscount Bolingbroke
(*Philosophical Works*, Essay 2, Vol. II, p. 79, 1754)

1.0 INTRODUCTION: THE HISTORICAL PRIMACY OF MATERIALISM

My purpose, here and in the present work as a whole, is to remedy as much as one can the lack of agreement mentioned in our epigraph; and to restore a little order to the very foundations of philosophy of mind (or *pneumatic philosophy*, in the charming idiom of a former time) as Bolingbroke's aspersions suggest that we should.

In the centuries leading up to Descartes most philosophers accepted, with at least philology squarely on their side, that the physical is the same as the natural. Bolingbroke may be read as blustering against the spread of Cartesian dualist notions that had introduced a deep confusion: human selves were created beings, subject to change and buffeted in the causal flow of things, so they were in a respectable sense *natural*; but they were not, for Descartes, *physical*. We shall see that there were others who were similarly affronted, in Bolingbroke's century, by Descartes's revolutionary solution to the problem of wresting free will – and so Christian humanity at large – out of the maw of all-engulfing mechanism. From the vantage point of the twenty-first century it may seem implausible to some that this Cartesian push was so distinct and so influential; and of course it is only approximately true – as always – that one philosopher single-handedly wrought such a revolution. But it is at least a *useful* fiction, with much truth in it.

Over the last century it has become a commonplace to diagnose and describe an

insidious Cartesian legacy, ripe for correction (we think of Gilbert Ryle, for example). Again an approximation; but again there is sufficient truth here. In a way, each of these three parts will be concerned with that legacy: Part 2 with neutralising a pervasive and generally Cartesian obsession with 'logical' foundations for modal inquiry and modal metaphysics; and Part 3 with exposing the error of a clearly identifiable contemporary brand of neo-Cartesian dualism – a property-based, not a substance-based, dualism of the material and the mental (or, very accurately, of the functional and the phenomenal).

Here in Part 1, though, we examine critically the Cartesian schism in the causal order of the natural world, which is to say simply *the world*; and this is an apposite prelude to our eventual main theme of modal thinking and its bearing on the ontology of consciousness. In this part I argue to the conclusion that all that is natural is best thought to be physical, and that this is best taken as necessarily the case. This is not an exercise in history of ideas, but we shall be exploring historically where this will provide explanations for current points of view or salutary examples of errors that we might want to be wary of. I do not claim finality for my findings in this preliminary study: I seek only to make a strong case for my preferred way of categorising things and properties.

1.1 DEFINING MATERIALISM – LEWIS

1.1.1 *Lewis as a principal theorist of contemporary materialism*

The late David Lewis was arguably, and perhaps paradoxically, one inclined to 'imagine spirit less an object of natural philosophy than body', in Bolingbroke's terms. He held that materialism is true, but only contingently so; so there are possible worlds at which there are spirits that are not in any sense physical. Are spirits, from the point of view of such a world, *natural*?

If we hold with Lewis that all properties are either fundamental or supervenient on fundamental properties, we will probably want, as he does, a definition of materialism (or equivalently, let us say, of physicalism) that brings into play patterns of coinstantiation of fundamental physical properties; but it is not yet settled how all-embracing the definition we arrive at will be. To derive it we have first to establish two intermediate definitions: of *fundamental properties* and of *physical properties*. The discussion following in this section and some way beyond draws on Lewis (1983a, 1986a, and 1994), since it is a respectable opinion that his are among the most formidable, elaborated, and committed presentations of materialism in the recent literature. Mainly for ease of exposition, I shall use expressions appropriate to the assumption that Lewis's modal realism is right, but nothing should hang on that decision at this stage. In Part 2 we have much more to say about the utility and plausibility of modal realism.

Some of the discussion that follows in this section, and throughout Part 1, is couched in terms of oppositions between the mental and the physical. But this is not essential to our immediate task, which is to examine materialist claims more generally. We do not address the mental as our main focus until Part 3, where we are in a position to apply the results we are aiming at here and in Part 2.

1.1.2 *Properties: fundamental and physical*

Fundamental properties

Here Lewis's usage seems to me clear and straightforward enough, so let me just collect the points of his terminology relevant to our purposes. *Perfectly natural* properties

(which I'll normally call PNPs, and which Lewis also calls *sparse* or *fundamental* properties) are those that 'carve at the joints' (1986a, p. 60), as opposed to the *abundant* properties, which are more or less gross and gruesome, and accordingly less or more 'natural' for Lewis. Sometimes Lewis singles out just these derivative, relatively unnatural properties as the abundant properties (as I shall), but sometimes he characterises *all* properties as abundant, with the PNPs constituting a very small minority of the abundant properties (*loc. cit.*). The abundant properties all supervene on PNPs; every set defines (or is identical to) some unique property, so properties are 'as abundant as the sets themselves' (*loc. cit.*). If we are well disposed toward talk of universals, we can say that for each universal there is exactly one PNP: 'Most simply, we could call a property *perfectly* natural if its members are all and only those things that share some one universal' (1983a, p. 347). *Things*, we may say on Lewis's behalf, are whatever can instantiate properties. Every thing must instantiate some PNPs (otherwise it could have no properties at all), and there are just enough PNPs to 'characterise things completely and without redundancy' (1986a, p. 60). Worlds are themselves to be counted among the things. PNPs are perfectly natural *simpliciter*, not with reference to a world (1986a, p. 61, note 44; I should point out that I am using the term 'world' in the Lewisian sense, to be discussed at length in Part 2).

All that has been suggested so far concerning properties can be applied with minor adjustments to the classification of relations as well – as attributes of pairs, triples, or generally n-tuples of things. Two things are *duplicates (simpliciter)* iff '(1) they have exactly the same [PNPs], and (2) their parts can be put into correspondence in such a way that corresponding parts have exactly the same [PNPs], and stand in the same perfectly natural relations' (1986a, pp. 61-62; Lewis adds that the second condition may not strictly be needed). Every PNP is an intrinsic property, but not *vice versa* (1983a, p. 357); and correspondingly every perfectly natural relation is *internal*, which is to say that it is supervenient on the intrinsic properties of its relata (1986a, p. 62). (In what follows I shall mostly discuss only properties, taking it as understood that much the same could be said concerning relations.) Properties *alien* to a world (and normally the actual world is assumed) are properties neither instantiated at that world, nor supervenient upon only those PNPs that are instantiated at that world (1983a, p. 364; there might be some doubt about this second clause as an accurate report of Lewis's

intentions, but I find it more straightforward and perspicuous than Lewis's own version).

Some unexplicated notion of supervenience has been assumed in the above without, I trust, leaving indeterminate anything that is crucial. Supervenience itself will be touched on throughout this part, and it will come under closer scrutiny in Part 3. Similarly, the theoretical explication of alethic modality that is proper before discussing supervenience theses is postponed till Part 2.

Throughout this work I shall say little of my own about the nature of properties *simpliciter*, since I am happy to endorse Lewis's account; at least I find that it is sufficiently clear and articulated for our purposes.

Physical properties

It ought to be remarked that some of Lewis's usages are not perfectly precise or consistent; but the above definitions, while they take some searching out and regimenting, do fall into good order readily enough. The same cannot be said for the definitions surrounding the notion of *physical* properties. I propose to trace a path through Lewis's observations leading to his final definition of materialism, in order to show what I consider to be infelicitous in that definition and in its presuppositions, before considering alternatives, including my own.

A typical Lewisian account of the connexions between the physical and the fundamental is as follows:

It is a task of physics to provide an inventory of all the fundamental properties and relations that occur in the world. (That's because it is also a task of physics to discover the fundamental laws of nature, and only the fundamental properties and relations may appear in the fundamental laws; see Lewis, 1983[...], pp. 365-70). We have no a priori guarantee of it, but we may reasonably think that present-day physics goes a long way toward a complete inventory. Remember that the physical nature of ordinary matter under mild conditions is very well understood [...]. And we may reasonably hope that future physics can finish the job in the same distinctive style. We may think, for instance, that mass and charge are among the fundamental properties; and that whatever fundamental properties remain as yet undiscovered are likewise instantiated by very small things that come in very large classes of exact duplicates. We may further think that the very same fundamental properties and relations, governed by the very same laws, occur in the living and the dead parts of the world, and in the sentient and the insentient parts, and in the clever and the stupid parts. In short, if we optimistically extrapolate the triumph of physics hitherto, we may provisionally accept that all fundamental properties and relations that actually occur are physical. This is the thesis of materialism. (1994,

pp. 412-3; my own added emphases are marked by underlining. This will be my practice throughout; see Conventions and Abbreviations, in the prefatory material.)

This looks on the face of it to be good common sense (at least it might be accepted as such among all *level-headed* philosophers of mind!). But something looking like a circularity might be detected in the excerpt which spoils it as grist for a definition. 'It is a task of physics to provide an inventory of all the fundamental properties and relations that occur in the world,' so if our appeal is to a completed physics – whether future, or merely entertained as an ideal – we might as well say that all natural properties are physical and be done with it, rendering the term 'physical' otiose. At least we might simply *declare* that all *actually instantiated* PNPs are physical properties. Some of us would be happy enough with at least this, but it is clear from much that Lewis has written that he is far from ready to accept even this restricted claim. Consider this from the earlier work he cites in the above extract:

That suggests that Materialism is, at least in part, the thesis that there are no natural properties instantiated at our world except those recognised by [presumably idealised] physics. That is better, but I think still not right. Couldn't there be a [perfectly] natural property *X* (in the nature of the case it's hard to find an example!) which is shared by the physical brains in worlds like ours and the immaterial spirits that inhabit other worlds? Or by thisworldly quarks and certain otherworldly particles that cannot exist under our physics? Physics could quite properly make no mention of a natural property of this sort. It is enough to recognise the special case applicable to our world, *X-cum*-physicality, brainhood or quarkhood as it might be. Then if by physical properties we mean those properties that are mentioned in the language of physics, a Materialist ought not to hold that all natural properties instantiated in our world are physical properties. (1983, pp. 361-2; my interpolation of 'perfectly' in the fourth line is justified by the statement I have underlined – there is only ever, in the Lewisian context, a question of physics giving a complete inventory of perfectly natural properties.)

This is hard to fathom. While the context here is an argument against a certain characterisation of physical properties as just those inventoried by physics, just what sort of natural property is *X* supposed to be, in this argument? Nothing so causally superfluous as a *mathematical* property, surely, or Lewis would not have found it hard to find an example. Now, take his rhetorical question regarding brains and spirits, remembering that what counts 'intuitively' as materialism is presumed true in all of this: *X*, not belonging within the purview of physics (however that is now to be delimited), presumably must make no difference to the history, in any terms *admitted* to be

physical, of the brain instantiating it. (If it did we would be *way* off track in defining materialism!) Does its instantiation make a difference to the history of the otherworldly spirit? If so, we can perhaps understand why *X* should be attributed to that spirit (however it is that we are to make sense of the spirit's *immateriality*), and we can perhaps understand its status as a PNP; but by what criterion would we ever attribute it to a physical *brain*? *X* is, let it be stressed, supposed to be a perfectly *natural* property, and these are crucially 'the ones whose sharing makes for resemblance, and the ones relevant to causal powers' (1983a, p. 347). But brains (physical by stipulation, in a stipulatedly materialist world), would resemble each other, and be related causally to other things in the world, entirely indifferently to instantiation or non-instantiation of *X*, which must be neither physical nor supervenient on physical properties (since PNPs are at base level, and supervene on no other properties).

Consider the second question raised in the extract, about the quark and some particle alien to our physics. *X* must not be a property whose instantiation makes causal differences in our world. (If it did make causal differences, it would clearly have to be mentioned by physics, and it is hard to see how a natural property of a quark, making a causal difference, should not on that account be deemed a physical property.) But then what would ever determine us to ascribe *X*, otherwise unnamed, and unknown to physics, to a quark or to anything at all? Physics, according to Lewis, 'aspires to give an inventory of natural properties – not a complete inventory, perhaps, but a complete enough inventory to account for duplication among actual things' (1983a, pp. 356-7). What causal powers could *X* (a 'natural' property of a 'physical' thing) have, then, compatible with the assertion that physics need not mention it in its inventory? None, it would seem; at least, none in a world that most of us would feel easy classifying as a *materialist* world; nor could the aspirations of physics 'to account for duplication among actual things' be met, if there are *X*-like properties.

What emerges from this analysis is that Lewis is not content to take PNPs – whether all of them, those actually instantiated, or those coinstantiable with actually instantiated PNPs according to some ensemble of actual 'laws of nature' – as essentially fundamental physical properties; yet he has not succeeded in delimiting the scope of the physical in some other way. Fundamental physical properties are *roughly* those

appealed to in *something like* the physics of our time: it is not clear whether Lewis affirms this by fiat or by faith. If by fiat, it is by a stipulation that ultimately leaves the very term 'physical' idling; if by faith, what *exactly* does Lewis have faith in? That current physics bids fair to discover the true *physical* PNPs? But what sorts of properties are *they*?

(In the preceding analysis of properties like *X* I have assumed that Lewis would stick by his introduction of them, and retain talk of them in his finished account of the physical. It may be preferable to take it as an interesting speculative turn, in only one of the relevant works, and not to burden Lewis with it. I nevertheless believe that it at least illustrates how even the best theorists can lapse momentarily when confronted with the problem of defining materialism – and this observation is most relevant to our present task. In what follows I refer again to *X*, but I hope that the proviso I suggest here will be borne in mind.)

'It is a task of physics to provide an inventory of all the fundamental properties and relations that occur in the world,' writes Lewis, as we have seen. We have also seen that he adds riders to the effect that physics may not be able to succeed in this task, and should not be expected to do so. Quine has another approach:

If the physicist suspected there was any event that did not consist in a redistribution of the elementary states allowed for by his physical theory, he would seek a way of supplementing his theory. Full coverage in this sense is the very business of physics, and only of physics. (1981, p. 99)

In Lewis's terms, this amounts to the assertion that no properties are ever discoverable that we should not expect to find among the PNPs in the inventory aimed at by physics; *all* PNPs instantiated at our world belong in that inventory. Put simply, all our PNPs are physical PNPs, for Quine. If this amounts to a redefinition of *physics* that renders the term 'physical' redundant, that is a consequence to be lived with. This is a benign redundancy, because for Quine there is no question of there being *non-physical* worlds, as there is for Lewis. What remains of substance and use, using Quine's approach, is a clearer vision of physical science for the benefit of its practitioners and critics. I am strongly sympathetic to Quine's view, and shall return to a deeper consideration of it later in Part 1.

For now, though, let us resume the story of Lewis's moves toward a definition of

materialism, working with the assumption that, for him, *some* world's fundamental physical properties constitute some proper subclass of that world's instantiated PNPs, and including some further observations regarding allegedly discoverable *X*-like properties.

1.1.3 *Lewis's definition of materialism*

I consider five successive formulations of materialism (from Lewis, 1983a; all references in this subsection that are undated are to this piece), the last of which Lewis accepts.

M1: Any two possible worlds that are exactly alike in all respects recognised by physics are qualitative duplicates. (p. 362)

M1 is given in terms of whole worlds 'in order to bypass such questions as whether mental life is to some extent extrinsic to the subject' (p. 360). Lewis rejects it because it makes of materialism a necessary truth, which he maintains is not what materialists want (certainly not Lewis himself). 'For instance,' he observes, 'our Materialistic world differs from a nonmaterialistic world that is physically just like ours but that contains physically epiphenomenal spirits' (p. 362). Lewis is perhaps using 'physically epiphenomenal' non-standardly here, since that phrase might standardly import a kind of supervenience on the physical. (See later discussion of meanings of 'epiphenomenal'; compare also Lewis's commitment to nugatory and unknowable 'epiphenomenalons' at worlds 'physically' indistinguishable from our world, but rationally believed *absent* from our world – a commitment that I argue is unsupportable, in Part 2.) He just means that those properties are unconnected to, and especially causally unrelated to, the physical properties instantiated at the world in question. We should note that if there could be properties like *X*, their instantiation might also render M1 a poor characterisation of materialism.

M2: There is no difference, *a fortiori* no mental difference, without some nonmental difference. Any two worlds alike in all nonmental respects are duplicates, and in particular do not differ in respect of the mental lives of their inhabitants. (p. 362)

Lewis faults M2 for two reasons (apart from its not respecting materialism's contingency): it never identifies the non-mental with the physical; and it denies that all

physical properties might also be mental. This is the alleged possibility of 'Panpsychistic materialism', which Lewis finds implausible but claims he cannot rule out, and which would have the consequence that there are simply no non-mental properties at some worlds (pp. 362-3). Perhaps a variant of M2 less vulnerable to these objections (though it may invite others) would be:

- M2': There is no difference, *a fortiori* no mental difference, without some difference characterisable entirely in terms of physical properties (all of which are non-mental). Any two worlds without differences so characterisable are duplicates, and in particular do not differ in respect of the mental lives of their inhabitants. (p. 362)

Moving to the next of Lewis's five candidate formulations of materialism:

- M3: No two Materialistic worlds differ without differing physically; any two Materialistic worlds that are exactly alike physically are duplicates. (p. 363)

Lewis rejects M3 as appealing to the definiendum in the definiens:

All we learn is that Materialistic worlds comprise [*sic*] a class within which there is no difference without physical difference. But there are many such classes. In fact any world, however spirit-ridden, belongs to such a class. (*loc. cit.*)

The last sentence of this may seem odd, but it is of course quite right (at least using Lewis's well-known principles for the construction of worlds – see Part 2 – and assuming with him that there *are* spirit-ridden worlds!). We simply consider the class whose only two elements are the spirit-ridden world in question and a world identical to it except for an extra instantiation of some physical property.

- M4: Among worlds that conform to the actual laws of nature, no two differ without differing physically; any two such worlds that are exactly alike physically are duplicates. (p. 363)

M4 is rejected also, first because it could hold 'at a world where materialism is false but where spiritual phenomena are correlated with physical phenomena by strict laws' (p. 363). Presumably those strict laws must be additional to the 'actual laws of nature' mentioned in M4 – a presumption that rules out there being a clause like 'that's all' among the actual laws. I take it that the qualification 'strict' accords with Davidson's gloss: 'A "strict" law is one which makes no use of open-ended escape clauses such as "other things being equal"' (Davidson, 1994, p. 231). If so, perhaps Lewis does not say

enough about this correlation. Surely it needs to be a correlation so tight as to amount to a kind of supervenience of *all* spiritual phenomena at a world on the physical phenomena of that world. (Let us also remark here what amounts to Lewis's automatic exclusion of spirit from the sphere of the natural, adverted to at the beginning of this section.) His second reason for rejecting M4 is that it 'fails to hold at a Materialistic, spirit-free world if the laws of that world do not preclude the existence of epiphenomenal spirits. Our world might be such a world, a world where spirits are absent but not outlawed' (p. 363). Let it be noted that Lewis, in the sources we look at here, is uncommitted on whether a world's pattern of PNPs determines that world's laws. He asks: 'Could two worlds differ in their laws without differing, somehow, somewhere, in local qualitative character?' (1986a, p. 14), and intimates that he discusses it elsewhere, inconclusively. We do in fact know Lewis to be an advocate of a 'Humean supervenience' account of laws of nature; but these matters are on the agenda for Part 2 and Part 3.

With that I propose to leave consideration of M4, also without resolution. It is a problematic formulation, but it will not occupy us unduly. What we should retain from Lewis's discussion of it is that perfect-duplicate worlds may, for all Lewis can tell us, differ in their laws, and this insight may be important later on. Now for the formulation that Lewis does find acceptable:

- M5: Among worlds where no natural properties alien to ours are instantiated, no two differ without differing physically; any two such worlds that are exactly alike physically are duplicates. (p. 364)

Lewis accepts M5 as satisfactorily capturing materialism as 'a restricted and contingent supervenience thesis' (p. 364). In introducing it he observes that '[if] our world is Materialistic, it is safe to say that some of the natural properties instantiated in any nonmaterialistic world are properties alien to our world' (*loc. cit.*). That does seem fair, but there are problems of a converse sort with M5. If we allow, as Lewis does, that there could be non-physical PNPs like *X* instantiated at a materialist world (as discussed above), and that for all we know some *are* instantiated in our world, then surely for all we know there are worlds that are physical duplicates of ours, and that have no alien properties instantiated, but that lack properties instantiated at our world – some properties like *X*. In that case, by Lewis's account there are, for all we know, worlds

without spirits (as our materialist world is), with no properties alien to our world, but that cannot be brought into the materialist fold. This seems quite contrary to any usual understanding of the notion of materialism, and to be considerably removed from his own rough and summary definition cited earlier: 'that all fundamental properties and relations that actually occur are physical. This is the thesis of materialism.' Furthermore, the worlds just described should on the face of it have a *stronger* claim to being materialist than our world has. For one thing, inhabitants of such a world are justified by Lewis's reasoning in discounting *our* world as a materialist world, though they count their own world as materialist, and both theirs and ours are equally without spooks!

Lewis is certainly ready to rule that some other worlds have their own physics (different from our physics); and, as we have seen, that some worlds are non-materialistic or unphysicalistic:

Physics has its short list of 'fundamental physical properties': the charges and masses of particles, [etc.], and maybe some more yet to be discovered. In other worlds where physics is different, there will be instances of different fundamental physical properties, alien to this world. [...] And in unphysicalistic worlds, the distribution of fundamental physical properties won't give a complete qualitative characterisation of things, because some of the 'fundamental' properties of things will not be in any sense physical. (1986a, p. 60)

(Another note in passing concerning *X*-like properties: according to the last sentence of this quote, if such properties are instantiated at our world then our world must be unphysicalistic, since such properties are by definition non-physical PNPs. But this surely can't be a consequence that Lewis intends.)

Since Lewis's most detailed definition of the physical even at our world is dependent upon the content of contemporary actual physics, rather than upon some characteristic shared by a privileged grouping of PNPs independently of the classifications we happen to have discovered or happen to favour, we have little guidance concerning the nature and scope of any *alien* physics. Is it definitive of physical properties (of any world) that they have a spatial aspect, or are only ever instantiated at spatial locations? Not if *X*-like properties are to be taken seriously (since quarks and brains might have them, and *they* are spatially located), or if in some worlds there are spirits (paradigmatically unphysical, for Lewis) that are located in spacetime.

(There is circumstantial evidence that Lewis allows among his possibilities spirits with spatial properties, along with the temporal properties he thinks they would have to possess: 'Our world might [taking this as an epistemic 'might'] be such a world, a world where spirits are absent but not outlawed,' as quoted above. Lewis is committed to our world being one at which space and time are constituted by different dimensions of a single spacetime manifold, so arguably to be consistent he must not allow that our laws permit beings which exist only in space, or only in time. That would be like a point in 3-space having some *x*- and some *y*-coordinate, but no *z*-coordinate. So if our laws allow for temporal spirits but not things lacking either temporal or spatial qualities, any spirits they allow for must have spatial properties. There is, however, strong evidence that Lewis does indeed believe space and time to be more separable, though they are not separable by the laws of our world. I shall return to this and related themes in Part 2.)

I conclude that materialism is not well defined by Lewis, mainly because of his failure to provide an independent definition of physical PNPs: one that is not hostage to the state of physics as we have it now, or can foresee ourselves having it, and that will enable us to decide which worlds are physical but with different physics from our own. There will be more to say about Lewis's views later in Part 1, and much more in Part 2.

1.2 DEFINING MATERIALISM – JACKSON

1.2.1 Basic terminology and assumptions

This section is dedicated to Frank Jackson's treatment of the foundations of materialism (as we shall persist in calling it, though he normally uses 'physicalism', with a slight differentiation from 'materialism', irrelevant to our present concerns).

Lewis has provided us with valuable concepts and terminology with which to standardise, clarify, and assess the claims of other theorists, and the detail we have gone into with his account should stand us in good stead as we proceed, including with Jackson. My discussion here draws on: Jackson (1994; referenced by some as '1993'); Braddon-Mitchell and Jackson (1996; for brevity I shall not mention Jackson's co-author in what follows; in any case the position in 1996 is very closely modelled on Jackson's own in 1994); and, only where the account varies in a way relevant to our concerns, Jackson (1998). I intend eventually to advance the discussion well beyond the merely definitional, partly because of the obvious intrinsic importance of doing so, and partly because such a move seems to be the best way of showing what I find to be certain quirks in the definitions examined. In the section after this one, a survey of other accounts (followed by my own) will draw on the work done in exploring both Lewis's and Jackson's important and influential treatments.

It should perhaps be noted that Jackson has come to be, if not a card-carrying materialist, at least a fellow traveller, after a shift away from his widely influential anti-physicalist and anti-functional pieces of the 1980s (such as his 1982 piece):

Once upon a time I was convinced that any adequate account of colour experiences required reference to qualia understood as properties over and above those that appear in the physicalists' story about our world. Nowadays I am much more sympathetic to physicalism. (Jackson, 1998, p. 101)

Like Lewis, Jackson classifies properties as physical if they are appealed to in the physical sciences, or are straightforwardly derivable from such properties:

Our official definition of a physical property (item, relation, etc.) is that physical properties are ones that appear in the physical sciences, broadly construed, or anyway are of the same general kind as required to give a complete account of inanimate nature. Materialism is then the doctrine that psychology supervenes on the physical so defined. But it is common, and no harm is done thereby, to allow an extended use of 'physical' to mean physical either according to the official

definition (or some near relative) or that which supervenes on what is physical according to the official definition. (1996, pp. 26-27)

The characterisation of 'physical property' is essentially the same in Jackson (1998), though it is more elaborately presented, and there is 'a more explicit emphasis on the microphysical foundations of physics as a whole (and so of physicalism itself):

[...] the clearly non-trivial claim that the kinds of properties and relations that are enough to account for everything below a certain size, and in particular below the size needed to have semantic or psychological properties, are, in suitable combinations, enough to account for everything, or anyway everything semantic and psychological. (1998, p. 8)

What is important for us is that physicalists have three reasonable things to say by way of explaining what they mean by physical properties and relations – they are those that we need to handle the non-sentient, they are broadly akin to those that appear in current physical science, they are those we need to handle the relatively small [...]. (*loc. cit.*)

So far Lewis and Jackson appear to be well enough in accord. Let us note, however, that though Jackson has it that we should assume the official definition when making materialist claims that certain properties supervene on physical properties (on pain of circularity), 'officially' physical properties do not automatically count as PNPs as defined by Lewis, since there are many properties appearing in 'the physical sciences, broadly construed' that are clearly not PNPs, but supervenient on them (*metallic* is an example; Lewis, 1983a, p. 347). Jackson's preliminary way of situating materialism as a metaphysical thesis is given in such observations as these:

Physicalism is the very opposite of a 'big list' metaphysics. It is highly discriminatory, operating in terms of a small set of favoured properties and relations, typically dubbed 'physical'; and it claims that a complete story, or anyway a complete story of everything contingent, about our world can in principle be told in terms of these properties and relations alone. It is miserly in its basic resources while being as bold as can be in what it claims. (1994, p. 26)

Physicalists variously express their central contention as that the world is entirely physical; as that when you have said all there is to say about physical properties and relations you have said all there is to say about everything, or anyway everything contingent including psychology; that the world is nothing but or nothing over and above the physical world; that a full inventory of the instantiated physical properties and relations would be a full inventory *simpliciter*; and so forth. (*ibid.*, p. 27)

As we have seen, a point of divergence between Lewis and Jackson is marked by

Jackson's suggestion, unmatched by Lewis, that the precise selection of properties to be considered physical is not of prime importance for a broad critique of materialism: 'What is important here is that there is a favoured list, not how a property or relation gets to be on that favoured list' (1994, p. 26). This suggestion of initial indifference is of consequence for my discussion of Jackson's analysis; so is his next suggestion: 'What will be important is the notion of a complete story,' since 'it is the physicalists' claim to have a complete story about the nature of our world which commits them to our world having a psychological nature if and only if that nature is entailed by the world's physical nature' (1994, p. 27). This is Jackson's *entry by entailment thesis*, for which he argues at length.

His progress towards a more satisfactory definition of materialism is via three candidate characterisations (see 1994, pp. 27-28), which are perhaps best thought of as supervenience claims:

- (A): Any two possible worlds that are physical duplicates (physical property and relation for physical property and relation identical) are duplicates *simpliciter*.
- (B): Any world that is a physical duplicate of our (the actual) world is identical *simpliciter* with our world.
- (C): Any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world.

(Let us note in passing that the parenthetic qualification in (A) is to be understood as having to do with *patterns of coinstantiation* of the properties and relations. I shall sometimes use the same abridgement myself, where the sense is clear enough – which cannot always be assumed, as we shall see in considering other theorists.) We need not dwell here on (A) or (B), which Jackson rejects on the ground that neither gives a characterisation of materialism as a contingent claim about our world. Regarding (B) he observes: 'Physicalists are typically happy to grant that there is a possible world physically exactly like ours but which contains as an addition a lot of mental life sustained in non-material stuff' (1994, p. 28). (C) is another matter; Jackson thinks it is the best characterisation available of materialism, but he finds that, taken by itself, it fails to have materialism discharge its obligation to provide a complete story of the actual world, since the notion of minimality in (C) is itself closely dependent on some

unexplicated notion of completeness (1994, p. 29).

1.2.2 A critique of minimality

Minimality is indeed at the heart of the matter. Here are two of Jackson's explanations of the notion:

What is a minimum physical duplicate? Think of a recipe for making scones. It tells you what to do, but not what *not* to do. It tells you to add butter to the flour but does not tell you not to add dirt to the flour. Why doesn't it? Part of the reason is at no-one would think to add dirt unless explicitly told to. But part of the reason is logical. It is impossible to list all the things not to do. There are indefinitely many of them: don't add bat's wings, don't add sea water; don't add... Of necessity the writers of recipes rely on an intuitive understanding of an implicitly included 'stop' clause in their recipes. A minimal physical duplicate of our world is what you would get if you – or God, as it is sometimes put – used the physical nature of our world (including of course its physical laws) as a recipe for making a world. (1994, p. 28)

In short, a minimum physical duplicate of our world contains nothing more that it *must* in order to be a physical duplicate. (1996, p. 24)

My own first objection to this is something like the reverse of Jackson's objection to (A), which is that 'our world is nowhere mentioned in (A)' (1994, p. 28). I maintain that in order not to prejudge things we first need to define *minimum physical duplicate* (henceforth M Φ D) independently of the standing of materialism and of the nature and constitution of the actual world; and it is worth our while also to take a step or two back in respect of *natures* and *duplicates*. My aim here is to develop rigorously articulated definitions that Jackson would be happy to substitute for his own. Very generally, then, in terms of some arbitrarily chosen class of properties Γ , and understanding a *nature* in the present context to be some rather general property of a whole world, we can define Γ *nature* like this:

Γ nature: The Γ nature of a world is the pattern of coinstantiations of properties it must have solely as a consequence of its patterns of coinstantiations of Γ properties.

For the moment we shall assume that this singles out for any given world a unique pattern of coinstantiations, which is open to question. Next, let us use Γ *duplicate* (Γ D) to signify a duplicate with respect to Γ :

Γ D: A world w_2 is a Γ D of a world w_1 iff w_1 and w_2 share the same Γ nature.

We can now define *minimal Γ duplicate* (MGD):

MGD: A world w_2 is an MGD of a world w_1 iff w_2 has among its patterns of coinstantiations of properties all and only those it must have to be a Γ D of w_1 , and no pattern of coinstantiations absent from w_1 .

Provisionally taking it then, as Jackson does, that the class of physical properties may be thought of as somewhat arbitrarily constituted for the purposes of broad analysis, we can use this unexceptionably general definition to get a definition of *minimal physical duplicate*:

M Φ D: A world w_2 is an M Φ D of a world w_1 iff w_2 has among its patterns of coinstantiations of properties all and only those it must have to be a physical duplicate of w_1 , and no pattern of coinstantiations absent from w_1 .

One novel feature of this definition is its clear requirement that *all* of w_2 's instantiated properties must also be instantiated properties at w_1 , which is not overtly obligatory by Jackson's version of minimality, though it certainly seems to be intended. If the choice of properties to count as physical were restricted to a selection from Lewisian PNPs the requirement would be redundant, since then 'officially' physical properties would be ruled not to have multiple possible realisations in some supervenience base: they would be truly fundamental.

Some consequences of the definition of M Φ D may serve to make its significance clearer. If w_1 has some physical properties and also some non-physical properties (these last being, that is, properties that are neither physical nor *supervenient* on any physical properties), and if w_2 is an M Φ D of w_1 , then: w_1 *cannot* be an M Φ D of w_2 ; w_1 *cannot* be an M Φ D of itself; and w_2 *must* be an M Φ D of itself.

Using the MGD-style of definition we can define other sorts of minimal duplicates too, if we like. In particular, *minimal psychological duplicate* (M Ψ D) could be defined like this:

M Ψ D: A world w_2 is an M Ψ D of a world w_1 iff w_2 has among its patterns of coinstantiations of properties all and only those it must have to be a psychological duplicate of w_1 , and no pattern of coinstantiations absent from w_1 .

This *seems* as if it should be satisfactory, provided we can get clear about which

properties we want to count as psychological. An M Ψ D of our world, for example, would be very much sparser than our world, since much at our world is in no way relevant to its psychological nature. But it will be a worry that both w_2 and w_3 could conceivably be M Ψ Ds of w_1 without being M Ψ Ds of each other. There may be several distinct ways, all with equal claims to *minimality*, of selecting from the properties of w_1 to found an M Ψ D of w_1 . Minimality seems not so obvious, now: the very same set of psychological properties may very plausibly have several possible 'realisations', founded on different selections from w_1 's non-psychological properties; so w_2 and w_3 might both be M Ψ Ds of w_1 , but not of each other – one or both having some pattern of coinstantiations the other lacks (though clearly all three would be psychological duplicates *tout court* of each other).

Since the psychological is the very matter at issue, an easy non-psychological example may be useful. Suppose that there is just one game of chess that ever happens at a certain world, which we may call w_{ch} , and that the moves constituting that game by themselves determine a *chess nature* (CD) for that world. Then we can define *minimal chess duplicate* (MCD):

MCD: A world w_2 is an MCD of a world w_1 iff w_2 has among its patterns of coinstantiations of properties all and only those it must have to be a CD of w_1 , and no pattern of coinstantiations absent from w_1 .

Now, the class of distinct MCDs of w_{ch} may be vast, since its chess nature could, we ought to think, be minimally realisable in a vast number of ways. So there would be many MCDs of w_{ch} that are not MCDs of each other.

So much the worse for M Φ Ds and MCDs, but we're not interested in those? Such insouciance will not do, though. Let it be stressed: the case warns us that a similar uncertainty affects our definition of M Φ D if what is to count as a physical property is not constrained to be a Lewisian PNP in this analysis (see Daly, 1998, p. 215, note 21, for a related point, concerning multiple alternative determinations of physical effects); yet our MGD schema had seemed to be a proper and rigorous development of Jackson's proposal. So once again, an analysis using Lewis's notion of PNPs is preferable.

1.2.3 Entailment by physical facts

Jackson argues compellingly that materialism stands or falls along with (C), with all of that formulation's imperfections. Now, with our definition of MΦD, we are properly equipped to look at Jackson's next move, which is to expand upon the idea that '(C) entails that any psychological fact about our world is entailed by the physical nature of our world' (1994, p. 31). As we have seen, Jackson's intent (at least in his consideration of materialism) is ultimately to show that 'the psychological appears in the physicalist's story about our world if and only if that story entails the psychological' (1994, p. 26):

Let Φ be the statement [...] which tells the rich, complex and detailed physical story which is true at the actual world and all the minimal physical duplicates of the actual world, and false elsewhere. Let Ψ be any true statement entirely about the psychological nature of our world: Ψ is true at our world, and every world at which Ψ is false differs in some psychological way from our world. If (C) is true, every world at which Φ is true is a duplicate *simpliciter* of our world, and so *a fortiori* a psychological duplicate of our world. But then every world at which Φ is true is a world at which Ψ is true – that is, Φ entails Ψ . (1994, pp. 31-32)

Jackson claims that this is a demonstration that '[a] putative psychological fact has a place in the physicalist's world view if and only if it is entailed by some true, purely physical statement.' I shall now argue that, though he has succeeded in demonstrating this, and so has succeeded in showing that '[any] putative psychological fact which is not so entailed must be regarded by the physicalist as either a refutation of physicalism or as *merely* putative' (1994, p. 32), this is not a consequence that should alarm materialists. In short, Jackson asserts that either:

- A: materialism is mistaken, or
- B: the complete physical story about our world entails all its psychological facts.

I am happy to accept the disjunction A or B – if defining their terms *my way!* – but I reject A, and believe that the complete physical story about our world entails all its psychological facts (and indeed, all of its facts *simpliciter*).

It is perhaps worthwhile exploring how Jackson elaborates his story concerning Φ and Ψ , since it may show something of the dangers attaching to starting with the actual world in one's deliberations concerning physicalism.

We have already seen, in considering minimal duplicates with respect to other sorts of properties (our examples used psychological and chess properties), that any well-regimented system for explicating minimum duplicates requires the 'fundamentality' of the properties in question: but Jackson does not first establish that his paradigmatically physical properties are fundamental. And there is a further procedural problem. It is reasonably easy to show that there can be a sentence described as Φ is (true at all and only w_a and its MΦDs) *only* if physicalism is true, so that the complete physical story about our world entails all its psychological facts, and that the world is an MΦD of itself. Here is one of many ways to show this:

w_a = the actual world
MΦD = minimum physical duplicate

P1: Φ is true at all and only worlds w_a and its MΦDs.
[Jackson's stipulation]

P2: There is a duplicate *simpliciter* of w_a , which we shall call w_{ad} .
[required to be permissible, by Jackson's procedure]

P3: Φ is true at w_{ad} .
[from P2]

P4: w_{ad} is an MΦD of w_a .
[from P3]

P5: w_a is an MΦD of itself.
[from P2, P4]

C: Physicalism is true. [from P5, which is a way of stating physicalism]

We might wonder about the propriety of any system purporting to show what physicalism amounts to if we can derive the *truth* of physicalism from that explicatory system. This may not be a problem if physicalism is taken as true necessarily, because then there will be no case of a world at which something like Φ is false. But if we hold, as Lewis and Jackson do, that physicalism is only contingently true, it may amount to a distracting infelicity, or worse.

If there *is* a Φ , (C) is true, so is materialism, Φ entails all the psychological facts at our world, and materialists should be happy enough. But if there is *no* Φ , materialism is simply false. The *statement* that there is a Φ entails that materialism is true, and that the

conjunction of all the physical facts at our world entails all the psychological facts at our world.

1.2.4 Possible modifications of the entry by entailment thesis

The non-materialist can simply say: 'Well, on reflection, I believe that there is no Φ ; but the materialist must be committed to there being a Φ , and she is therefore committed to the belief that the conjunction of the physical facts at our world entails all our psychological facts. It is just that entailment that I deny, and which any materialist would come to deny too, on a proper consideration of examples – those involving zombies, or qualia, for example.' The practical and epistemological trouble with this obvious response, of course, is that even if materialism were false at our world, at an M Φ D of our world (call it w_m) there are counterparts of our resolute materialists claiming that Φ_m (though like us they call it simply ' Φ ') entails the psychological facts at their world. And they are right! Their behaviour, their arguments, and what passes for their mental life are indistinguishable from those of their counterparts at our world. (Many will assert that there are phenomenal features of our experiences that they must lack; but that cannot have a bearing on the matter of their behaviour and the arguments they deploy, which will be strictly indistinguishable from ours for all discursive purposes. The parallel difficulty in relation to phenomena themselves will be considered in Part 3.) But if materialism is mistaken at our world – at their world it *isn't*! But our materialists use the same arguments as *they* do, and have the same evidence as *they* have (or at least behave and talk as if they do, remembering that sentences are 'physical objects', according to Jackson; 1994, p. 24). Our materialists should therefore not be persuaded by anyone seeking to deploy Jackson's systematisation against the truth of materialism.

Perhaps we should reconsider the *stop* clause in Jackson's formulation (call it 'STOP'), so that a 'spooky' world (call it w_s) can be made to have its equivalent of Φ (call it Φ_s). I say that Φ and indeed Φ_s , as Jackson would characterise them, cannot be 'purely physical statements', which is what we are aiming at. If Φ is true at all and only our world and its M Φ Ds, it is reasonable that it include STOP. But, I claim, the inclusion of STOP, understood as it needs to be understood, does nothing to certify Φ 's purity as a

physical statement. Turning now to Φ_s , it would have to tell the whole *physical* story of w_s , and must leave out *all* the information about all the non-physical properties. But surely one item of unavailable information is that there *are* some non-physical properties in w_s . A 'keep-going-but-with-something-non-physical' clause could hardly be construed as a purely physical clause – its very articulation uses the term 'non-physical'. But what is STOP, if not the negation of such a keep-going clause? Surely if a clause is *physical* then so is its negation: both tell about the physical state of affairs. And, equally surely, if a clause is non-physical its negation is also non-physical.

So even a reconstructed Φ_s must fail. Either Φ_s does not include STOP, in which case it is true at worlds other than w_s and its M Φ Ds, or it has one of two sorts of stop clause (not distinguished by Jackson):

STOP₁: There are no more properties than those listed already.

STOP₂: There are no more physical properties than those listed already.

If Φ_s has STOP₁, it is false at w_s . If it has STOP₂, it is true at worlds other than w_s and its M Φ Ds (since it is true at innumerable non-materialist worlds other than w_s). So there can be no Φ_s , in any version.

To summarise and conclude: The materialist view that physical facts make true all other facts about the world – as a contingent claim – is not captured perfectly by Jackson's articulation of it. *Some* arbitrarily constructed classes of facts designated as physical will not support the materialists' claim to completeness; the facts should be selected, not arbitrarily, but rather from the class of PNPs. If that is done, and if definitions are tightened and applied with due rigour, materialism – provided its PNPs are properly selected – is well articulated. I would add that it then has an excellent claim to accounting for all the world's facts.

I now move on to a consideration of other definitions of materialism, including my own. These will be all the more clearly delineated for our consideration of Lewis's and Jackson's accounts.

1.3 DEFINING MATERIALISM – OTHER APPROACHES

1.3.1 Campbell

In this section I briefly review a loosely assorted sample of other typical but often less systematic approaches to the problem of characterising materialism. Some of them amount to mere asides to other concerns, but all, I hope, contribute to the general picture that is emerging of a serious difficulty in modern philosophy of mind – a difficulty which is an impediment to progress, and which calls for a certain radical move that I shall put forward in the section following this one. I exclude for the moment any consideration of those who are in sympathy with that radical proposal, and confine attention to theorists who hold that there is a principled distinction to be made. We start with Keith Campbell.

Campbell's engaging introduction to the philosophy of mind (1984) is rich in definitional and foundational points of interest:

In calling the body "material" I mean that it shares the properties common to the most familiar objects of our environment, such as shoes, ships, and lumps of sealing wax, e.g., mass, position, volume, velocity; that it is composed entirely of the recognized material stuffs, e.g., carbon, nitrogen, oxygen, phosphorus; and that its responses to physical influences satisfy only the laws governing all matter, whether found in living organisms or in inanimate systems. (p. 16)

This is reminiscent of Jackson (see especially the second excerpt in the previous section, from Jackson, 1996) and of Lewis, in its appeal to something well-known to us all, hardly needing further elucidation. But then:

Physics changes from one year to the next, with the result that the sorts of events which conform to physical laws can change from one year to the next. What the claim to materiality amounts to changes as physics changes. (Campbell, *loc. cit.*)

Of course many others have made this and related points (e.g. Lewis: '[Materialism] was so named when the best physics of the day was the physics of matter alone. [...] But it would be pedantry to change the name on that account'; 1994, p. 413). Campbell makes of the point a principle he calls the 'relativity of materialism', which he says looms large when we turn to consider the mind: 'The claim that, for example, the mind is the same thing as the (purely material) central nervous system, inherits the relativity of all materialism' (p. 17). It may reasonably be doubted that any materialist would on

reflection really want her doctrine to drift with the current of physical theory. I maintain that materialism might be more like Jackson's version of it – that there is some favoured class that can remain *ill-defined* as far as broad analysis is concerned – but I add that we must *eventually* fix it if any progress is to be made; against both Jackson and Campbell I maintain that it makes a crucial difference to our evaluation of materialism by what means the constitution of that class is fixed. More on this in the next section.

Campbell considers another complication:

A material thing need not always behave in accord with physical law. A stone which flies through the air under a divine propulsion would violate the dynamical laws governing the motion of stones. But it would remain a material thing for all that. [...] The brain, without ceasing to be material, can act under the influence of an immaterial mind. What is necessary is that all *physical* forces acting on the object should always have their normal effect, even if this is supplemented, from time to time, by supervening forces. (p. 17; note that 'supervening' is not to be taken here in its technical philosophical sense.)

Here is a clear bias in favour of physical *forces* as somehow privileged physical items, a partiality which Campbell goes on to explain somewhat obscurely: In a brain physical events occur; some are subject to physical laws; any of the physical events that are not so subject, lying 'even beyond the latitude allowed in quantum physics', may be of two sorts: those that are outcomes of 'divine or spiritual action', or those that are undetermined. A brain in which some of the latter occur 'is not at every point subject to the laws governing material things. It would not be a *purely material object*' (p. 18). Now, what makes this ruling decidedly odd is that a brain in which some of the former sort of physical events occur (occasioned by the action of some *immaterial force*) is to be taken as a purely physical object, provided only that every physical force acting on the brain is also efficacious, in its proper, physical-law-governed way. But this seems arbitrary. Suppose a part of the brain suddenly twitches, under the influence of some spookish whim, and without the application of any physical force. This twitch is to count as a physical event in a purely material object. But suppose the spookish whim were rather to suppress (or perhaps to moderate) the efficacy of some physical force that would otherwise have caused just such a twitch. A physical force would have been thwarted, and the continued quiescence counts against the purely physical status of the brain. What makes the difference? Campbell gives no principled answer. He adds a final observation on the nature of the material, in order to broach epiphenomenalism:

If events other than physical ones occur in an object, then that object is not purely material. But if these non-physical events make no difference to what happens [*sic*: but must what happens always be physical?], if they are idle, then so far as the physical events and qualities are concerned, the object can conform entirely to physical law. (p. 18)

Campbell then wants to designate the object (with its epiphenomenal properties) 'material plus'. Two questions arise: First, why *exactly* should this object count as material plus, while a quiescent object in the presence of thwarted physical forces is relegated to 'not purely material' status, and an object that has become the plaything of poltergeists is as purely material as they come? And secondly, what is it *exactly* that makes the events in this last excerpt 'other than physical'?

There are too many such questions emerging from Campbell's treatment. Admittedly his is a work aimed at merely introducing the subject (though it has often been cited in the journal literature); the arbitrariness and counterintuitive nature of the judgements I have examined here are more starkly exhibited than many in the literature because of this pedagogic orientation. The case is truly instructive. No firm foundations are established; and no touchstones are uncovered by which materiality is reliably to be discerned.

1.3.2 Tye

Tye (1996, p. 43) provides the following summary of his candidate definitions of the term 'physical', which I have relabelled for clear reference:

- T1: Something is physical just in case it lies within the domain of physics.
- T2: Something is physical just in case it lies within the domain of physics, chemistry, molecular biology, and neurophysiology.
- T3: Something is physical just in case it is described in some theory adequate for the explanation of nonliving matter.
- T4: Something is physical just in case it is either ultimately constituted of or ultimately realized by something in the domain of microphysics.

T1 is rejected as being too restrictive. We should recall here Jackson's concern to offer an extended as well as an official definition – though T1 is even more restricted than Jackson's official definition, which appealed to 'the physical sciences, broadly

construed' (Braddon-Mitchell and Jackson: 1996, p. 26). T2 is nearer to that official definition; Tye is dissatisfied with it because there seems to be no principled way to stop – if *gene* and *neuron* are to count as physical terms, why not *crocodile* and *continent*? We would need an account of which sciences are physical and which not (p. 39). But to this the thoroughgoing materialist is entitled to reply: 'They are all physical! Just what is your problem with our Unified Science?' Tye does not say; but his T3 is offered as another way to restrict the physical (again recalling Braddon-Mitchell and Jackson). He rejects T3 also because there may be properties only ever instantiated in certain living brains, and figuring in neurophysiological laws, and surely they should count as physical. T3 is vitiated in other ways, too. The qualification *adequate* is not apt, since presumably more than one true theory is adequate for the explanation of non-living matter but does much more than that – some may even feature extensions that account for minds, angels, and the futures market. Perhaps *barely adequate* would make the required restriction. T3 (along with Jackson and Campbell's similar provisions) depends on our having a clear and independent understanding of what should count as *living*. (This term is in any case hazardously close to *minded*, which is the very term in dispute. An easy reminder of this is had by thinking of the term *animated* – *lively*, or having *anima*, which is mind.) In some contexts, this would be a pedantic worry indeed, but not in our context, in which we are to assess claims about zombies, robots, programs, and ectoplasms of one stripe or another. If more warning were needed that the scope of the living is highly problematic, we need only look to the outlandish but seemingly possible realisations, mentioned by Leslie (1996, pp. 221-2, p. 230), of systems very plausibly counting as living. There are related and previously canvassed worries about what should count for present purposes as matter: are fields, or spacetime itself, included as matter? How about whatever we would wish to count as matter at worlds with different physics from our own world, or at Lewis's proposed unphysical worlds, with their immaterial substances? By what principle are those properties to be excluded?

Tye ultimately expresses acceptance of T4, with glosses on how 'constitution' and 'realization' are to be understood, which is in terms of supervenience. He offers a 'proposal on behalf of the physicalist: that phenomenal states are both perspectively subjective and physical [in the sense of T4]'. This proposal need not detain us now, but

the matter of constitution or realisation by something in the domain of microphysics is of central interest. The problem with T4 is its reliance on a settled microphysics for the actual world. We are given no guidance as to what should count as physical for a world *simpliciter*, nor concerning whether all worlds should count as materialistic (in relation, of course, to *their* microphysical properties, however *they* are to be identified). As I have suggested earlier, this can fairly be taken as evidence that our understanding of the situation at the actual world is incomplete. If a world is non-materialistic, some of its properties must be ultimately realised in or constituted by non-microphysical properties. By what *principle* are we to decide whether a property is microphysical or not? An appeal to the actual state of physical theory for this world is of no use, as can be seen when we attempt to broaden our scope to encompass other worlds. Without knowing what makes a world – any world – materialistic we cannot be sure that we have a sound warrant for claiming that ours is.

1.3.3 *Chalmers*

David Chalmers' views on the relations between mind and the physical domain will be examined closely in Part 3. Here I shall just characterise them with respect to the extent of the physical, and the nature of the claim he takes materialism to be making, so that they can be conveniently set beside the accounts given by other theorists. (All my references to Chalmers, unless otherwise marked, are to his 1996 monograph.) He holds that the physical properties are:

...the fundamental properties that are invoked by a completed theory of physics. Perhaps these will include mass, charge, spatiotemporal position, properties characterizing the distribution of various spatiotemporal fields, [...] and so on. The precise nature of these properties is not important. If physics changes radically, the relevant class of properties may be quite different from those I mention, but the arguments will go through all the same. Such high-level properties as juiciness, lumpiness, giraffehood, and the like are excluded, even though there is a sense in which these properties are physical. In what follows, talk of physical properties is implicitly restricted to the class of fundamental physical properties unless otherwise indicated. I will sometimes speak of "microphysical" or "low-level physical" properties to be explicit. (p. 33)

There are clear echoes of Lewis here (the provisional list of physical properties; the appeal to a notion of fundamental properties, which we may plausibly take Lewis's way) and of Jackson (the distinction between official and extended definitions; the

confidence that the exact constitution of the class of physical properties does not affect the present broad analysis – it can be taken as somewhat arbitrary for now). He does differ drastically from Lewis on certain points. We have seen that Lewis holds all PNPs (and so *a fortiori* all fundamental physical properties) to be intrinsic (1983a, p. 357); Chalmers goes quite the other way. He is most inclined to accept the view that *no* physical properties are intrinsic, and indeed that '[t]here is only one class of intrinsic nonrelational property with which we have any direct familiarity, and that is the class of phenomenal properties' (p. 153).

Concerning the broader domain of physics he writes:

Of course, there is a sense in which the physics of the universe *must* entail the existence of consciousness, if one *defines* physics as the fundamental science from whose facts and laws everything else follows. This construal of physics, however, trivializes the question involved. [...] For our purposes, it is best to take physics to be the fundamental science developed to explain observations of the external world. (pp. 118-9)

We have seen that Quine, for one, takes physics to aim at unrestrictedly full coverage, which 'is the very business of physics, and only of physics' (1981, p. 99). This is as close as one could wish to defining physics so as to have it entail all the facts of the world; it is an approach I endorse, and will defend in the next section, as far from trivial in the present confused state of play with the demarcation of the physical and the non-physical. For the moment let it be noted that setting physics up as Chalmers does, with an appeal to an external world seeming to require that there must also be a distinct *internal* world, determines the course of his subsequent discussion so that a sort of dualism is bound to emerge – as indeed it does. He pre-empts the course of the dispute at least as radically as he alleges that his opponents do.

The best reason Chalmers gives for holding that physics is more restricted than the Quineans will allow is that, even when it is considered as restricted in his way, it appears to be already causally complete. Confronting the proposal that experience, if it be as fundamental as the physical (which in his opinion it is), might best be thought of as *included* in the physical domain, he observes that:

[...] physics forms a closed, consistent theory even without experience. Given the possibility of a zombie world, there is a clear sense in which experience is superfluous to physics as it is usually understood. (p. 128)

If only there were a 'usual' understanding of physics to invoke! Chalmers' zombie world will loom large in my analysis in Part 3, to elucidate these definitional and other more substantial problems more clearly. Chalmers goes on to note that the issue may be merely terminological, and that his favoured dualistic apprehension of reality may in the end be supplanted by a kind of neutral monism, but 'it remains the case that if a variety of monism is true, it cannot be a *materialist* monism. It must be something broader' (p. 129). So much for the arbitrariness with which the physical properties, and so materialism itself, may be specified in any preliminary analysis.

1.3.4 Van Fraassen

Van Fraassen's analysis (in 1996; all references are to this piece) is a refined but caustic attack on the very idea of materialism ('science's "presumptive materialism"', to use his phrase; p. 151) as a non-stipulative position of any worth or standing. He starts with a rough characterisation of it with what he calls *the Thesis*: 'matter is all there is'. He then teases out at some length how this might (or might not) have meaning:

What would count as something that is not material? Descartes said that matter is extended and mind is not; mind thinks. But if that is not a stipulative definition, it is certainly wrong. Else we would have to say that Hertz's massive point particles, if they exist, are not material. [...] It may be unfair to take Descartes as our whipping boy. But more recent putative statements of materialism do not, it seems to me, fare much better. Typically they start from some version of received scientific opinion, perhaps with some anxiety about being up to date. [...] they will say that everything is composed solely of elementary particles. If we take this seriously we shall, I wager, once more land in an untenable historical parochialism. When Newton introduced forces in addition to elementary particles, did he deny the Thesis? Forces are not composed of particles. When Huyghens's waves-in-the-aether theory defeated Newton's particle theory of light, was that a set-back for materialism? [...] When a recent article in a physics journal bore the title "Particles do not exist", was that a denial of materialism? [...] Surely not. But if materialism were really, purely and simply, some such thesis as that everything is composed of elementary particles, I could not so readily say "Surely not"! (pp. 164-5)

He discerns two main moves that '*soi-disant* materialists' have made. First, some (like UT Place and David Armstrong) have come up with specific empirical claims concerning the roots of the psychological in the physical – at least, in what counts as physical in current science. Van Fraassen considers three preliminary questions to be put to these theorists. The third (pp. 166-7) is most relevant to our concerns. It may be paraphrased like this: What if on further investigation the empirical claim is not borne

out? Would that be the end of materialism? It seems unlikely that the proposers would accept this. But if they did not, in what sense were they right to characterise their materialism as empirical?

The second move materialists have made is to express a conviction that the Thesis will be 'compatible with science, *whatever science comes up with*' (p. 167), though they will *claim* that they would surrender their materialism if it turned out that science was inadequate to explain some phenomenon. But, says van Fraassen:

holding the Thesis, they make the bold conjecture that this will never happen. That *what* would never happen? If that question cannot be answered with a precise and independent account of what material factors are, there is still one option. That is to nail a completeness claim to science, or to a specific science such as physics. (p. 167)

The example given is Jack Smart, but we may be put in mind of some of the theorists examined earlier – Lewis in particular.

Van Fraassen goes on to consider the more resilient and less definable '*spirit of materialism*', which endures despite changes in physical theory and fashion, and which promises to reveal to the philosophical investigator what is really at the heart of materialist claims. He gives the example of those we would want to call materialists welcoming gravity, when it was introduced, and greeting electromagnetic fields with equanimity when they were first mooted. How did they *know* that the new explanatory means were still physical means, when they seemed categorically different from what went before? Concerning Newtonian forces as causes of change of momentum, van Fraassen points out that:

it could be added consistently that these causes are immaterial, spiritual – even mental, if Mind does not have to be someone's mind. If instead the forces are said to be material just like the extended bodies so classified before, the materialist must seemingly have some rather mysterious kind of knowledge: a *knowledge-that* the newly introduced entities have the *je ne sais quoi* which makes for materiality. (p. 169)

Van Fraassen's diagnosis, in view of this resilience, is that, rather than with any specific empirical claim, materialism is to be identified with a cluster of attitudes, including a tendency to accept the ontology suggested by the prevailing science, and a readiness to accord science a completeness, even though that is called into serious question by its record of compromised successes and catastrophic changes of direction. These

persistent attitudes explain the ease with which materialists are able to adapt to changes in the sciences. Despite assuring us that 'this does not reflect badly on materialism; on the contrary, it gives materialism its due,' van Fraassen draws the inference that only 'the confusion of *theses held* with *attitudes expressed*,' which yields what he calls in picturesque Hegelian parlance *false consciousness*, can account for the conviction that science requires presumptive materialism (p. 170).

If materialism is not to be the hallmark of properly conducted and systematised science, what is? The coherence and distinctness of naturalism itself is brought into question:

Even to diagnose what it is to naturalize something is far from easy. To identify what naturalism is, apart from something praiseworthy, I have found nigh-impossible. [...] I venture to assert: we see here too a position that only purports to be a factual thesis. Most likely it cannot be identified with any factual thesis at all, but derives all its strength and support from attitudes that engender affinity with certain theses at each historical stage. (p. 172; cf. Kim Sterelny, 1990, p. xi: '“Naturalism” is a term in philosophy so vague that it is in danger of becoming merely honorific.' But Sterelny does allow the term at least one useable sense, as we shall see.)

I find van Fraassen's trenchant and heretical critique of materialism most refreshing. It may most readily be taken as an attack on those philosophers who say, for all that they know not *what* they say, that they are good materialists. But it ought to be clear that van Fraassen's piquant philippic against materialism will discomfit also philosophers who seek to distinguish themselves as *non-materialist*, like Chalmers, while claiming at least to know *what* they are not. If it means little to claim to be a materialist, it means just as little to claim not to be one. And if van Fraassen is right about the vacuity even of naturalism, most of the philosophers examined so far have a serious job of revising to do. Again, Chalmers can be singled out, since he is vociferous in rejecting materialism and espousing a view he calls 'naturalistic dualism'. The burden of Part 3 will be to show that this is a radically incoherent conception.

But now it is time to distinguish myself from the philosophers I have surveyed, and reveal my own hand. I have hopes of achieving some show of heresy – perhaps even a brighter taint of it than van Fraassen has managed for himself.

1.4 DEFINING MATERIALISM – A RADICAL PROPOSAL

1.4.1 *Physics before Descartes – a useful ancient myth*

What stands out in most of the definitions reviewed so far is their lack of system and their stipulativeness, keyed to a generally uncritical acceptance of prevailing pre-philosophical notions of the physical and the non-physical. Recall the parenthetical remark of Lewis's noted earlier, to the effect that materialism was, to be pedantic, ill-named by our modern lights, but the name has stuck despite the contents changing and threatening to continue to change. Van Fraassen uses this as one example to motivate his exasperated question:

If the "physicalist" or "naturalist" part of this philosophical position is not merely the *desire* or *commitment* to have metaphysics guided by physics – i.e. something that cannot be captured in any thesis or factual belief – then what is it? (1996, pp. 169-170)

Set Lewis's observation up next to his belief that some worlds are unphysical, his belief that some worlds have different physics from ours, and his assertion that physics has a brief to provide a full inventory of *all* the fundamental properties instantiated in the world, and the tensions should be apparent. Elements of the same ad hoc approach have been amply noted in Jackson, Campbell, Tye, and Chalmers, in as much as they address the matter directly at all.

That this tentativeness and provisionality impede progress ought to be equally apparent. Why are philosophers still at pains to point out that there is a good question about the scope of the physical, but at the same time content to wave it away and go on to characterise materialism, as if this task could be achieved satisfactorily without going through the arduous preliminaries? (See for example Jackson, 1994, p. 26: 'A fair question is how to specify precisely the notion of a physical property. I am not going to answer this fair question. Roughly, I will mean what is typically meant'; and we have seen the rest.) Such is the tug of the arch-intuition seemingly at work in even the very best philosophers: that common sense and time-honoured opinion are to be rescued first. But where conservative approaches have been doctrinaire, patchwork, impossible to articulate in detail, and tentative, and when they have led to apparent paradoxes, it may well be that some especially recalcitrant weed needs uprooting. That weed is the nettle I

propose to grasp, and twist hard, in this section. To do this, I need to go back well before Descartes.

If folk ideas are to be accorded such respect, let me be so folk-like as to examine the rough, dictionary-style etymology of some of the terms of interest – an approach having much in common with the cross-cultural and linguistic investigations of Kathleen Wilkes (1988) and points of contact with the etymological investigations of Martin Heidegger (1959). The Greek word φυσικα (along with the associated verb φυσω, or φυσειν in its infinitive form: having to do with *production, birth, coming into being*) is a source of the cluster of English words with *physics* at its core. In Greek it denoted, roughly, the nature of something, or, very simply, the way it is. Φυσικα, from which our *physics* is straightforwardly derived, is the study of the natures of things in the world. So general is the term that it is highly likely that the English verb *be* itself has the same Indo-European source as φυσικα (the Greek 'root' is φυ-; the reconstructed Indo-European root is **bheu*, which also yields the important Sanskrit root *bhū*; see Partridge, 1966, and Monier-Williams, 1899). The original Greek physics was a very broad-ranging and inclusive study indeed, though the term φυσικα itself apparently acquired along the way many subsidiary meanings. The same sort of problematic multivalence is to be observed with such very broad 'nature' terms as the Chinese *tao* (Watts, 1975, pp. 37-55), and indeed our own English *nature* itself, both of which have some interesting relevance to the present definitional question.

Heidegger's analysis of φυσικα alone occupies five pages ('for words and language are not wrappings in which things are packed for the commerce of those who speak or write. It is in words and language that things first come into being and are'; 1959, p. 13). In the course of the analysis he disputes the propriety of its translation as *natura*:

This basic Greek word for the essent [translator's coinage for *das Seiende*: roughly, that which exists] is customarily translated as "nature." [...] But with this Latin translation the original meaning of the Greek word *physis* is thrust aside, the actual philosophical force of the Greek word is destroyed. [...] *physis* originally encompassed heaven as well as earth, the stone as well as the plant, the animal as well as man, and it encompassed human history as a work of men and gods [...]. But if, as is usually done, *physis* is taken not in the original sense of the power to emerge and endure, but in the later and present signification of nature; and if moreover the motion of material things, of the atoms and electrons, of what modern physics investigates as *physis*, is taken to be the fundamental manifestation

of nature, then the first philosophy of the Greeks becomes a nature philosophy, in which all things are held to be of a material nature. [...] Thus the Greeks become essentially a higher type of Hottentot, whom modern science has left far behind. (Heidegger, 1959, pp. 13-15)

But, we may protest, this is equally unfair to Greeks, Hottentots, and the very concept of *natura*. And to read *nature* as 'the motion of material things, of the atoms and electrons' is not necessarily what those who would equate φυσικα and *natura* have in mind at all. Rather, both the natural and the physical are to be deemed as all-encompassing as Heidegger claims that original φυσικα is.

The crucial point is that the physical, or equivalently the natural, was simply and broadly identified with the *causal* – the succession or web of concrete events that constitute the *world*, as opposed to such static and abstract domains as the mathematical, or Platonic *ideality*. The point is well made by Adam Smith, in a passage written just two decades after Bolingbroke's that serves as our epigraph:

In the ancient philosophy, whatever was taught concerning the nature either of the human mind or of the Deity, made a part of the system of physics. Those beings, in whatever their essence might be supposed to consist, were parts of the great system of the universe, and parts, too, productive of the most important effects. Whatever human reason could either conclude or conjecture concerning them, made, as it were, two chapters, though no doubt two very important ones, of the science [*i.e.* physics] which pretended to give an account of the origin and revolutions of the great system of the universe. (Smith, 1776, Book 5, Chapter 1, Part 3)

And he goes on to decry how the moderns (*qua* post-Cartesians, we may presume; see next subsection) have rent this unified domain and study asunder:

But in the universities of Europe, where philosophy was taught only as subservient to theology, it was natural to dwell longer upon these two chapters than upon any other of the science. They were gradually more and more extended, and were divided into many inferior chapters, till at last the doctrine of spirits, of which so little can be known, came to take up as much room in the system of philosophy as the doctrine of bodies, of which so much can be known. The doctrines concerning those two subjects were considered as making two distinct sciences. What are called Metaphysics or Pneumatics were set in opposition to Physics, and were cultivated not only as the more sublime, but, for the purposes of a particular profession, as the more useful science of the two. (*loc. cit.*)

Early enough in modern English a distinction is indeed made between natural existents and those non-natural existents that would enter into Smith's chapter concerned with the divine, at least (as opposed to the human mind) – as if the world (itself presumably

'natural') were not the sum of all there is. We see a parallel distinction between the physical and the non-physical; for example: 'Physike, which is the studie of naturall things: metaphysike, which is of supernaturall things' (an illustrative citation from 1586 used in the entry for *Physic* 1., OED; the deplorable abuse of the term 'metaphysics' is widespread among newcomers to the study of philosophy, who can, it seems, cite eminent authority, though the actual OED entry for 'metaphysics' puts things right). Indeed, for most of their histories the *nature* group of terms and the φυσικη group run parallel, with the Greek φυσικη being esteemed as more learned than the common Latin *natura* (cf. the 'class'-distinction between *ethics* and *morality*). Why the modern divergence, so that it is possible for a philosopher to claim to be an adherent of naturalism though not of physicalism or, in the case of Kim Sterelny, to make a special point of claiming to be both? ('My approach is not just physicalist, it is naturalist'; Sterelny, 1990; he maintains that there is a methodological side to the term that remains meaningful, to do with its connecting philosophy seamlessly to the 'natural' sciences.) The answer could well be just this point – that the scope of physical science has become uncertain, for reasons to be explored shortly. Meanwhile, it is reasonable to suppose that for most of the last twenty-five centuries the domain of the science called 'physics' was the entire natural world – that causal whole we call the world, *simpliciter*.

To epitomise, a useful history-of-ideas myth might go like this: If there were philosophers of mind in the ancient Greek world they were physicalists (and naturalists, like Sterelny). As Bertrand Russell points out, matter was 'primarily anything we could touch, though the first step toward mystification was taken when Empedocles included air' (1927a, p. 131). Leaving aside Leucippus and Democritus (who as philosophers were so mulish as to insist on something solid and gross-matter-like at the base of absolutely everything), the addition of such soft, elusive, and rarefied matter was generally taken as a great explanatory advance. Since there was no aid yet to be had from neurophysiology or from the various functionalisms (some would say we are still waiting), the mysterious air was plucked at avidly as the most subtle and ill-understood manifestation of matter: eminently suited, therefore, to account for what seemed especially mysterious about living things, including ourselves.

Like all good myths, this is more than a *mere* myth. The main words in very many

languages for the presumed vital principle in animals and humans have to do with air, wind, or breathing. We could start with the word *animal* itself: Latin *animale*, a thing furnished with *anima* (from which French gets *âme*), which is roughly the equivalent of *soul*, and cognate with Greek άνεμος, meaning *wind*. Other easy to find examples include Latin *spiritus*; Greek θυμος (related to Sanskrit *dhūmas* and Latin *fumus*, both meaning *smoke*; see Partridge, 1966) and ψυχη; Hebrew *rū'kh* and *nēšāmā*; Sanskrit *âtman* and *prāna*; and English *breath* itself. All of these import both some sort of life-force, life-principle, motive force, or inner selfhood, as well as literal breath. (In the case of *prāna*, the principle is applied essentially to all living things, including plant life. To lack *prāna* is to be non-living. Ψυχη is also a more encompassing term than many have thought.) *Rū'kh* is discussed by Locke (*Letters to the Bishop of Worcester*, p. 457).

One of the most elaborate developments of this shift employed the term πνευμα (foundation of *pneumatic*, which we have seen already), standardly translated as *breath* or *spirit*. This term is also discussed by Locke (*loc. cit.*). It was applied to a fluid, presumed real, located in space and time, coursing through real vessels in the body, whose presence and nature was supposed to account for movement, digestion, thought, and just about everything we as living things get up to (see notes on the opinions of Erasistratus in Galen; *On the Natural Faculties*). This doctrine and its many ramifications and refinements persisted up until Renaissance times (indeed, beyond: see Culpeper, 1653, and New Agers everywhere).

Locke's remarks concerning *rū'kh* and πνευμα have been criticised by Richard Aaron:

Locke adduces the materialistic accounts that accredited authors, such as Virgil and Cicero, give of the human spirit. It is air or fire or breath, and we find terminology of the same sort even in the Scriptures. So that to regard the mind of man as substantially material is no new doctrine. (Locke's appeal here seems a little unfair. Such writers hardly used those terms to defend the materiality of the mind. All their terminology was material and they obviously chose the least solidly material, air and fire and the like, to describe mind.) (Aaron, 1955, p. 145)

But from another point of view it is Aaron who is being unfair, in a manner reminiscent of Heidegger's noted earlier. Why should the ancient authors so *readily* choose the terms they did? Surely because there is a tendency of both the ancient and the modern

temperament to hypostatise mind – to identify it with some *bona fide* substance. At least the ancients referred to by Locke and Aaron avoided proposing new *categories* of substance for mind. Perhaps they speculated that whatever variety of material it is that underwrites mind is bound to be pretty motile and mutable (as a certain breed of moderns still think), and what is more so than air, fire and breath? It quite *properly* never occurred to these authors to propose any *immaterial-yet-causal* substance for mind, since the notion would have struck them as inexpressible and incomprehensible. Even for us moderns, it has taken us many centuries to come to the idea, and about three to disabuse ourselves of it (so far!).

Anyway, there were mundane problems in accounting for the subtleties of the soul with networks of pneumatic plumbing. The evidence from dissections, for example, simply failed to support the hypothesis. When confronted with such an impasse, apply the epicycles – or in this case, make the fluid in question even more *subtle, fine, only analogous* to breath proper. But of course, it remains *physical*. How could it be otherwise? Everything in the world is physical! By the time of Descartes, however, the Copernican revolution was in full swing, and the time was ripe for a similarly momentous reordering, to deliver comprehension of the microcosm.

1.4.2 *Descartes's physics and after – a useful modern myth*

The myth delivering to us Descartes's new disposition of the world is too well known to rehearse in detail here. What we do need to remind ourselves of is the fact that most of his contemporaries saw Descartes as a dangerous mechanist – in respect, that is, of almost all of the living world. And indeed, his expulsion of a permeating, vaporous soul in favour of standard gross mechanical explanations of the human body must have seemed strikingly deviant. A notable example:

[I]n a cord ABCD, if one end D is pulled so that the other end A moves, the exact same movement could have been brought about if one of the intermediate points B or C had been pulled, and D had not moved at all. In similar fashion, when I feel a pain in my foot, physiology tells me that happens by means of nerves distributed throughout the foot, and these nerves are like cords which go from the foot right up to the brain. When the nerves are pulled in the foot, they in turn pull on inner parts of the brain to which they are attached, and produce a certain motion in them; and nature has laid it down that this motion should produce in the mind a sensation of pain, as occurring in the foot. (Descartes, *Meditations*, VI)

Animals were sheer mechanisms; humans were almost entirely sheer mechanisms, but somehow possessed of an *immaterial* soul (a *what* soul?) – somehow safely sequestered away, somehow out of the loop. This was a means of preserving the soul from the depredations of naturalistic determination and the like, but equally physics was liberated to pursue explanation of virtually the whole world as if it were as mechanistic as clockwork. Regardless of the particular vicissitudes of Cartesian physics proper, the effect of this revolution on seventeenth-century and later scientific thinking is inestimable.

It must all have seemed quite a bargain. The ruinous cost is much more evident now – though not yet the full extent of it, I suggest. It is perhaps possible to view many of the main strands in twentieth-century philosophy as struggles with the Cartesian legacy, as mentioned earlier. One of these struggles concerns method: the effort to come to a new understanding of the relations between possibility, ontology, and thought, to be addressed at length in Part 2. In particular it is profitable to view the current difficulty with the demarcation of the physical as such a bequest. If mind is indeed beyond nature (for the moment supposing that notion to be coherent), there is a causal realm that is not physical. But Descartes did not enlighten us concerning the *precise* nature of the boundaries (unless in terms of a jury-rigged physics lacking a notion of momentum) – though definite boundaries there would have to be – and provided a perilous precedent: if mind is not natural, there may be still other domains of genuine, causal reality that are also not natural. Not to be a naturalist came to seem a workable option for theorists in many fields of philosophical inquiry. Recent manifestations of this uncertainty are reminiscent of the moves made by the original pneumatic theorists (though perhaps the early adopters of Empedocles' radical supposition of substantial air provide the better analogy). Some of our moderns have recourse to the most subtle and mysterious features of matter to trade across the great divide between mind and the 'physical' world. What better explanation can there be for the flourishing of that college of neo-pneumaticists whose speciality is the attempt to exploit the mysteries of quantum mechanics to explain mind? Beguiled by the common Cartesian inheritance of modern thought, they take the essential *difference* of mental phenomena as a given, and no matter how piously and productively mechanistic their thinking on other topics, mind must forever remain a deep vortex of problems for them, absolutely demanding the

reassuring obscurity afforded by mixing the inscrutabilities of neurophysiology with those of the even more baffling quantum domain. Of course, the whole world is a quantum world; let's incorporate our specific failure to understand mind into our general failure to understand the world as a whole. (Some examples are: Penrose, 1989, 1994; Lockwood, 1989; Hodgson, 1991.)

This traffic at the ill-defined frontier between the 'physical' and the 'non-physical' goes both ways. In some of the standard interpretations of quantum theory, of course, an appeal is made to unexplicated but *assumed* consciousness to negotiate certain conundrums in the physical theory (Goertzel, 1992).

The world has been made safe for physics, by an almost complete exorcism. But a whole new domain has sprung into existence in which some of the perennial problems have been quarantined. Their term is now up, and they are released to wreak vengeance on an unready philosophy. How to place mind in the world was a question adroitly and ingeniously avoided by Descartes, but it's back now as the *hard problem* par excellence (Chalmers, 1996, xii-xiii). Freeing physics has fettered philosophy.

1.4.3 *Actual materialism*

Whether the two myths enlarged upon above convey the literal historical truth behind the contemporary confusion about materialism is not a matter of concern; the intent has been to fix ideas, and to allow a sharper focus on a certain alternative view, which I now present.

Let us take for granted Lewis's analysis of fundamental properties – the perfectly natural properties (PNPs) on which all other properties supervene, as a matter of *a priori* truth, as Lewis himself invites us to do. We can alternatively translate the whole affair into terms of Armstrongian universals, as Lewis suggests we can; but here I shall adopt Lewis's way. Let us also take very seriously the first task Lewis gives to physics: it has to provide an inventory of *all* the PNPs instantiated at our world. Now, we have the option of simply declaring that this task is achievable in principle. In short, we can *define* physics with respect to the actually instantiated PNPs, so that all of them count as physical.

This simple move is likely to be resisted on many fronts, and to be taken as a naïve

avoidance of the issues. But it is really just to take seriously, as I say, one of the rulings Lewis (preceded by Quine) gives concerning the scope of physics, and to play down his others, which are after all incompatible with it.

Lewis, as a professing materialist, is unlikely to be much perturbed by the suggestion that ours is *bound* to be a materialist world. But he would certainly be unhappy with the stipulative route to this conclusion; materialism gets to be true but trivial. I shall argue shortly that this should not occasion alarm, and is in a way inevitable. For the moment let us just ask: How and when should we expect physics ever to abdicate? And confronted with what *evidence* should materialists, about mind or any other domain, ever be predicted to throw in the towel? As we have already noted, Quine and van Fraassen (with opposite motivations) find physics compassing whatever may come its way, adjusting even its deepest assumptions to accommodate new phenomena. (Van Fraassen asserts that nothing could stand in science's way: 'there cannot be such a thing as a true "science stopper" at all'; 1996, p. 158; see also *ibid.*, p. 177, note 9.) We have seen just this resilience and plasticity many times in the actual history of science. Air was taken as matter, and the definition of matter changed. The curious behaviours of the amber (ἤλεκτρον in Greek, – that's where we get our word *electron* from) and the lodestone were taken not as indications of supernatural anomalies (*what* anomalies?): physics shifted ground and ushered them into its expanded stronghold. The very geometry of our space was shown to be non-Euclidean (though this was previously thought *logically* impossible), space and time became spacetime, 'atoms' came to be composite, Newtonian–Laplacean determinism broke down, and the whole sorry demise of classical physics unfolded – and lo, we have all (well, most of us) managed to retain a belief that the nature of the world is essentially physical. A miracle? Certainly not. Physical knowledge advances, often in discontinuous leaps; I suggest that the assumption we have all been implicitly working with is not that it could ever come to light that physics needs supplementing, but that there *is* a true physics (at least one!) to be discovered (*pace* van Fraassen), that is properly *by definition* an account of the whole causal nature of the world.

One philosopher who has recently come to a conclusion similar to mine is Chris Daly. He finds, after critically reviewing a number of alternatives in the literature, and citing

in support such theorists as Israel Scheffler (1950), Noam Chomsky (1972), and Tim Crane and DH Mellor (1990), that there is unlikely ever to be a satisfactorily principled way of distinguishing physical from non-physical properties:

My overall conclusion is that the notion of a physical property is not well-defined. The Kripke-Putnam account of natural kinds does not successfully carry over to the expression 'physical property'. A family resemblance, or exemplar-based, definition of 'physical property' [...] cannot provide a sufficiently precise definition of the kind needed in analytical metaphysics. Similarly, defining the expression 'physical property' by reference to the methods of and principles of physical science leaves the distinction between physical properties and all other physical properties without metaphysical interest. (Daly, 1998, p. 212)

I suggest that Daly's conclusion is similar rather than identical to mine because he does not take the extra step beyond consideration of materialism as a contingent thesis that I take in the next subsection, and because he is clearly reluctant, as I am not, eventually to accept the claims of a frankly materialist monism – also to be explained shortly.

There are also a few troubling methodical issues in Daly's otherwise trenchant treatment of the topic. There is insufficient rigour in his handling of the crucial details concerning properties and their 'existence' at worlds:

We have the notion of a property existing at a world. For example, the property *being red* exists at this world. I will take the notion of a property existing at a world to be a primitive notion. [...] If a property exists *contingently*, it exists at some possible worlds [...] (*ibid.*, p. 197)

But for all the appearance of clarity here we are never certain whether Daly's properties exist at some world by being *instantiated* at that world or merely by being *instantiable* by the laws of that world (though there is some circumstantial evidence that Daly intends *instantiated*). This is a further manifestation of insufficient attention to the modal dimensions that will occupy us at such length in Part 2. Nor is even the pivotal *prima facie* distinction between physical and non-physical properties made at all lucidly, so that we are left at the end with worries about quite what has been achieved. Consider the following statements (all derived with minor rewording from statements on Daly's p. 198):

S1: Physicalism is the thesis that (at least) all actual individuals are physical individuals, and that all actual properties are physical properties.

- S2: Physicalism needs a distinction like this: every property is either of type (1) physical or type (2) not physical.
- S3: Type (2) properties include *being-painful*, *being-a-stagflationary-economy*, and *being-a-Truman-supporter*.
- S4: According to physicalism every property of type 2 which exists at the actual world is 'a physical property at the actual world'.

Now, even without examining context, given that the terms in these statements pretend to univocality there is clearly a problem here. The unexplained classification of *being viscous* as a chemical property (perhaps along with the choice of *being made of cellulose* as paradigmatically biological) does not relieve our apprehension. I diagnose the very pervasive tendency not to keep a sufficient distance from a conventional doctrine and vocabulary, when the very issue is the standing of that doctrine and the propriety of that vocabulary.

1.4.4 Necessary materialism

An immediate objection to the *automatic materialism* established in the preceding subsection might be that it, like its rivals, gives no principled means of classifying other worlds as materialistic or not. Granted that this world is materialistic, by whatever means we have arrived at that conclusion, the whole point about materialism for most of the theorists we have heard from earlier is that it is a contingent thesis, true at this world but not at all worlds; or to put it differently, *the world* might not have been one at which materialism is true.

The best definitional move (or stipulative move, if you don't like it!) seems to be to have it that all worlds are materialistic. We declare the inventory of PNPs instantiated at any world to be the list aimed at *by the physics of that world*; it too is a world at which materialism is true, since *its* ideal physics is automatically true and complete. So some worlds (let it be assumed, for now) have different classes of instantiated PNPs from ours (and the classes at some might even be disjoint from the class at ours, for all we know); with Lewis we might say that there are worlds at which physics is different, but we should not concur with him that some worlds are non-materialistic, because that is ruled out by our definition. And the correct modal observation to make about *the world* (without rigidifying our reference, of course) is that it might have been one at which

physics differs from the physics that happens to be true – not that it might have been one at which materialism is false. And we allow in all of this that Lewis is on track in the observation regarding PNPs that we cited earlier: they are ‘the ones whose sharing makes for resemblance, and the ones relevant to causal powers’ (1983a, p. 347).

Of course I am well aware that this is a radical proposal, and by the end of Part 2 it will develop into an even more radical one. I have already justified giving such a drastic move serious consideration; it remains for me to show the evidence that this is the *best* way to use the word ‘materialism’, and does not commit us to vicious vacuity (which vacuities automatically are, by some accounts), or land us with the very undecidabilities with which the alternatives are rife.

1.4.5 *Consequences, objections, and replies*

Concerning the vacuity objection, I think we can accept with van Fraassen that the notion of materialism tends that way no matter which way we turn it. John Earman’s conclusion (at the end of a final section headed *Physicalism versus Physicalism*) lends support here:

The above remarks prompt but do not answer the question: Is there a way to construe the doctrine of physicalism that makes it worthy of philosophical debate? (1975, p. 567)

On the basis of a review of the literature appearing in the two decades since Earman wrote, I am ready to give this question a negative answer, but with a twist: there may well be no construal of materialism *by which it has coherent alternatives* with which to advance the philosophy of mind. If others claim there is such a construal, I have no objection. I only ask them not to shy away from providing the details. I agree with Lewis and Jackson that properties dependent or supervenient on physical properties should by extension be accorded physical status, so let all properties instantiated at a world be physical (according to the true, completed physics of that world, remember). I think I am driven to this conclusion, given my acceptance of Lewisian *a priori* supervenience of all the higher-order properties of a world on the PNPs of that world. Can those holding an alternative view be as clear as this? Again I invoke Earman’s judgement:

The task of providing a useful and non-question-begging criterion for sorting higher-order properties into the physical and the nonphysical seems to me near hopeless and, perhaps, ill-founded. (*loc. cit.*)

And again I accept this, but go further: on the evidence of the literature since Earman wrote, the task *is* hopeless, and it *is* ill-founded, and what’s more this is *necessarily* so. For myself, I am happier to live with a vacuity than with a confusion lacking any prospect of resolution, and to move on to questions with better foundations and more susceptible of informative treatment.

Another who has examined the scope of the physical (in the context of an analysis of neutral monism, to be discussed in Part 3) and who draws a radical conclusion, differing somewhat from my own, is Gaëlen Strawson:

Insofar as I am any sort of materialist, then, I am an agnostic materialist: our current conception of the physical is fundamentally incomplete on its own terms. Quite independently of the mind-body problem, it is a commonplace that there is a sense in which our ordinary concepts of space, time, and matter are profoundly inadequate and partial representations of the nature of the reality to which they are a response. [...] Is agnostic materialism so uncommitted as to be vacuous? Does it turn the word ‘material’ (or ‘physical’) into nothing more than a descriptively empty synonym for ‘real’? Is an agnostic materialist really just a “?-ist”? (1994, pp. 98-99)

I am happy to answer the underlined question in the affirmative, and see this affirmation as tending towards liberation rather than towards absurdity.

Does all of this make *me* a materialist? I might protest at the question, being unhappy with the terminology, which suggests that anyone could coherently claim to be anything else! Perhaps though, given the *benign* vacuity of materialism in what I take to be its best reading, it can make little real difference whether along the way I call myself a materialist or not – but I mislead less if I do so style myself. And I do hold that actual physical science, with the worldview it has developed and is developing, promises more than any other investigative program to issue in explanations of our single, seamless world. I also hold, and shall argue at considerable length in Part 2, that the reach of physical science extends surprisingly far beyond the actual world. For these reasons I am content enough to label my position *necessary materialism*, rather than merely, say, *necessary monism*. Certainly I would not wish to be associated with anything called *neutral monism*: the qualifier ‘neutral’ suggests a default dualism in the background,

which I do not discern.

As for Strawson's '?-ism', it should be clear, from my views stated earlier concerning the origins of the relevant words and concepts, that the likes of Strawson and myself have as well-founded a claim as anyone else to use of the more complete designation 'materialist'. Strawson writes on this classificatory point:

Why do I call myself a materialist, rather than a "?-ist"? My faith, like that of many other materialists, consists in a bundle of connected but unverifiable beliefs. I believe that experience is not all there is to reality. I believe that there is a physical world that involves the existence of space and space-occupying entities that have nonexperiential properties. I believe that the theory of evolution is true, that once there was no experience like ours on this planet, whether panpsychism is true or false, and that there came to be experiences like ours as a result of processes that at no point involved anything not wholly physical or material in nature. Accordingly, I believe that however experiential properties are described, there is no good reason to think that they are emergent, relative to other physical properties, in such a way that they can correctly be said to be nonphysical properties. (*ibid.*, p. 105)

My differences with Strawson concern both the alleged unverifiability and the apparent contingency of the claims he makes on behalf of materialism. He leaves relatively unexplored the modal underpinnings of materialism, and I think that these are central to establishing a case against most dualist arguments, which themselves typically rely on modal analyses, as we have seen and will continue to see in Part 2 and Part 3. We note in this last excerpt from Strawson his focus on spatial qualities in his characterisation of material reality. Spatial or spatiotemporal qualities, at least since Descartes, have of course dominated attempts to characterise the physical. I have had little to say about them here, taking it for granted that contemporary physical science is so thoroughly spatiotemporal in tenor that it was hardly necessary to be explicit on the point. But consider this suggestion from Russell:

Let us coin a word, "chrono-geography", for the science which begins with events having space-time relations and does not assume at the outset that certain strings of them can be treated as persistent material units or as minds. Then we have to ask ourselves first: can the science of matter, as it appears in physics and chemistry, be wholly reduced to chrono-geography? (1927b, p. 294)

Russell goes on to suggest that chemistry is indeed reducible to physics, in principle, and physics (very plausibly, at least) to his chrono-geography. But from my own point of view it appears that Russell's understanding of the term 'physics' (like Heidegger's, exhibited earlier) is too closely bound to the physical science of the early twentieth

century, with its advances brought about by Einstein's epochal relativity theories and the astonishing new work being done on the sub-atomic structure of matter. It must have seemed that these developments provided new and highly specific characterisations for physics, and though the vision was not yet (as it still is not) of a truly unified and 'finalised' science, the hope that this would soon be achieved was much in the air. From beyond the century it is perhaps clearer that our notions of material reality must make fewer presumptions. Russell's science of chrono-geography looks eminently identifiable with physics itself. An analysis of the precise nature of spatiotemporal relations, along with an assessment of their claim to primacy, would be needed to justify this claim completely. This cannot be attempted in the present thesis, of course, though a good deal will be said about the nature of spatiotemporal reality in Part 2. (Further discussion of Russell's views will be undertaken when we address neutral monism in Part 3.)

None of the foregoing talk of commitment to materialism being trivial and in the end unavoidable will satisfy someone who objects to the scenario of even the arch-dualist Descartes being cornered into an acknowledgment of materialism's claims. Of such an objector I ask: When was Descartes *ever* satisfactorily classified, or pinned down to a fully-worked-out metaphysics? Consider the formidable awkwardness of his ontology: two distinct real domains, both *causal* through and through, one spatiotemporal and utterly incapable of instantiating mental properties (though Locke caused a sensation by suggesting that it was in God's power to override this restriction; see *Essay*, 4.3.6), the other capable of instantiating mental and temporal properties (only?), and utterly forbidden any spatial properties. Yet there is supposed to be causal commerce between these domains – *in* time, and *at* a spatial location! One may reasonably judge that it is simply not a coherent position, despite the opinion of most eminent theorists of our time that the position is 'logically' possible (see Part 2). As Paul Teller (who argues cogently but on different grounds from mine that materialism is not contingent) points out concerning realisation of functional states by supposed immaterial spirits,

if we did have a "medium" which related inputs to outputs with systematic causal connections, what reasons would we have for saying that this medium was not physical? (Teller, 1984, p. 157)

Questions dealing with causation, along with spacetime (or space and time) and laws, will be addressed again in Part 2 and Part 3, as I have said. But now let us see more of

Teller's analysis:

We think materialism must be contingent because we can do something that has the form of imagining it turning out to be false, and because we have to work very hard to establish that it is true, work to which empirical research often seems relevant. But we can likewise do something that has the form of imagining Hesperus turning out to be distinct from Phosphorus, and we had to work very hard and empirically to establish that they are one and the same. Until Kripke everyone assumed that these facts about the identity of Hesperus and Phosphorus demonstrated that the identity is contingent. That was a mistake. I am suggesting that we are making an analogous mistake about the contingency of materialism. (*ibid.*, p. 155)

As for those objectors who maintain that a principal task of philosophy is to account for our strongest intuitions regarding the nature of reality, I remind them that intuitions can be schooled and shaped, as we have just seen, and are not nearly as uniform, universal or immutable as might be supposed. I have already mentioned in passing Wilkes' project of scrutinising words in various languages denoting something akin to our concepts of consciousness and mind. If Chinese lacks an exact equivalent of our word 'consciousness' (very recent in Western cultures and languages in any case), is Chinese *defective*, or do speakers of Chinese rather have radically different categories for making sense of the world (Wilkes, 1988)? I shall spare the reader the obvious and overworked Sapir-Whorf-style analysis that could be done here, and prescind from any speculation about the direction of fit between our intuitions and our vocabularies. Suffice it to say that intuitions are diverse, and so is vocabulary; and at this point in our inquiry into matter and mind I favour re-focusing on the task of 'limning the true and ultimate structure of reality' (Quine, 1960, p. 221), rather than any more exploration of history-of-ideas myths. There will be words enough devoted to a resumption of that exercise when we later come to home in on the notion of consciousness.

Perhaps the point that a thoroughly universal (and therefore necessary) materialism is counterintuitive misses the mark for a more elementary reason, anyway. We may not have any *philosophical* intuitions worthy of the respect accorded to our other intuitions. Some of the best breakthroughs in the history of the discipline have been viewed as heretical (in fact I have given Descartes's own revolution as a prime example), but that is just how philosophy can be expected to make progress. The overwhelming heuristic success of Descartes's program should not compel us to think that it issued in new truths *directly*, nor even, as discussed above, that it is ultimately even coherent. It did

good service in advancing physics for a time, and at least provoking philosophy into worthwhile productivity. But to advance now we may well need to let it go – bathwater, baby that refuses to grow, and all.

The pervasive fallibility of intuitions is a major theme of this thesis in more than one way and at more than one level; we may take as a reminder the case of *action at a distance*, which has been considered by many (under the influence of Greek philosophers of various schools, especially the Democritean atomists, as revived by Gassendi) to be unquestionably impossible (see extended discussion in Part 2; and further generalised treatment in Part 3).

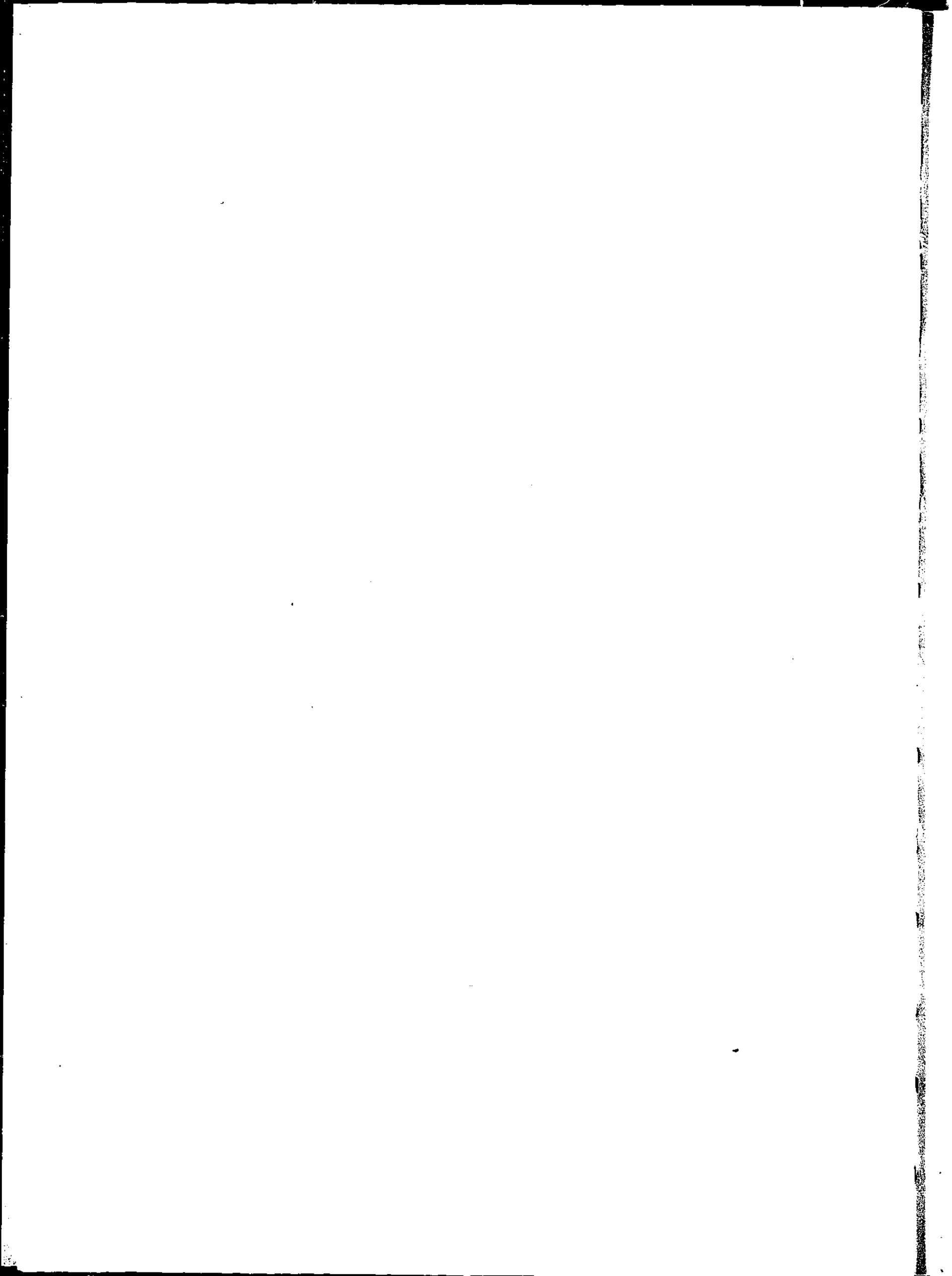
I shall be using points from this first part in what follows, and returning to give further argument for the idea of a necessary materialism using the conclusions of my later investigations. The close relevance of this approach to materialism for the investigation of mind and consciousness will, I hope, become more apparent as we proceed. In Part 2 we shall explore modal epistemology and metaphysics. There I shall again be taking a Lewisian perspective, though again with serious reservations – this time about Lewis's views concerning the range of the possible. Part 2 will dwell in particular on space, time, and causation (after a fashion); and we shall look at some good reasons to resist uncritical acceptance of allegedly 'Humean' theory, just as this first part warned against swimming with the still prevailing Cartesian current.

This part and the next, while long and substantial in their own right, may be read as preliminaries – necessary for getting our house in order, the better to meet head on and unashamedly the threat to philosophical progress posed by the latest dualists, in Part 3. It seems that Cartesian sympathisers arise in every era, and ours, with its unprecedented burgeoning of activity in the philosophy of mind, is no exception. But we should remember that materialism – indeed, *necessary* materialism – was here first. It was the very air that the pre-Cartesians breathed.

Part 2



Necessity and
Nature



PART 2

Necessity and Nature

It is observed by arithmeticians, that the products of 9, compose always either 9, or some lesser product of 9, if you add together all the characters of which any of the former products is composed. Thus, of 18, 27, 36, which are products of 9, you make 9 by adding 1 to 8, 2 to 7, 3 to 6. Thus, 369 is a product also of 9; and if you add 3, 6, and 9, you make 18, a lesser product of 9. To a superficial observer, so wonderful a regularity may be admired as the effect either of chance or design: but a skilful algebraist immediately concludes it to be the work of necessity, and demonstrates, that it must for ever result from the nature of these numbers. Is it not probable, I ask, that the whole economy of the universe is conducted by a like necessity, though no human algebra can furnish a key which solves the difficulty? And instead of admiring the order of natural beings, may it not happen, that, could we penetrate into the intimate nature of bodies, we should clearly see why it was absolutely impossible they could ever admit of any other disposition?

— David Hume (as the character Philo)

(*Dialogues Concerning Natural Religion*, 9, 1779) ~

2.0 INTRODUCTION: A NEW LOOK AT MODALITY

In Part 1 I argued against the customary assertion that materialism is at best contingently true. I argued that we should construe the physics that underwrites materialism at our world as the all-compassing study of causal reality at our world, no matter what turns causal reality is found to take; and, in effect, that at any world with any discernible regularities worth speaking of, there had better be a physics for that world as well. In short, I argued for what we shall now call a minimally specific *Ur-physics*, encompassing all of the modal vicissitudes to which concrete reality is subject, and under whose banner materialism could be construed as utterly necessary. A corollary might be that naturalism and materialism are one and the same, and that naturalism – difficult to explicate at the best of times – is best regimented to be a kind of necessary default assumption in all of our philosophical investigations of reality.

If we do not proceed in some way like this, I suggested, we risk a descent into incoherence, paradox, or heuristic *culs de sacs*. Examining one particular descent is the topic of this whole work, and we arrive at the hub of that topic in Part 3, where we

consider an important recent dualist approach to explaining consciousness in a predominantly 'physical' world. But there is much more preliminary work to do, which is the task of this, the largest and I think most heterodox of the three parts.

Our theme is *Necessity and Nature*. It is all very well to achieve what we have up till now, but when we try to fill in the content of our Ur-physics, so far we draw a blank. While we are not embarked on a program of investigation in physics itself, which is altogether *ultra vires*, we may be able to define limits or outlines to discipline our modal, Ur-physical speculations. Typically physicists themselves do not stand in need of such a restraint; some need a broadening in certain of their speculative horizons. Carl von Weizsäcker, for example, a nuclear physicist who savoured philosophy of a Kantian flavour, held that the exact disposition of the spectral lines of hydrogen is 'logically' necessary and deducible *a priori* (von Weizsäcker, 1981).

As Hume's Philo suggests in the passage I have chosen as this part's epigraph, there is no way for us to establish with certainty the modal status of the 'whole economy of the universe'. That is as much as to say that the broad contents of Ur-physics elude all 'human algebra'. Perhaps von Weizsäcker is right in his claim that the nature of hydrogen is entirely determined by the broadest necessity, but mistaken in claiming that this nature is deducible – at least by any means available to beings like ourselves. Daniel Dennett identifies what he calls *Philosophers' Syndrome*: 'mistaking a failure of imagination for an insight into necessity' (1991, p. 401). But there is also a syndrome disposing us the opposite way: mistaking an exuberance of imagination for an insight into the possible. This is at least akin to, and may even be a separate manifestation of very same tendency as, the predisposition of early mapmakers to place sea-dragons in any distant uncharted waters. It is humanity abhorring a vacuum – a primitive Parmenidean (or Ptolemaic, as I shall be suggesting) impulse toward plenitude, served by assuming that our place in things is special, and that elsewhere is different.

Nowhere is this impulse more floridly expressed than in philosophers' speculations concerning how things stand at *other worlds* – those forever unnavigable regions simply begging the metaphysical enthusiast to run riot with conjectural cartography. Richard Mason has called to account the notion of logical possibility itself, concluding this way:

[B]eginners to philosophy are often bewildered by their teachers' assertions that it

is *logically possible* to fly unaided to the moon, or that any work on artificial intelligence is pointless because its goals are *logically impossible*. I am saying that the beginners are right to be bewildered. That kind of super-science is phoney. We had better look at it closely, then stop it. (1988, p. 23)

His own close looking is a careful analysis of what evidence people use for claims of logical possibility, which is by way of claims to imaginability and conceivability. This evidence he finds wanting; and while I shall not follow Mason's own precise route to suspicion of this inflated notion of 'logical' possibility, my efforts here work to something like the same end as his efforts.

David Lewis was taken as a major champion of materialism in Part 1. We continue homage to him now as a pioneer in regimenting modal discourse, and while there will be tough dissent concerning his views on the *range* of the possible, the core principles of his modal realism will be vigorously defended as by far the best foundation for modal investigations.

2.1 WHY MODAL REALISM?

2.1.1 Explication of modal realism

Modal realism (MR), as we shall understand the theory here, is most fully articulated in Lewis's remarkably original and compelling *On the Plurality of Worlds* (1986a). The theory starts with the claim that *all possible worlds, including the actual world, are equally real*. Some (including recently Alex Pruss, 2001a) have called this theory of Lewis's *extreme* modal realism, reserving the more general term to embrace also any more or less non-fictionalist account of possibilia; but this seems unfair, almost pejorative, in view of the direct nature of Lewis's fundamental claim compared with the compromises offered by others. MR may usefully be analogised to more familiar and widely accepted realisms, for example realism about other minds than one's own (which few think *extreme*!), or realism about times past and future, as opposed to presentism.

Following straightforwardly from Lewis's basic realist commitment is an *indexical account of actuality*. If all worlds are equally real, then the actual world can only be distinguished as *this* world, rather than by any ontological criterion. So too, according to realists about past and future times, the state of the world *now* has no distinctive ontological status – it is just *this* state. And realists about other minds – most of us, after all – have a similar semantics for the pronoun *I*.

The distinction we now commonly make between the real and the actual in explicating MR is in fact not favoured by Lewis himself. In *Anselm and Actuality* we find this passage:

Prior slips here in presenting the indexical analysis (as a tall story). He writes, 'this word "actual" must not be taken as signifying that the world in question is any more "real" than those other worlds...' But 'real' (even in scare-quotes) is presumably indexical in the same way as 'actual.' Hence we can no more say that all worlds are equally real than we can say that all worlds are equally actual. (Lewis, 1970, p. 88, note 7)

And it comes as a surprise to note that Lewis is quite consistent through the years in not applying the epithet *real* to his worlds. We, however, shall follow the reasonable convention that Prior follows, and maintain the distinction adumbrated above, as it is unequivocally licensed by Lewis's own term *modal realism*, even though he regretted having given his trademark theory this 'bad name' (1986a, p. viii). (What name would

have been better? Perhaps *Modal Copernicanism*; see 2.4.1.) Note that my term *total reality* and John Leslie's very near equivalent *Reality*, both of which will be introduced later, rely on this same convention. *Actual* and *real* can be distinguished by their histories to make sense of their uses with separate meanings: *actual*, as deeply embedded in Latin, French, and English usage, has often had a connotation of the *practical* and the *relevant* about it that is often absent from *real*, deriving from its connexion with *act* – almost *praxis* – and sometimes *presence*. So the only world with which we have all our causal dealings is most aptly called the *actual* world, whereas the Latin *res* at the root of *real* was always capable of indicating states of affairs removed from immediate concerns. *Real* may be taken, then, as a mere existential intensifier, somewhat like *genuine* or *fair-dinkum* (in Australian speech), and this is benign and fitting enough.

Of course, still less than by Lewis himself is such a sharp distinction between the terms *real* and *actual* drawn by theorists of other persuasions (see an example from David Armstrong, in 2.1.4).

In terms of modal logic, MR provides a *model for S5*, a system which itself has a strong claim to be the best formalisation of our standard modal intuitions. In our general modal thinking, most of us want every possible world to be *possible with respect to* (that is, in modal-logical terms, *accessible from*) every other world, and this is just what S5 offers and what MR clearly delivers ontologically. (I am not here speaking of the Lewisian logic of *counterparts*, which as Stuart Brock reminds me does not fit S5. I mean just the intuition that makes us want possible scenarios to be possible *simpliciter*, to put things informally. In other words, we would want to say that which worlds we take to be *at all* 'possible' is not determined by which world we find ourselves at. That is what MR captures very straightforwardly.)

Lewis is also an uncompromising advocate of a very permissive *principle of recombination*, according to which anything can stand in any spatial or temporal relation to anything else provided no contradiction arises in any true description of the arrangement. Any of the contents (strictly, counterparts of any of the contents) of any world may notionally be arbitrarily recombined with any others, and the result corresponds to a real world. This principle is allegedly Humean, but I shall argue, in

agreement with Galen Strawson, that it arises from an excusable but ultimately unsupportable reading of Hume. I therefore call it not *Humean* recombination (though Lewis himself makes the attribution directly enough; 1986a, p. 87) but *Lewisian* recombination (LR). For now the important point is that this and any other principles of recombination (and for that matter most worldmaking principles) are independent of the basic ontological account of possible worlds. As we shall see in 2.2 and elsewhere, many critics tilt objections at Lewis's MR that could only be effective as attacks on LR. Very often Lewis answers these critics with the observation that the objections tell equally, if at all, against the objectors' own favoured modal ontologies, just because the *manner* of worldmaking is orthogonal to the *matter* of basic ontology. These are usually very effective responses from Lewis, and I shall develop some more of the same sort on Lewis's behalf. The whole question of recombination principles and other worldmaking moves, as opposed to the question of basic ontology, is of central importance to our investigations, and will be treated in detail in 2.3.

Lewis may be taken to hold that *a world is a large concrete particular*, composed mereologically of smaller concrete particulars. Lewis expresses uncertainty about this claim because he is unsure how the terms *concrete* and *abstract* are properly to be used (1986a, pp. 81 ff.), but most readers start off reasonably happy with the notion of concreteness and will not immediately share Lewis's worries. We shall return to this theme in 2.2.4, and at the very end of Part 2. For now, just one slight qualifier: I do not mean that Lewis is not committed to abstracta existing at worlds, only that the causal histories at his worlds can be given in purely concrete terms, by a preliminary and approximate understanding of 'concrete'.

The relation between objects that fixes them as parts of the same world – the *worldmate* relation – Lewis generally claims amounts to *spatiotemporal relation*, or at least mere spatial or mere temporal relation. No objects that are *not* related spatially or temporally are parts of the same world; no objects that *are* so related are parts of distinct worlds. A fairly telling critical attack has been mounted against this account of the worldmate relation; Lewis himself is ready to consider alternatives (such as causal connexion, which he already takes to be restricted to worldmates), and we shall look into this also in 2.2.5 and 2.4.4.

So ordinary concrete objects, concerning which modal claims are standardly made, are *worldbound*: there are no transworld individuals. This is the basic and official Lewisian line, and it is allied to the doctrine of *counterparts*, touched on just now in passing: modal claims about (worldbound) objects are made true (a way of putting it that I fear can be seriously misleading!) by the features and fates of their counterparts at other worlds. We shall not have much to say about this important feature of Lewis's comprehensive modal theory, nor about *hydras* (metaphysically ungainly entities with parts not all at the same world – as real and as 'concrete' as any entities, in Lewis's system, but hardly relevant to our interests here; see good discussion in Bigelow, 1990, p. 209), though we shall deal with worldboundedness itself again in 2.2.4.

Perhaps it is clear from this exposition of MR – carefully distinguishing it from LR, as formulated – that it is a very serviceable model by which to form and to regulate our modal intuitions; at least, I say that it is, and I intend to show how those who do not make use of it fall into error and paradox far more easily than those who do take the initial counterintuitive plunge that Lewis beckons us toward. More than a few philosophers have come to feel that investments in 'ontological seriousness', as John Heil calls it, return handsome dividends:

I believe we have a right to be suspicious of anyone who embraces the formal apparatus of possible worlds while rejecting the ontology. Indeed, I think we might be more suspicious of formal techniques generally, when these are deployed to answer substantive questions in metaphysics and the philosophy of mind. So long as we remain at a formal level of discourse, it is easy to lose interest in what might ground our claims. And this, I think, has led to the kind of technical sterility characteristic of so much contemporary analytical philosophy. (Heil, 1998, p. x)

I share this dismay at the 'technical sterility' of much that passes for philosophical inquiry in our tradition. (For me this is also tinged with disappointment at a widespread obsession with the *fascinating impasses* that the use of ineffective techniques brings us to. If Mason's students are bewildered by talk of 'logical possibilities', what must they make of many philosophers' ataractic acceptance of their failure to *solve problems*, once and for all?) Heil claims inspiration from CB Martin; the very title of their joint essay 'The Ontological Turn' (1999) captures a tidal change in analytical philosophy to which I am more than happy to trim my own rudder.

I emphasise that most of the work that I shall put MR to does not require actually

executing the doxastic *entrechat* that Lewis incites us to: all we need to do for the most part is think *as if* we believed MR to be a correct account (which in a way is what the advocates of the modal theories discussed in the next two subsections effectively do, though not always well, I claim). Nevertheless I shall offer some reasons along the way for accepting MR as literally true – reasons independent of Lewis's own. Mainly in order to demonstrate the superiority of full-blown MR in the capacity just discussed, we now turn to an examination of some alternatives to it, and then to a defence of MR from some of the more weighty objections.

2.1.2 *Ersatzism*

In much of our discussion of modality (especially now that we approach ersatzism) we shall have to distinguish standard world-based talk and the much less usual unrestricted talk that takes all worlds and their contents as on an equal modal footing. This is the mode of talk Lewis uses when he claims that there *are* talking donkeys (1986a, *passim*). It is also very closely analogous to the tenseless use of verbs, readily accepted in philosophical discourse. Usually the context will make it obvious which mode we are in, as it often does in the case of tenselessness; where it does not make it obvious, I shall try to be explicit. But I ask my readers to be indulgent: some quibbles may have to do only with a wrong understanding of the mode we are in.

Ersatzism has it that there is necessarily exactly one *concrete* real world – the actual world – and an indefinitely large number of real *abstract* worlds, which are the merely possible worlds, each of which represents a way the actual world could have been but is not. (The polemically coloured term *ersatz* is Lewis's, to make a distinction from his own 'genuine' modal realism; e.g. 1986a, pp. 136 ff., from which I draw much of the characterisation to follow in this section.) From the vantage point of actuality – of this very world, rigidly picked out – the ersatzist should say that all concrete objects are *actual*, and all are *actual*. (And note: if ersatzism is right, that is the only vantage point anyone *has*, speaking unrestrictedly!) Exactly what kinds of 'real' abstract objects non-actual objects and worlds are is a further question: Lewis broadly distinguishes *linguistic* ersatzism (pp. 142 ff.; the non-actual worlds are maximal descriptions, or maximal consistent sets of sentences, or some other linguistic presentations of ways the actual world might have been) and *pictorial* ersatzism (pp. 165 ff.; the non-actual worlds are

like pictures presenting ways the actual world might have been). He has specific ontological objections to both these versions of ersatzism, while considering the linguistic to be at least supportable (and partly just a matter of philosophical preferences), unlike the pictorial. A third grouping Lewis labels *magical* ersatzism (pp. 174 ff.; non-actual worlds are some or other kind of unanalysable, structureless, abstract primitives). Common to all three brands of ersatzism so distinguished is the problem that they do not reduce, and so do not explain, the very modal dimensions that possible-world theory is intended to account for. In most of what follows I shall intend by *ersatzism* linguistic ersatzism, the most serious opponent to MR. We do not need to examine the arguments for or against the ersatz proposals in close detail; for *many* of our purposes, as I have said, the choice between them and MR is not crucial. I have said that an assumption of MR guides intuition best of all, and that this heuristic distinction is its first obvious attraction for the work we want to do, though it is not the only attraction. Armstrong, whose theory of modality differs radically from Lewis's, acknowledges this strength in realist theories of possible worlds generally (Armstrong, 1989, pp. 6, 17). There are two important and related respects in which MR goes further than and does yield different hard theoretical results from ersatzism, however.

First, ersatzism obviously treats actuality as something much more than indexical. The nature of this primitive actuality and the problems arising directly from it will be addressed in 2.1.4, and will loom large when we come to examine anthropic approaches to modal doxastics in 2.4, and its sequel in 2.5.

Second, whatever non-indexical reason there may be for *this* way the world could be being 'actualised', rather than some other way, is thoroughly elusive. (We should note that for ersatzism there is presumably an abstract object of the same category as all of the merely possible worlds corresponding to the actual world – its *abstract double*, we might call it.) On the face of it, there can be no principle (certainly no *causal* principle) to decide in favour of *this* way rather than some other way that does not also rule out the other ways as *impossible* ways – as ways the world could *not* in fact be! There is no analogue of these difficulties in MR. We shall examine how an ersatzist might at least go some way towards achieving what we shall easily achieve using MR when we look at some work by Peter van Inwagen, in 2.4.2. Perhaps explaining possibility is the easy

part for theories of modality: it is the actuality that makes for hard going.

2.1.3 *Combinatorialism and other fictionalisms*

Fictionalist accounts of modality (see for example Rosen, 1990) however committed or uncommitted they are to employing the possible-world formalism common to MR and the various kinds of ersatzism, they certainly shun any sort of commitment to *real* possibilia. Rather they take possibilia to be mere inventions or conventions to enable modal discourse, which itself has practical utility. Many of the troubles that Lewis finds in ersatzism (particularly the problem of distinguishing possible from impossible representations) are present also in fictionalism. What I shall be urging in 2.3 against tendencies in all possible world accounts (including the 'logistic' principles that Lewis adopts) regarding their excessive dependence on linguistic and logical means of worldmaking I shall intend as applying *a fortiori* to all forms of fictionalism.

At least in the case of Armstrong's combinatorialism, a frankly fictionalist account (Armstrong, 1989, pp. 49 ff.), we may add the objection that it is founded on an implausible logical basis. While MR and the ersatzisms pay respect to the widely held intuitions formalised in S5, Armstrong's theory does not. Only properties that are actually instantiated may be recombined to generate possibilia; a consequence of this already implausible suggestion is that, from the imagined point of view of a fiction in which certain properties instantiated in the actual world are *not* instantiated, the actual world and a vast class of other perfectly plausible fictions are not possible, since they have, from the fictional, imagined point of view, alien properties.

I propose generally to set these fictionalist theories aside as unlikely to be productive as we search for secure foundations for modal belief, whatever other merits may be attributed to them.

2.1.4 *Neo-Spinozism: lip-service possibilism?*

There are many current theorists who pay only lip service to possibilities beyond a very circumscribed modal perimeter. These include writers who take seriously some sort of a hard distinction between ways that *this* world could be (perhaps articulated in terms of a conventionally established notion of natural possibility as distinct from logical possibility) and descriptions of worlds alleged to be logically coherent but implicitly too

'outlandish' to amount to 'genuine' possibilities. Yet such writers are firmly committed to a full-blooded *logical* modality – it's just that, overtly or covertly, they are unwilling to have any world *be* in a way that is 'merely' logically possible! I call this tendency *neo-Spinozism*, though perhaps *crypto-Spinozism* would be apt also. Spinozism is the doctrine that only the actual world is possible; the modern variants I speak of ring a curious change on this, deeming some features of the world essential to it and effectively modally invariable while, with apparent arbitrariness, 'according other features free mutability.

The tendency I discern is hardly systematic or visible enough to amount to a position, unlike MR, the ersatzisms, or the fictionalisms. I shall discuss it further in 2.4.1, where we begin to examine indirect but potent ways of founding modal beliefs. An incipient form of it may be diagnosed even in Armstrong's avowed 'actual-world chauvinism': 'The possible is determined by the actual, and so, saving recombination, cannot outrun the actual' (1989, p. 56), and in such *obiter dicta* as: 'Such a notion [*sc.* causation by transcendent entities] is perhaps barely possible...' (1980, p. 152). But what is *bare possibility*? In the context, we are invited to entertain a proposal as *logically* but not *genuinely* possible, I suggest. Here is another example of uncertainty regarding the relations between some logical and some genuine possibility, perhaps with some admixture of epistemic or doxastic modality, which also illustrates less than lucid usages for *actuality* and *reality*:

The actual is expanded so far that every possibility for a space-time is actual in some island universe. There are no *mere* possibilities. Nor would I deny the logical possibility that the actual has these swollen dimensions. But I see no reason to think that this is the way that the actual really is. One of the possibilities for the actual is that every possible spatio-temporal system exists. But it is only one possibility. (Armstrong, 1989, p. 17)

The context is, of course, censure of Lewis's modal realism; but one rarely detects such a strong scent of equivocation in Lewis's work itself, and I suggest that this is largely because of the clarity and definiteness lent by a commitment to MR. As for the related business of epistemic possibility being surreptitiously conflated with logical possibility (or what I prefer to call broad possibility, or possibility *simpliciter*), this will be addressed in Part 3. There I shall also argue that Chalmers is a prime neo-Spinozist, and that this concealed tendency of his thought is a philosophical fatal flaw. I suggest that

his not taking the ontological prerequisites of possible world discourse seriously is strikingly like the mote he finds in other philosophical eyes: the sin of not taking consciousness seriously. There would be no problem if he were not to use such talk; but he does in fact rely on it: and he *needs* to do so.

Now though it is proper to survey the dissenting opinions concerning modal realism, and to answer them in turn.

2.2 DEFENDING MODAL REALISM

2.2.1 Counterintuitive?

The first objection most people have to MR is that it is catastrophically counterintuitive. Most people, if they think about it at all, consider actuality and reality to be one and the same: the actual world is the only one that truly matters to them, since what happens to them only happens in the actual world – so it alone is *real*. It is strange even to have to express such a thing, so basic is the sense of it. But there is a perfectly sound response available to this objection. Why, after all, should we imagine that we are equipped with any reliable intuitions at all about modal metaphysics? And what enhanced utility is there in preferring positions favoured by intuitions of utterly spurious provenance? One is reminded of a very famous Wittgenstein anecdote, worth repeating for a new context:

Upon meeting a friend in the corridor, Wittgenstein is alleged to have said: 'Tell me, why do people always say it was *natural* for men to assume that the sun went round the earth rather than that the earth was rotating?' The friend responded, 'Well, obviously, because it just *looks* as if the sun is going round the earth.' To which Wittgenstein replied, 'Well, what would it have looked like if it had looked as if the earth was rotating?' (Lockwood, 1989, p. 15, with citation of Elizabeth Anscombe and Tom Stoppard.)

Wittgenstein's interlocutor might have answered acutely: 'One thing is certain: it *didn't* look or feel to our ancestors as if the earth is rotating!' to which Wittgenstein could have replied: 'But the earth *was* rotating, wasn't it? So it *must* have looked and felt as if it was!' In fact, of course, the earth would look precisely the same in either kinematic condition. And so it is with the many-world ontology: we bring to consideration of it precisely zero *a priori* warrant for our intuitions for or against, so it is best to put these pre-philosophical intuitions firmly aside.

That said, it is interesting to examine the situation of Lewis's MR in Lewis's own terms. Lewis himself is a keen supporter of many common-sense intuitions, odd as such an assertion may seem in the context of evaluating his MR. A great deal of Lewis's effort in *Plurality of Worlds* (1986a) is spent supporting pre-philosophical or pre-existing philosophical convictions – often the very ones that MR can be deployed against with devastating effect. We shall see some of this paradoxical conservatism of Lewis's as we proceed (in 2.2.6, for example), and it is telling that Lewis believes a great deal in and

consequent upon MR to be in fact not counter to common intuitions and usages anyway, even if its core tenet itself is counterintuitive:

A worthwhile theory must be credible, and a credible theory must be conservative. It cannot gain, and it cannot deserve, credence if it disagrees with too much of what we thought before. And much of what we thought before was just common sense. Common sense is a settled body of theory – unsystematic folk theory – which at any rate we *do* believe; and I presume that we are reasonable to believe it. (*Most of it.*) (Lewis 1986a, p. 134)

Here it may be interesting to touch on just one clear example of his tendency to save intuitions. A common objection to MR is that, if every possible life situation really occurs at some class of worlds or other, there can be no point to doing the right thing, no matter how the 'right thing' (prudentially or morally) is to be established. After all, in the overall order of things every possible response to every situation necessarily occurs willy-nilly. This fact is unaffected by what happens at *this* world. A straightforward but intuition-disregarding response is: 'So much the worse for ethical theory! Metaphysics trumps ethics.' But this is not Lewis's way; the details need not detain us, but the conclusion of the half-dozen pages he devotes to the objection is indicative of his conservatism (though perhaps also of a certain tendency to forensic overstatement):

An ethics of our own world is universalistic enough. Indeed, I dare say that it is already far too universalistic; it is a betrayal of our particular affections. If my modal realism has any bearing at all on matters of value and morality, it pushes me toward common sense, not away. (1986a, p. 128)

2.2.2 *A wildly inflated ontology?*

Modern astronomy and cosmology deliver a universe far more extended in spacetime than could commonly have been stomachied by earlier generations (though of course even in ancient times there were some who conjectured that the universe is truly vast – or even infinite in extent). We in the twenty-first century are inured to all that. But Lewis (along with quantum many-world theorists and certain cosmologists, all with their own good reasons) now postulates vast numbers of such universes. *Non sunt multiplicanda entia praeter necessitatem?* Lewis clearly has a case to answer. There is, however, an elegant theoretical simplicity in MR which compensates for an abundance of entities sufficient to induce an apoplexy in *Venerabilis Inceptor* Occam. There is only one broad kind of real world, rather than the two kinds required in any species of

ersatzism – abstract (merely possible) and concrete (actual). And the simplicity goes further than an uncluttered inventory of kinds, because there is no problem of accounting for *this* way, of all the ways for a concrete world to be, being 'actualised', as mentioned earlier. With ersatzism, actuality remains as unproblematic as it always was. (Of course there is huge variety in the *contents* of the Lewis worlds: but that is a problem for LR and the related principles Lewis uses to stock his worlds. It is a ground for rejecting those worldmaking principles, but as I shall argue it has little to do with MR proper. For ersatzism, there is an exactly parallel problem: and why should it help that the ersatzist jumble of contents is *abstract*?)

We might compare the relations between MR and ersatzist accounts to those between the nuclear theory of atomic structure developed in the early 20th century (by James Chadwick and others) and its predecessor, 19th-century atomic theory (associated with John Dalton). The more modern theory allows a much greater number of token basic particles in the world with the pay-off that there are a mere three types of basic particles (protons, neutrons, and electrons); in the older theory there were around ninety basic and irreducible types of atoms to be found in the world, with the dubious pay-off that there were fewer tokens of the basic types. Moving from ersatzism to MR we reduce the number of basic types from two (abstract worlds and concrete worlds) to one (concrete worlds), but with this move (as opposed to the move away from Daltonian theory) the total number of tokens remains the same. (If anything, MR postulates exactly one fewer, because ersatzism presumably requires an abstract double of the concrete actual world, as we have seen; or, some might prefer to say, a concrete 'implementation' of exactly one abstract world that underwrites that abstract world being the actual world.)

We shall return to Occamist questions in 2.2.6 and beyond, but in the meantime we need to see how others have objected to the extravagant number of entities that Lewis allegedly postulates on *non*-Occamist grounds. The response developed just now might do some work against some of these objections also, but they deserve the separate treatment we shall now accord them.

2.2.3 *Too many worlds?*

Some philosophers have objected that there are strict logical reasons for rejecting

Lewis's huge number of real worlds as impossibly inflated. Peter Forrest and David Armstrong (1984) argue that Lewis's principles for worldmaking can be demonstrated to permit too much. One version of the difficulty is this: a new world is constructible by Lewis's principles from duplicates of any worlds already allowed to exist – and so also from duplicates of *all* worlds. But such a new 'maximal' world must itself, of course, be real. So it both is and is not a member of the set of all real worlds – a style of paradox widely held to be unconscionable. Other versions lead to contradictory assignments of cardinality to the number of real entities.

Lewis' answer to such objections is complex, but in the end he concedes the need for a qualified principle of recombination to replace the first formulation: duplicates of anything can coexist with duplicates of anything else in the formation of a world, 'size and shape [of spacetime] permitting' (1986a, p. 89). The proviso is spelt out like this:

The only limit on the extent to which a world can be filled with duplicates of possible individuals is that the parts of a world must be able to fit together within some possible size and shape of spacetime. Apart from that, anything can coexist with anything, and anything can fail to coexist with anything. (pp. 89-90)

Lewis wrestles inconclusively with the problem of what these supposed limits on spacetime might amount to; Nolan (1996) argues that the objection fails, and rehabilitates an unrestricted LR. But I propose to pass over the detail of these arguments and counterarguments, which are highly technical mathematically and so far unresolved (see also discussion of Pruss, 2001a, shortly). I observe three things:

First, it is interesting that Lewis chooses to restrict his recombination principle this way rather than question anything more fundamental in it – as if this general restriction, while necessarily having specific consequences of *some* unfindable sort or other, should in no way bring into question the firm commitment to LR itself. We shall examine reasons for being very suspicious even of the qualified principle, and make much use of that suspicion, in 2.3 and later.

Second, Lewis's preferred resolutely spatiotemporal approach is evident in the kind of general restriction he favours, though the Forrest–Armstrong objection itself need not be construed in spatiotemporal terms; and we shall look at problematic consequences of this aspect of his theory in 2.3.3, where suppositions in his treatment of space and time are brought into question.

Third, one is justified in insisting that the Forrest–Armstrong argument, along with others like it, is again not an objection to MR proper, but to unrestricted LR (before Lewis's very general modification of it, that is). I have said that MR and LR are orthogonal, and I agree with Lewis that objections trading on features of LR are a problem for the various ersatzisms, commitment to which must, after all, entrain a commitment to *some* sort of recombination principle also. Pruss (2001a) writes against Nolan's support for LR, with complex mathematical argument relying in part on the physics of various kinds of actual types of particles and how these might occupy locations in space, and concluding that in MR using LR-style principles, the totality of worlds would – most counterintuitively, once more – not constitute a set. While Pruss's position is ingenious, it relies on many speculations regarding what is 'logically' as opposed to 'physically' possible, and tacks to and fro a little too much for us conveniently to pursue its detail. But of interest with regard to my contention is this comment:

It is worth noting that the worries involved here are somewhat less pressing if one does not take possible worlds to be *concretely* existent as they are in EMR [sc. Lewis's MR]. For although it is very plausible that the well-defined collection of substantial existing material objects such as Lewis's possible worlds would form a set, it is somewhat less plausible that the collection of all merely possible worlds considered as pure possibilia is a set. The cardinality argument as an objection thus weighs more heavily against EMR than against other theories of possibilia, whether ersatz, Leibnizian, or whatever. Nonetheless, the argument does demonstrate that any reasonable theory of possibility needs to deny that there is a *set* of all possible worlds. (Pruss, 2001a, p. 119)

But the difference in plausibility appealed to here is spurious. For a start, the concrete entities under discussion are not occupants of a single world – each *is* a world. Why should we think that our meta-set-theoretic intuitions – notoriously fallible even under favourable circumstances – are at all reliable in this new and unfamiliar context? More pointedly, Pruss's own cardinality argument against LR (and so, as he thinks, against MR) is conducted explicitly in terms of particles that share a common spacetime manifold, and it is largely the cardinality properties of the aggregate of points forming the manifold that give rise to the difficulties that Pruss finds for Lewis. But whatever are the spacetime intuitions that may be compelled to do duty in considering concrete entities sharing a manifold, these cannot be pressed into service in respect of concrete entities *not* sharing a common manifold. Next, we might agree that both the ersatzist

and the Lewisian are exactly to the same extent under a compulsion to agree to the assertion that: 'There are n worlds, and the aggregate of them is not a set.' But if logical criteria are to have any purchase, both theories are equally vulnerable to attack on the basis of this assertion, since impossibilities and implausibilities regarding abstract and concrete entities are equally to be established by purely logical means (or so we are told!). If MR and ersatzism differ in any of these respects, ersatzism (which is ersatz realism, remember) must abandon the very claim that it would have us accept: that it is realism enough.

Stripped of the distinction he essays, Pruss's argument assails Lewis's position no more trenchantly than it does the ersatzist camp.

We shall touch on objections like Forrest and Armstrong's again in 2.3 and 2.5; now we examine more closely this matter of the concreteness of Lewis's worlds.

2.2.4 *Concrete worlds? Duplicates?*

In Lewis's MR all worlds and all of their contents (which is to say all of their parts, and we mean *parts relevant to causal history*) are concrete entities. But for most philosophers all concrete entities are essentially contingent: there are modal truths concerning how they might vary or how they might fail to exist, or have distinct duplicates. These features of concreteness, many claim, are incompatible with what those of us equipped with the standard 'S5 intuitions' expect of possible worlds, each of which is ultimately a non-contingent entity, along with all of its worldbound constituents (using our terms in non-worldbound fashion – analogous to the tenseless fashion for verbs, recall). When she says 'there is a possible world at which there is a chain of gold a light-year long', the ersatzist intends that the abstract entity mentioned is a fixed feature of one particular world: itself a non-contingent entity (we shall allow ourselves to say 'a *necessary* entity'), of the same broadest category as a mathematical or logical construct. Construed as worldbound, world-indexed, and 'modally isolated', the 'abstract chain' is not a contingent entity, so no *concrete* entity could stand in for it by any shift of theory, since to be concrete is to be contingent. At least, according to this line of reasoning, it would be a most counterintuitive metaphysics that allowed the concrete to be necessary, as is sometimes canvassed in the search for alternatives to a

God as the necessary foundation of the world's existence in discussions of the cosmological argument:

But farther: why may not the material universe be the necessarily existent Being, according to this pretended explication of necessity? We dare not affirm that we know all the qualities of matter; and for aught we can determine, it may contain some qualities, which, were they known, would make its non-existence appear as great a contradiction as that twice two is five. (Hume, *Dialogues Concerning Natural Religion*, 9. Here Cleanthes is speaking. See also the epigraph to this chapter; both that and this excerpt from Hume will be discussed in 2.3.2.)

But objections to MR of the sort outlined above derive from some slippage in the meaning of *concreteness*, which in a way is sanctioned by Lewis himself (see 2.1.1). Clearly concrete entities are contingent, if 'contingent' is to be meaningful at all! Any of them can vary, or fail to exist. But for all possible-world theorists this is either a matter of concrete entities having transworld identity locating them at some but not all worlds or, in the case of Lewis in particular, of their being worldbound but necessarily having equally worldbound and concrete counterparts at other worlds. For Lewis, concrete entities *qua* worldbound individuals are *all* unequivocally necessary entities; they *only* have modal variability – or are commonly deemed to have it – courtesy of counterpart relations.

It should therefore be immediately ruled out that a world or a part of a world be subject, *qua* world or world-part, to modal variation. Especially, worlds themselves ought to be thought exempt from simple duplication as worlds (though we have no principle so far that will rule out intrinsic duplicates of them featuring as proper parts of other worlds). If it has application anywhere, the principle of identity of indiscernibles surely has application in this context. Nevertheless the problem of indiscernible duplicate worlds recurs noisomely for Lewis. Armstrong argues that Lewis's own 'Humean' principle of recombination ought to settle the matter in favour of duplicate worlds for Lewis:

His possible worlds are, relative to each other, 'distinct existences' in the Humean sense. It is a mark of a distinct existence, at least according to Hume, that it permits the existence of any further distinct existence, including discernible duplicates [*sic*; *sc. indiscernible duplicates*]. I believe, indeed, that this is a true principle of the theory of possibility. But in a Realistic treatment of possibility, what is permitted exists. So should there not be indiscernible worlds? (Armstrong, 1989, p. 22)

Armstrong goes on to argue that there is then a problem of determining the cardinality of each presumably infinite class of duplicates of mutually indiscernible worlds, with

very damaging implications for the integrity of Lewis's theory as a whole.

Lewis himself is unsure what to think about such duplicates: 'I have no idea whether there are indiscernible worlds...' (1986a, p. 84). The matter is of some consequence, as are more general questions concerning 'distinct existences'; these will be touched on in 2.3 and 2.5. For now, though, it will be valuable to make a more developed and plausible case specifically against these putative indiscernible duplicate worlds.

If concrete worlds are, in the sense I have supplied above, necessary entities, as I maintain and as Lewis must agree (to the extent that he is happy with the very notion of concreteness), they are not subject to any 'Humean' recombination principle. As worlds, they cannot after all be thought to exist together at the same world! And this is what recombination would demand. Certainly they are 'distinct existences', if any necessary entities can be so classified. But are any necessary entities distinct existences in any sense *that is relevant here*? It is hard to find uncontroversial examples. If the natural numbers are entities, they are presumably necessary entities (*pace* EJ Lowe; see below), and each is discernible from every other. But is this sufficient for *distinctness*? Our standard criterion for distinctness is well enough settled in the case of ordinary concrete objects: *any object A and any object B are mutually distinct iff it is true of each that it can exist without the other*. (This formulation neatly shows the circularity of 'Humean' principles of recombination, by the way.) But this very plausible criterion cannot be applied to the natural numbers, since if they exist at all they exist necessarily, so none can exist without all of the others existing also. (Lowe, 1996, p. 117, worries about worlds at which there is only one simple concrete object; if an object is not distinct from its unit set, the number 1 is the only number to exist at that world. But, as Horatio reproaches Hamlet, *'twere to consider too curiously, to consider so*. In any case, perhaps in Lowe's almost-minimal case there would be also the set whose two members are the single concrete object and the number 1, so that the whole cascade of natural numbers can indeed gush forth.) Worlds in a system of modal semantics are not numbers, but if they are entities they too are necessary entities, and as such cannot be modally shuffled around as ordinary concrete objects, by proxy arrangements with counterparts, can be. If they exist, they all exist *together* (however that is understood) or they all exist *apart* (however that could be understood!); so recombination of them –

and, *a fortiori*, duplication of them – makes no sense. At least we have not seen an *argument* that it makes any sense.

To summarise: according to the only coherent reading of MR, worlds are not subject to modal variation – together they constitute the *ground* of modal variation for entities *at* worlds, which themselves vary modally only as they are construed as ordinary objects, having other-worldly counterparts; as such, worlds are necessary entities if they are entities at all, and so are not subject to 'Humean' recombination principles, which can only apply to ordinary objects as I have just limned them.

Failure to differentiate between the two relevant points of view here (the *worldbound* versus the *ontologically unrestricted*) are a commonplace, especially when there is discussion of non-actual but 'really' concrete entities. One further example will suffice. In a closely reasoned and formally sophisticated piece Robert Koons writes:

A composite or aggregate object cannot exist necessarily, since a constituent part of a mere aggregate can exist in the absence of the rest of the aggregate. Hence, an aggregate must have parts that exist only contingently, which means that the aggregate as a whole must exist contingently. (Koons, 1997)

And later in the same article, apparently referring to the above passage:

I have argued above that an aggregate like the cosmos [*sc.* the putative *multiverse*] cannot exist necessarily: there are infinitely many sub-aggregates that could have existed in its place, each sub-aggregate being just like the actual cosmos except in lacking one or more actual universes. (*ibid.*)

There may be several ill-advised moves here (difficulties with *aggregates* as opposed to *mere aggregates* for example), but the one of most interest for us is the one that takes us from a single world in the first passage to the ensemble of universes (or of worlds, as subsequent discussion more or less shows Koons to include). It should not convince us, and would simply beg the question against a modal realist reduction of modality, if that were Koons' intent – which, to be fair, is not *exactly* the case.

So we have seen that objections based explicitly or – more often – implicitly on the concreteness of Lewis's worlds are prevalent and polymorphous, but an application of the strictures presented just now is, I believe, sufficient to nullify at least the majority of them.

2.2.5 *Arbitrary worldmates? Arbitrary worlds?*

What is it that makes two real concrete entities features of, or parts of, the same world? The problem of the worldmate relation is perhaps best presented as emerging from the problem of the concreteness of worlds, though it must, as I contend, occur in analogous form in ersatz formulations as well. For an ersatzist, a simple declaration that two objects are worldmates is normally all that is forthcoming, but we are entitled to ask why this fiat should be sufficient. In Lewis's MR, the ensemble of all worlds may be understood as one large concrete object – the *maximal hydra*, we might say. Given the acceptability of this move, Lewis's worldmate relation constituted of spatiotemporal relations is open to question. Why should we not say that there is just one actual world (as Armstrong asks, for example; 1989), comprising all of the spatiotemporally isolated items Lewis distinguishes as *worlds*? What makes spatiotemporal relations so privileged in the domain of modality (when there are enough problems in the analysis of spatiotemporal relations in their own right)?

Among these problems are some arising when we contemplate so-called *dark matter*, thought to interact only gravitationally with the matter that we can observe. Suppose that there is a world with *super dark matter*, which interacts not at all with the 'other' matter at that world. But by what warrant could we claim that super dark matter is in spatiotemporal relation with 'ordinary' matter at such a world? If we believe in absolute spacetime, what warrants our saying that the two sorts of matter occupy the same manifold? But then, even if we do find a way around this problem, what becomes of the presumed primacy of spatiotemporal relations in determining that super dark matter and ordinary matter are worldmates?

Such concerns are serious, deeply affecting the individuation of worlds, as Lewis is well aware. To press a little further: when we contemplate any two worlds with indistinguishable spacetime manifolds but each with its distinct configuration of material entities occupying spatiotemporal locations, there may be no sufficient warrant for our saying that these are two distinct worlds. Lewis cannot distinguish this two-world description from an alternative by which we are contemplating just one world with two disjoint sets of contents, and causal relations only within each set. (It would of course be counter to Lewis's general system and lethal for MR if we were posit

haecceities for spacetime manifolds as a solution.)

Here we need not delve further into the details of the many objections of this kind: many are well set out in Bigelow and Pargetter (1987), and many others are astutely raised and adroitly refuted by Lewis himself (especially in 1986a). In the end I think we can most easily vindicate Lewis with some bold general considerations – though these would not be palatable to Lewis himself. Perhaps all candidate worldmate relations are *benignly* arbitrary. Perhaps the very notion of a *world* is ultimately itself picked out as merely a great procedural convenience, which suggests that all modal truths are relative truths, dependent on a prior choice as to how the multiverse is to be 'sliced'.

Now, here is a contention to be reckoned with. Beyond the sublunary business of finding a semantics for S5, *sub specie aeternitatis* there simply is what we may call *total reality* (cf. Leslie's use of the simple term *Reality* to designate whatever domain is more inclusive than some limited *universe*; 1989, pp. 46 ff. *et passim*). At least this certainly is literally true if MR is literally true. Total reality is as it is 'necessarily', through and through, just as all of its components are, as argued above. Worldbound beings like us, necessarily confined to and necessarily in 'causal' relation and so direct causal epistemic relation with only a minute portion of total reality, find reason to speak of *contingency* (along with causality itself, and cognate concepts), though there is nothing answering to it in total reality itself.

I call this necessity of total reality *Ur-necessity*, and it is useful to think of it as belonging alongside the *Ur-physics* established in Part 1. Using this brace of concepts we can take hold of reality more securely than with the customary concepts that serve for everyday purposes. The fact that nothing resembling either of these *Ur-concepts* is normally appealed to when viewing *sub specie quotidiana* that portion of reality we call *the world* is not relevant to their ultimate philosophical respectability. Bigelow distinguishes the two points of view like this:

If by 'the world' we mean the totality of all things which are contingent (in the Lewis sense of 'contingent'), then for Lewis the world is a Spinozistic one, with no accidental properties. The things in the world are contingent, in the Lewis sense – because although they do exist in some of those parts of the world which are maximally connected under chains of spatial, temporal or causal relations, they do not exist in other such maximally connected parts of the world. But in Lewis's metaphysics none of these things, indeed no things whatsoever, are contingent – if by that we were to mean that it is logically possible that they might not have

existed, with this term intended as a wide-open, unrestricted, Quinean quantifier ranging over all the things there are. (Bigelow, 1996, p. 151)

Very well; but it is one thing to acknowledge such a consequence of Lewis's way with worlds – indeed, an essential characteristic of it – and another to take seriously the *further* consequences. While not claiming to be in a position to refute varieties of *meta-Spinozism* (as we shall call it) such as Lewis's, Bigelow presses for exploration beyond them:

It is a good idea to see if we can sustain a theory in which there are things which are *contingent*, not in the Lewis sense, but in the sense that it is logically possible that they *might not have existed*, in the very widest-open Quinean sense of 'existed'. (*loc. cit.*)

Let us note that the meaning of 'it is logically possible that they *might not have existed*' is obscure, and that it would be in most contexts. (I am, by the way, assuming that Bigelow does not carelessly intend 'logically possible' to mean what some people covertly take 'epistemically possible' to mean; in that case he would simply be saying 'in the sense that, for all we [can] know, they might not have existed'. This is not plausible in Bigelow's context of attempting to 'sustain a theory'. See discussion in Part 3.) Take cabbages, for example. At least *they* exist in the least problematic way: concretely, now, in our region of the actual world. What can it mean to say *it is broadly possible that cabbages might not have existed*? (I solicit the reader's indulgence as I substitute *broadly possible* for *logically possible*; most philosophers will think this can make no difference.) But the time has come to speak of cabbages unrestrictedly – indifferently concrete and, as it were, 'ersatz'. And let us also speak of unicorns – concrete and 'ersatz'. It seems at first that Bigelow's position would have to go like this:

$C = \text{is a cabbage}$

$U = \text{is a unicorn}$

$\exists x Cx \wedge \Diamond \sim \exists x Cx$

$\Diamond \exists x Ux \wedge \Diamond \sim \exists x Ux$

Since we are assuming S5 throughout, these reduce unproblematically to:

$\exists x Cx \wedge \Diamond \sim \exists x Cx$

$\Diamond \exists x Ux \wedge \Diamond \sim \exists x Ux$

But these are just to say that there are, contingently, cabbages; and that there might be, and equally might not be, unicorns. Bigelow clearly cannot have intended that we strive toward establishing such homely truths as these.

He must therefore mean something else, perhaps that we eschew both MR and ersatzism and also maintain a vigil for some account according to which $\Diamond\Diamond$ does not reduce to \Diamond . But this suggests a more telling point, I think, regarding this passage from Bigelow. If we are genuinely to welcome this Quinean sense of 'existed' with all that it imports, Bigelow's plan looks like holding out for the retention of a primitive modality – or perhaps a recalcitrant iterable modality – in the face of *any* plausible reduction of modalities. After all, the contents even of ersatzist worlds *exist* in that they can be quantified over in the broad Quinean way; at some of them there are ersatz talking donkeys, just as in some MR worlds there are said to be concrete talking donkeys. If this is not so, then ersatzism has certainly fallen short of its primary goal. For both concrete and ersatz talking donkeys, as worldbound individuals (or as worldbound *parts* of individuals, to be perfectly catholic), it is true that they have existence of *some* sort (abstract or concrete) as a matter of the broadest necessity. But now it is as if we were invited to search for a way in which some possibilia should only be *contingently* possible: at least, this seems to be an inevitable consequence of Bigelow's suggestion, in its unrestricted context.

So even Bigelow, being among the few to give expression to the inevitable meta-Spinozist implication of Lewis's MR, then succumbs to the feeling that there may *yet* be, even in the MR context, some sense in which total reality may *vary*. But there cannot be, any more than there can be a sense in which mathematical realia may vary.

The moral might be, once more, that a cosy and plausible preservation of common-sense notions apt for everyday discourse is quite unsustainable *sub specie aeternitatis*. And according to MR, when we are speaking most unrestrictedly it is completely vain to employ notions like necessity and contingency at all. This is something like the limiting case of an observation of Quine's concerning common usage:

[...] if people thought that almost everything that happened at all happened by necessity, then [...] they would have little occasion to use the adverb 'necessarily'; mostly it would go without saying. (Quine, 1976, p. 68)

Presumably if we think that, *sub specie aeternitatis*, absolutely *all* truths are necessary truths, the term *contingent* can find no application *sub specie aeternitatis*; but then, nor can *necessary*, strictly speaking, have such application, because there is absolutely nothing to contrast it with. In Part I I argued that the term *materialism* strictly has no

application because there is no coherent immaterialism. so I now argue that there is strictly no Ur-necessity because there is no Ur-contingency. *Necessity*, if applied to concrete realia, is just our awkward quotidian way of expressing the state of things *beyond* the quotidian, somewhat as mystics and novice four-dimensionalists alike are wont to speak clumsily of all times being *present*.

On its very broad meta-Spinozist reading, then, Lewis's entire programme offers one very serviceable way of regimenting both day-to-day and philosophical speculation about the totality of reality and our places in it; and while there may be other ways that will do just as well, they will have to stake and defend their claims, alongside Lewis's MR with all of its auxiliary apparatus. In particular, all forms of ersatzism are vulnerable: they not only have actuality as an irreducible feature embedded in total reality, but it is they that cling to contingency as itself irreducible. One licentious way to put this is that the one actual world is irreducibly *contingently* actual, for if its actuality were necessitated there could be no other possible worlds! (Whatever principle it is that necessitates a certain world's being actual must equally exclude all other worlds from being actual, but such a *necessary* exclusion from actuality can be nothing other than a relegation to impossibility.)

To add to the ersatzist's woes, let us note that if actuality is a primitive, and is necessarily – and implausibly, I say – such as to inhere as an essential global property in exactly one world, by what warrant can we rule out *other* essential and necessary properties that may also inhere in exactly one world? And why should we then think that this and indeed an endless series of mysteriously differing *schmactualities* must all apply to the one mysteriously, contingently, *actual* world? The ersatzist, once lost in the luxuriance of his abstract vision of total reality, will find that Occam's razor is no machete with which to clear a way home to a single concrete actuality.

We shall explore some related themes in 2.4.4, but it is now proper to comment briefly on the role of the razor, now that it has been unsheathed.

2.2.6 *Occam defied?*

We have seen that there is a ready answer to those waving Occam's razor at Lewis's many worlds, and also their parts (see 2.2.2); but now the Occamists regroup once more

around this smaller quarry of parts of worlds. Forrest (1982) puts forward the *epiphenomenal objection*, to be dealt with fully in 2.5.2, to the effect that Lewis, as a modal realist, is not entitled to use Occam's razor to avoid a commitment to causally nugatory so-called *epiphenomenalons* haunting the actual world, while others like Forrest himself are so entitled. This amounts to asserting Lewis's special incapacity to justify induction, counter to our experience that induction is well-founded (or, pragmatically, well enough); so broadened, and varied, the objection is also deployed by Holly Thomas (1993) and Alex Pruss (2001a, 2001b). I shall defend Lewis, using weapons that he denies himself, in 2.5. Questions arising from justifications of induction will be crucial to establishing the revised view of alethic modality that I arrive at in the conclusion to the present part.

So much for objections to MR proper. There are many others; I have surveyed the recurrent themes, and suggested ways of dealing with them. But I also continue to maintain that the worldmaking principle LR is entirely orthogonal to MR itself. Other world-permitting or world-restricting principles could be put in its place (indeed, I shall be suggesting some). Lewis quite early modified his own principle under pressure from the Forrest–Armstrong objection, as we have seen; and ersatzists may themselves use some version of the principle favoured by Lewis (though they typically do not bother at all). We now turn to consider this separate question of recombination principles, and worldmaking in general.

2.3 RECOMBINATION RECONSIDERED

2.3.1 *Lego™-worlds or logo-worlds?*

Lewis's recombination principle LR supposedly has precedent in Hume, whom Lewis calls 'the greater denier of necessary connections' (1986b, p. ix). It is easy to see how parts of the *Treatise* and the *Enquiry* can be read as suggesting that there can be no necessary connexions between *distinct existences*, so that any two distinct entities could either coexist or not. Certainly Hume has much to say about supposed necessary connexions, as he also has more than a little to say about distinct existences, but it may not be what the majority have thought it is.

Something that may concern us is that many, supplying the word *necessary* where Hume omits it, go on to gloss that word as equivalent to *logically necessary* – which very expression I mean to impugn as we proceed. Mackie provides an example in the very Humean domain of causation:

What we can learn from Hume's discussion is that cause and effect must be logically distinct occurrences (or 'existences'), that it must be logically possible that either should occur while the other does not. (1974, p. 287)

In the next subsection (2.3.2) we shall explore this common reading of Hume and discover that this 'Humean' precedent for what I call Lewis's *Lego-worlds* (after the highly recombinatorial and paradigmatically distinct Lego™ blocks beloved of the very youngest philosophers) is quite problematic. In any case, we now have to note the genuinely Humean foundation of another strand in Lewis's worldmaking, issuing in what I call his *logo-worlds*. These are worlds that are first imagined, then scrutinised to ascertain that their description does not yield any logical contradiction, before final acceptance as genuine. We may take it that all Lego-worlds are also logo-worlds (since any Lego-world is ideally imaginable and has no description yielding a contradiction), but that not all logo-worlds are Lego-worlds. (This is because LR does not by itself, according to Lewis, deliver all possible worlds; 1986a, p. 92, where he rather cryptically informs us that 'We can't get the alien possibilities just by rearranging non-alien ones. Thus our principle of recombination falls short of all the plenitude of possibilities'. Cryptically, because this has more of an Armstrongian than a Lewisian tang to it; earlier the idea had seemed to be that we patch together 'parts of different possible worlds' to

yield a new possible world; p. 87. In any case, alien basic 'natural' properties will indeed not be accessible simply by recombination of known elements of the actual world, so it does look as if LR needs to be supplemented. Let us note in passing the oddity that Armstrong's actual-world-chauvinistic approach he calls *combinatorial*, where Lewis's approach is thought of as *recombinatorial*, as if the elements to be combined must be already 'to hand'. The reverse of this labelling would seem to be more apt.)

Lewis (1986a, pp. 90-91) adopts a suggestion of Quine's for the formation of worlds. There is a world corresponding to every mathematically describable pattern of 'occupied' points in any coherently mathematically describable spatiotemporal manifold. So the range of worlds is by no means restricted to worlds with 'grainy' matter such as ours is presumed to be. Matter in some worlds is distributed continuously (that is, where matter occurs its distribution is continuous, and as it were *arbitrarily* marked off from absences of matter); in some worlds (like ours) it is systematically grainy, and so more 'discretely' distributed in dense bundles; in others it has other mathematically describable distributions, including mixed types. This may appear to be the ultimate in Lego-world principles of recombination, according to which each block is point-sized. But even it will not deliver different *species* of matter (which Lewis allows on 'logical' grounds), nor 'immaterial' entities (which Lewis explicitly allows), nor even much to do with fields, energies, and other naturalia, it would seem (though Lewis does suggest it will work for fields). The determination of possible spacetime geometries themselves is also not accomplished by the Quinean principle.

I shall have something to say about the propriety of all such principles of recombination and compositional worldmaking; but let us now address worldmaking by logical means. The logical criterion by which putative worlds are judged – that a description yield no contradiction – is, I claim, quite insufficient. In short, as Lewis himself hints, the usual basis for conceiving something to be possible (and then perhaps constructing further possibilia by some application of a recombination principle) is our experience of the actual world:

To imagine a unicorn and infer its possibility is to reason that a unicorn is possible because a horse and a horn, which are possible because actual, might be juxtaposed in the imagined way. (1986a, p. 90)

But for a great deal of worldmaking, such as any involving allegedly possible properties alien to the actual world, we lack this basis. I say that we lack any *warrant* for a commitment to certain alleged possibilities. If we proceed from actual horns and horses to merely possible unicorns, we have our warrant for the operation by *ostension* – not by any knowledge of the nature of horns and horses, nor by any insight into what it is that *makes* horns and horses possible. But ostension of alien basic properties is not possible by definition (any more than *construction* of alien basic properties is possible – again by definition). I shall return to this question of warrant for modal belief in 2.3.4. Note once again, though, that this should be viewed as a problem for worldmaking in general: it has nothing directly to do with MR, but rather with LR and its allied principles. It is simply a problem that is brought clearly into focus using the assumption of MR, and that should certainly count *for* the theoretical potency of MR, not *against* it. Indeed, a case could perhaps be made for the problem of warrant being more severe for ersatzism, because the alien basic properties are to be thought uninstantiated by any concrete entity, though they must still somehow be ‘available’ for instantiation by concrete entities.

2.3.2 *Distinct existences, necessary connexions, human algebra*

The traditional reading of Hume’s supposedly well-known dictum that is taken to inspire LR is, as we have seen, ‘that there can be no necessary connexions between distinct existences’. Here are the excerpts from Hume’s *Treatise* that seem to come nearest to the wording just given:

- T1: It has been observ’d already, that in no single instance the ultimate connexion of any objects is discoverable, either by our senses or reason, and that we can never penetrate so far into the essence and construction of bodies, as to perceive the principle, on which their mutual influence depends. ’Tis their constant union alone, with which we are acquainted; and ’tis from the constant union the necessity arises. (*Treatise*, p. 400)
- T2: If perceptions are distinct existences, they form a whole only by being connected together. But no connexions among distinct existences are ever discoverable by human understanding. We only *feel* a connexion or a determination of the thought, to pass from one object to another. It follows, therefore, that the thought alone finds personal identity, when reflecting on the train of past perceptions, that compose a mind, the ideas of them are felt to be connected together, and naturally introduce each other. (*Treatise, Appendix*, p. 635)

- T3: In short there are two principles, which I cannot render consistent: nor is it in my power to renounce either of them, viz. that all our distinct perceptions are distinct existences, and that the mind never perceives any real connexion among distinct existences. (*Treatise, Appendix*, p. 636)

The following passage from the *Enquiry*, while not being at all similar in wording, is perhaps reasonably near to the intent of these *Treatise* passages:

- E: Hence we may discover the reason why no philosopher, who is rational and modest, has ever pretended to assign the ultimate cause of any natural operation, or to show distinctly the action of that power, which produces any single effect in the universe. It is confessed, that the utmost effort of human reason is to reduce the principles, productive of natural phenomena, to a greater simplicity, and to resolve the many particular effects into a few general causes, by means of reasonings from analogy, experience, and observation. But as to the causes of these general causes, we should in vain attempt their discovery; nor shall we ever be able to satisfy ourselves, by any particular explication of them. These ultimate springs and principles are totally shut up from human curiosity and enquiry. (*Enquiry*, Section IV, Part I, para 26)

T1 suggests that there *are* ultimate connexions between objects, but that they are epistemically unavailable; ‘the necessity’ arises from the perceived ‘constant union’ of objects. This passage is a exegetically highly problematic. Galen Strawson, whose exceedingly thorough and minutely evidenced advocacy of an *epistemic* reading of Hume on necessity is an inevitable recourse here, analyses it at some length (Strawson, 1989, p. 156 and pp. 161-3). Strawson argues that in ‘from the constant union the necessity arises’, *the necessity* is to be taken as meaning *the impression of necessity*, not the objective necessity that Strawson takes Hume as accepting.

T2 begins with a conditional, and the broader context is consistent with the hesitancy evident here. The epistemic reading is nevertheless well supported: human understanding *cannot discover* the connexions. T3 is more definite about there being distinct perceptions (species of distinct existences), and again provides support for a purely epistemic reading. The drift of E is more circuitous.

A survey of these passages, and of the whole texts of the *Treatise* and the *Enquiry*, reveals that Hume nowhere speaks of a *necessary connexion between distinct existences* in those words, but where he does use words near to this formula, the intent is always either clearly epistemic or else quite vague. In the passages cited the putative connexions are said to be *real* or *ultimate*, not *necessary*. It is interesting to observe how

infrequently the word 'logical' and related forms appear in the *Treatise* and the *Enquiry*. Mackie's version (see above) is indicative of a modern tendency to see necessity in broadly logical, perhaps *de dicto*, terms. (See also this, from Robert Nozick: 'Hume's lesson that there are no logical connections between events'; Nozick, 1981, p. 144.) But only equivocal precedent for such talk may be discovered in Hume.

So much for the predominantly negative evidence adduced in support of reading Hume epistemically. We shall examine the question of what ought to qualify for us as distinct existences in the first place, which seems to me to be the crux of the matter, when we consider Lewis's worldmaking again in some detail (next subsection: 2.3.3). But now let us concern ourselves with the epigraph to the present part, in which a hidden necessity, inaccessible to any 'human algebra', is postulated in the order of things. It is generally accepted that, roughly, Philo represents Hume's own opinions. And here Philo is supported by another party to the dialogue, Cleanthes (see the excerpt cited in 2.2.4), with dissent from no one. (We may set aside as irony the shallow show of dismissiveness surrounding the passage we are examining; or it may well be just the result of a reflex precaution on the author's part, given the sensitivities of the time – and this despite the fact that Hume reserved the *Dialogues* for posthumous publication.) A standard interpretation of the Philo passage, at least, would have it that Hume is arguing for a kind of *natural* necessity (which is Strawson's general interpretation of Hume; he cites the Philo passage in a footnote, but comments little on it, except to say that it is 'striking'). I submit, however, that the very idea of a *separate natural necessity* is foreign to Hume. Philo introduces talk of mathematical necessity, which we may take to be necessity *simpliciter* – necessity of the broadest, least circumscribed kind. And then:

Is it not probable, I ask, that the whole economy of the universe is conducted by a like necessity, [...]

The word *probable* in 18th-century usage may not be as forthright as the current sense would suggest. OED records an alternative older sense: 'Such as to approve or commend itself to the mind; worthy of acceptance or belief; [...]' (so that one could answer a question with 'probably yes, probably no'); but against this evidence is the countervailing fact that *probable* could still, at the time Hume wrote, mean 'capable of being proved; demonstrable, provable. Now rare.' (OED). The modern senses of this

still multivalent term perhaps have forces intermediate between these two, and it is the sense that seems to predominate in Hume, though some slight evidence can be found in the *Treatise* and the *Enquiry* for his sometimes intending each of the older senses.

As for *a like necessity*, this is frustratingly ambiguous: a necessity of equal strength and species; or rather some necessity – perhaps indeed a robust *natural* necessity – that is merely *akin to* the mathematical necessity just mentioned, in that it is concealed from us? But the collateral evidence of the Cleanthes passage (which the Philo passage follows quite closely upon) weighs in favour of the former interpretation, with the wording 'as great a contradiction as that twice two is five'. The immediate continuation in the Philo passage also confirms the unrestricted-necessity interpretation:

[...] though no human algebra can furnish a key which solves the difficulty? And instead of admiring the order of natural beings, may it not happen, that, could we penetrate into the intimate nature of bodies, we should clearly see why it was absolutely impossible they could ever admit of any other disposition?

Starkly opposed to this view of things are, of course, passages such as the following from the *Abstract* to the *Treatise*:

It is not any thing that reason sees in the cause, which make us *infer* the effect. Such an inference, were it possible, would amount to a demonstration, as being founded merely on the composition of ideas. But no inference from cause to effect amounts to a demonstration. Of which there is this evident proof. The mind can always conceive any effect to follow from any cause, and indeed any event to follow upon another: whatever we conceive is possible, at least in a metaphysical sense: but wherever a demonstration takes place the contrary is impossible, and implies a contradiction. There is no demonstration, therefore, for any conjunction of cause and effect. (*Abstract*, pp. 650-1)

At least here it seems to be affirmed that human knowledge *is* a guide to what is possible, in a certain uncertain sense: just what is the 'metaphysical sense' in which what we can conceive must be possible? It would perhaps be a mistake to read it as equivalent to any of the modern notions of metaphysical necessity – conceptions fraught with difficulty in their own right. (And let us recall OED's sense 1b for 'metaphysical', which was very much current in the mid-eighteenth century when Hume penned the above passage: 'Applied with more or less of reproach to reasoning, ideas, etc. which are considered over-subtle, or too abstract.') It may be taken as nearer to modern notions of *logical* possibility, or even of *epistemic* possibility (which in some of its versions would then unveil a tight circularity in Hume's suggestion) – notions that I

shall be arguing in Part 3 are unruly and often subtly conflated by neo-Spinozists like Chalmers.

Here is how we should all very much like to interrogate Hume: 'Well, could the *world* really be such as to allow anything to follow anything else? Or do you just say that we cannot rule on the matter, because of our in-principle epistemic limitations?' Perhaps he would have to think hard before answering, but he would have saved us much anguished debate if he had come clean.

I shall pursue this foray into Hume exegetics no further. In sum, and on balance, I find the Strawson line quite compelling, though I must reiterate the qualification that the epistemically inaccessible but *real* necessitation that Hume attributes to the course of events in the world is by no means shown to be anything like a modern separate natural necessitation, which is a restrictive qualification on broad necessity; and the evidence strongly suggests that he never thought in terms of such a qualification. Events follow other events in the ways they do as a matter of raw *undifferentiated* necessity, to whose determinations we have no epistemic access. (Or, if I am wrong about this, they follow as a matter of sheer brute contingent fact – which in the course of our discussion of real possible worlds will become even more extravagantly implausible than it was before we set out.) Efforts toward the principled differentiation of necessities into types are examined in Part 3.

Ultimately it needn't matter to us what Hume thought or did not think; I only want to deny that he fits comfortably into the role of precursor to those offering modern *profligate* worldmaking principles.

We examine the modern suite of restricted modalities in Part 3, so I choose not to dwell in detail on putative natural necessity here. But this is a suitable occasion to excavate something that I allege is very like the hidden necessitation Hume speculates about, with an example from Chalmers concerning the famous ideal gas law (sometimes confused with the more restricted Boyle's law; see for example Bas van Fraassen, 1989, p. 94; it is in fact a combination of Boyle's and Charles' laws). It is worth reproducing and analysing at length, since Chalmers refers to it often, and it becomes relevant again in our deliberations in Part 3:

[T]he pressure exerted by one mole of a gas systematically depends on its

temperature and volume according to the law $pV = KT$, where K is a constant (I pretend for the purposes of illustration that all gases are ideal gases). In the actual world, whenever there is a mole of gas at a given temperature and volume, its pressure will be determined: it is empirically impossible that two distinct moles of gas could have the same temperature and volume, but different pressure. It follows that the pressure of a gas supervenes on its temperature and volume in a certain sense. [...] But this supervenience is weaker than logical supervenience. It is logically possible that a mole of gas with a given temperature and volume might have a different pressure; imagine a world in which the gas constant K is larger or smaller, for example. Rather, it is just a fact about nature that there is this correlation. (Chalmers, 1996, p. 36)

There are two preliminaries. First, *no* gases are ideal gases (gases in which the size of the molecules is *incommensurably* less than the mean intermolecular distance, and in which there are no attractive forces whatsoever between the molecules), so the discussion is a pretence, or at least an idealisation, from the very start. That may not be a matter for concern, or it may indeed turn out to be worth pondering. Second, R , not K , is used as the standard notation for the constant in this form of the gas law. K (or k) is used when the quantity of gas involved is not fixed as one mole, but rather as one molecule. The standard constant K is the Boltzmann constant, and is definable as R/N_A , where N_A is Avogadro's number, discussed below. In what follows I uniformly use Chalmers' K , meaning by it what is standardly meant by R .

Note the first underlined portion. Chalmers is quite confident of a certain 'logical' possibility; but let us analyse. (In the following treatment, original with me so far as I have been able to discover, I rely upon general theory from Halliday *et al.*, 1993.)

A mole of a gas is plainly defined as Avogadro's number (about 6.023×10^{23}) of molecules of that gas, so it at least is immune to any modal modifications. Facts about *pressure* are by definition entirely determined by facts about force per unit area; force is in turn entirely expressible in terms of the mass and the acceleration of the particles involved in exerting the force, and of course acceleration is only a matter of rate of change of velocity, which itself is rate of change of position: accounted for by duration and length.

We should take *temperature* to be the equivalent of *mean kinetic energy of constituent molecules*, and as a rigid designator, what's more. Kripke's classic argument (1980; his view may differ now) that *heat* is to be understood this way works well to establish this result. I set aside E. Nagel's way of understanding the term, in pre-Kripkean fashion, as

non-rigid: 1961. pp. 342 ff. Lest it be thought that Chalmers intends 'temperature' to be taken some non-standard way, we should note a certain explicit statement somewhat later in his book: '[...] high level properties such as temperature and memory are all logically supervenient on the [micro]physical'; p. 178. On the other hand he also endorses a competing view: '[A]s has been pointed out by Wilson (1985) and [PS] Churchland (1986), many physical phenomena that are often taken to be paradigms of reducibility (e.g., temperature) are in fact multiply realizable'; p. 364, note 20. Let us remark, in any case, that the present discussion takes place firmly within the domain set by the standard kinetic theory of gases, so that it would be very far-fetched to introduce alternative realisations of temperature, pressure, etc.

Leaving out the detailed derivation of the formula given below, facts about temperature are therefore exhaustively analysable as facts about *mass*, *duration*, and *length*. Facts about volume are uncontroversially and necessarily constructed out of facts about length. So the fundamental spatiotemporal and physical variables entering into the constitution of the invariable relation pV/T are mass, duration, and length. Now, the dimensional characteristics of p , V , and T go like this:

p (as force / area):	mass / (duration ² × length)
V :	length ³
T (∝ mean kinetic energy):	[½ ×] mass × length ² / duration ²

(It is, of course, proper to ignore the '½' in the expression for temperature, since this is an analysis of the fundamental *variables* entering into the dimensional constitution of the relation; we shall do so from now on.) So the dimensional characteristics of the ratio p/T are derivable like this:

$$\begin{aligned}
 p/T: & \quad \frac{\text{mass / (duration}^2 \times \text{length)}}{\text{mass} \times \text{length}^2 / \text{duration}^2} \\
 & = 1 / \text{length}^3
 \end{aligned}$$

And for the whole relation:

$$\begin{aligned}
 pV/T: & \quad \text{length}^3 / \text{length}^3 \\
 & = 1
 \end{aligned}$$

This means that no matter what unrestrictedly (or 'logically') possible changes are assumed to occur in the nature of mass, length, and duration, or in the relations between these, the value of pV/T cannot be affected: the dimensional analysis of the relation pV/T shows that it reduces to a dimensionless, pure number. It is fixed for all relevant worlds: all the worlds for which one can speak of mass, length, and duration and mean anything like the same as we mean for this world.

Now, to settle more imaginatively what all this means, and to ascertain whether Chalmers' confident modal assertion that K could vary can be salvaged, here is a thought experiment. Holding V constant, a mole of any ideal gas has the ratio p/T fixed. Take a mole of (ideal) gas, confine it in a rigid can, and heat or cool the gas, monitoring the temperature and the pressure. The pressure would, as a matter of fact, change in direct proportion to the absolute temperature (that is, the temperature measured from absolute zero). But if it is supposed to be 'logically' possible for K to differ between possible worlds, we must suppose it to be possible for K to vary over time (under LR-type assumptions, shared by Chalmers, about how possible worlds are constructible out of spatiotemporal parts or stages of other possible worlds); so we must be able to imagine the pressure of the mole of gas we are heating in our rigid can remaining constant as we increase the temperature. Well, it might *seem* as if we can do just that. But what is happening to the *constituent* variables entering into the higher-level relation p/T (with V imagined to be fixed) as the temperature rises? Temperature is defined in the low-level terms of mass, duration, and length, and so is pressure, so something must be happening with the relations between *these* three variables. But what? What ought we to be imagining at the level of mass, duration, and length, if our imagining of the pressure remaining constant is to be clear and distinct, or complete and coherent? It turns out that there is *nothing* that could account for the observation we think we can coherently imagine, since the relation pV/T is dimensionless: it would be like coherently imagining a pure number changing its value, and this we patently cannot do. To get closer to the case in hand, it would be like trying to imagine the density of a consignment of matter changing, while its mass and volume stay constant. To get closer still, it would be like trying to imagine a consignment of matter having a temperature of lower than absolute zero. These quasi-imaginings are ruled out definitionally, and so are *logically impossible*, if this term is to mean anything useful. So with our relation

between temperature, volume, and pressure – the only important difference being that it is harder to demonstrate the logical impossibility, as I have here laboured to do.

(Why does the constant K occur at all in formulations of the ideal gas law? Simply because the relation is expressed in some units or other, and a constant is employed to express by means of an equation the invariable nexus between p , V , and T . Another, closely related, equation is often used, which is free of any such constant:

$$p_1 V_1 / T_1 = p_2 V_2 / T_2$$

where p_i , V_i , and T_i represent the pressure, volume, and temperature of any one consignment of ideal gas at time i . The formula – technically an *equation of state* – simply says that for any consignment of any ideal gas a certain nexus between these variables obtains, *no matter what*. And we have seen that, once we have settled the meanings of these variables, the nexus holds as a matter of mathematically underwritten ‘logical’ necessity – not mere natural necessity, whatever that could turn out to be, as Chalmers asserts.)

The moral of this disquisition concerning the idealised physics of gases is not just that Chalmers has made a slip or two: it is, much more importantly, that necessary statements can *dangerously easily* be mistaken for contingent statements, even in very simple cases, delivered in the rigorous and standardised language of physics. It *seemed* that all we had to do to establish a contingency was make a plausible declaration, this time concerning K . We have seen that this does not suffice. It may be helpful to come back to this example, as Chalmers himself does, as we progress to examine far less readily understood affairs involving the consequences of facts that are agreed to be within the domain of standard physical science, no matter how that is delimited. I hope that what I have undertaken in the last few pages, while not claimed as the work of a ‘skilful algebraist’, in Hume’s phrase, nevertheless reveals in a fragmentary way how more is necessitated in the universe than we might have been convinced was the case.

Generally the conclusion might be more to do with the illegitimacy of deriving metaphysical conclusions from epistemological premises, à la Descartes. But that is a topic of later sections, and of Part 3.

2.3.3 *Problems specific to Lewis’s worlds*

Continuing our critique of worldmaking moves, we turn now to some close analysis of specific kinds of worlds that Lewis himself proposes. Some are purportedly generated simply by LR, and some by appeal to alien properties as well. The broad kinds of worlds I examine are *spirit and strange-stuff worlds*, *odd space-and-time worlds*, *Quine worlds*, and *anomalous macro-object worlds*.

Spirit and strange-stuff worlds

For Lewis, there are ‘immaterial spirits’ at some worlds. (In 1986a, p. 73, he is not sure about this, but in conversation in 1997 he was firmly committed to the existence of a Berkeleian world, and so to an immaterial world with at least one spirit. He was also committed to the possibility of Cartesian egos.)

Neither Lewis nor anyone else has *shown* the notion of an immaterial but substantial spirit to have any sort of respectability. For one thing, as discussed at some length in Part 1, it may simply not make sense to speak of *anything* being immaterial, let alone complex causal structures subject to change. Paul Teller (1984) asks in what sense they would not be countable as physical (and so as reducible to something that we would *not* wish to call spirit). The mere fact that people have spoken for some centuries about spirits, and comparatively recently started speaking of them as immaterial, does not demonstrate that these are warranted in our actual ontology; why should we think they are any more warranted among our possibilia? If we suppose existence in time but not space to be essential to spirithood, the respectability of the notion of spirithood is dependent on that of separable space and time (see below), an unsuspected and arguably fatal complication for any commitment to spirits to have to support. What if we had had a long history of protestations of belief in more than one kind of spirit, say, a thousand – each kind made of a completely different brand of ectoplasm, each as different from all the others as it is from matter? (This might have been laid down in some central scripture of our dominant religion, for example.) As things stand we can, some have thought, at least *make sense* of ectoplasm, which has seemed to mean that ectoplasm is ‘logically’ possible. But a thousand completely different kinds? That smacks of excess! What unfathomable occult qualities could conceivably underwrite such a wealth of difference? No, we are more likely to rule such excess out as impossible. I suggest,

however, that this is only because it is a new notion, not kneaded into common sense over the centuries. It is arguable that a thousand kinds of immaterial stuff plus one kind of material stuff in a world is scarcely more outlandish a proposal than that there is just *one* kind of immaterial stuff, to inventory along with material stuff. The move away from monism being accepted, on the basis of presumed imaginability, by what principle could we stop at a mere dualism (cf. Jackson, 1994, p. 25)?

Dieter Birnbacher, in his defence of epiphenomenalism, is happy to endorse a somewhat related expansionary proposal regarding different species of epiphenomena:

If consciousness is an epiphenomenon of physical processes, why should there not be – under suitable conditions – epiphenomena of conscious processes, or perhaps even an ascending series of epiphenomena which are at each level dependent upon those at the level below, as mental events are dependent on upon events in the brain? The idea of epiphenomena of epiphenomena is anything but an absurd one. Epiphenomena are normally so defined as to have no causal powers, but in so defining them one is thinking of a presupposed material level. One could perfectly well understand the concept of an epiphenomenon in such a way that epiphenomena never have causal effects on the level 'below' them, but can still influence phenomena at the emergent level 'above'. (Birnbacher, 1988, p. 31)

This is odd indeed, considering that Birnbacher explicitly endorses Broad's three conditions essential to epiphenomenalism, the second of which he gives as: 'No mental event has any part in the causation of another mental event' (*ibid.*, p. 25). It is a very strange metaphysics that would allow epiphenomena not to cause anything at their *own* level, but somehow to cause 'higher-level' epiphenomena, themselves endowed with a similar mix of causal inefficacies and powers. Dennett makes a related point with characteristic forthrightness: 'Anything that has no effects whatever in the physical world surely has no effects on the function of anything' (1991, p. 402; but strictly this strong point may be secured for all situations only if necessary materialism is assumed). So the presumption is as absurd as imagining that matter had the power to produce epiphenomena, but not the power to affect other matter. Birnbacher's amusing fancy serves as yet another example of a careless imagining, too easily presumed coherent.

To turn things another way, why should we not speak of the possibility, mentioned earlier, of more than one fundamental kind of *material* stuff coexisting in a world? The words are available to us, and it seems that no contradiction is discoverable; but of course the real problem is that we have no actual-world or principle-driven model for

such a doxastic commitment: no *warrant*, as we shall systematically call it in 2.3.4.

Odd space-and-time worlds

Could space and time be more separate than they actually are, in any interesting and robust sense? This can be taken as a question about the dominant pre-Einsteinian view of space and time associated with Newtonian physics. Again we believe something possible because it was thought coherent (indeed, thought true) for long enough for it to pass as common sense. But it is worth remembering that one reason for Newton's view being superseded is that attempts to construe a Newtonian world furnished in a way complex enough to resemble the actual world have led to manifest contradictions. What better basis could there be for ruling out a class of worldmaking speculations?

Space and time may be no more coherently separable than the different spatial dimensions may be separable from each other. For neither sort of separation is there any model or warrant in the world as we observe it. And again we can press further than Lewis to explore just how acquiescent we should be in this sort of worldmaking manoeuvre: Could all dimensions be temporal, and none spatial? In worlds like ours in that they have at least three spatial dimensions and one temporal dimension, could something exist in space but not in time, as it is alleged that it is possible for something to exist in time but not in space? Closely relatedly, could something exist in *one* of our spatial dimensions but not in another? How sure are you that you can coherently imagine these things? And what if you can? Do your imaginations, by being coherent, represent genuinely possible worlds?

Yet it is remarkable how confidently philosophers' assertions have been on these matters, especially while ruling on the matter of spirits discussed above:

Very arguably, no object could exist in space without also existing in time, but the reverse is not true, since non-extended and non-located Cartesian egos are at least metaphysically possible. (Lowe, 1996, p. 111)

They are 'at least metaphysically possible'; does this mean that the world *could* have included Cartesian egos? If so, must we not concede that our world contains them, for all we know? Must we not concede that we cannot rule out that we are Cartesian egos? Lowe might use some 'natural' necessity to finesse out of such a conclusion; but we shall discuss the question of 'natural' necessity later, and especially in Part 3.

To return to the thread of argument, though: if we accept that space and time are necessarily bound together as spacetime, some more work may still be needed to show that this in turn necessitates that anything *existing* in time must also exist in space. Michael Lockwood (1989, pp. 71-78) argues, as had Russell before him (Russell, 1927a, pp. 384-5), that this is indeed the case. Against Wittgenstein and Ayer, who in their different ways argue for the impossibility of locating mental events in space, and so the impossibility of mental events being identical to brain events (which *must* exist in space), Lockwood adduces basic relativistic considerations (pp. 72-78). In short, if mental events occur separated by an interval of time, this is with respect to certain frames of reference. Their having relations of time at all between them necessitates that, at least considered in some other frame of reference, they are separated in space, or bear spatial relations to each other. To resist this inference would be to resist the fundamental relativistic assumptions themselves. And to do *that* you would need a powerful warrant!

Quine worlds

This is a very general and productive worldmaking move, introduced earlier in the context of *Lego-worlds*. Lewis justifies it this way:

This is just an appeal to recombination. But we are no longer applying it to smallish numbers of middle-sized things, horses or horns of [*sic*: read 'or'] heads. Instead, we are applying it to point-sized things, spacetime points themselves or perhaps point-sized bits of matter or of fields. Starting with point-sized things that are uncontroversially possible, perhaps because actual, we patch together duplicates of them in great number (continuum many, or more) to make an entire world. (1986a, pp. 90-91)

This really does look plausible – indeed, unassailable. But it rests on assumptions that are open to question. Granting for argument's sake that spacetime itself is indeed continuous (or *could* be, if it is not), it is worth while looking more closely at the notion of 'point-sized bits of matter' (let's barbarously call them *minimissima*, by extension of the concept of *minima naturalia* in scholastic and corpuscularian philosophy; see Peter Alexander, 1985, p. 66), since it is almost certain that to get the worlds he wants Lewis needs, despite the hesitation in the excerpt above, something more than just points of spacetime: he needs them variously occupied or unoccupied by something that is not itself mere spacetime. I shall take it that matter and fields can be treated the same as each other, here.

Minimissima, to be seeds for the prolific worldmaking Lewis applies them to, must be capable of independent existence, indifferently with or without other minimissima as worldmates, immediately neighbouring or otherwise. For example, a world having as sole contents a single minimissimum must be possible if minimissima are to be possible at all. So must worlds whose spacetime is entirely occupied by wall-to-wall minimissima – worlds saturated with *gunk*, to borrow Lewis's term for continuous, non-atomic matter (Lewis, 1991, p. 20). Lewis suggests that minimissima may be actual, for all we know. We are not equipped to rule them out, that's true. But does current physics give us any reason to believe that isolated minimissima (that is, having no immediate spatiotemporal neighbours) exist? Anecdotally from physicists, I think not. In the absence of such a direct warrant from physics, we may ask whether the existence of point *parts* of the extended particles that are known to exist are near enough to minimissima to provide a proper precedent for Lewis's compositions.

Do a left half and a right half of a proton (or of some other, more fundamental, spatially extended particle) count as distinct existences? If so, then by LR there is a world at which they exist spatiotemporally separated, or in which only one of them exists. But what warrant have we for the judgement that they are distinct existences, in the first place? Only that they are located differently, and that our experience of macroscopic phenomena is that macroscopic things can indeed be separated, because they *do* often come to be separated: parts of things go out of existence and leave other parts behind, and sometimes new parts grow. But is it legitimate to extend this expectation to the very small *constituents* of macroscopic reality? Protons are, after all, things of a radically different kind from macro-level things, all of which are *composed* of things like protons. We have no model or precedent for this assumption of uniformity across levels, and no independent principle by which to justify it. (Indeed, the behaviour of the very small constituents of the everyday actual world is considered utterly discordant with common sense. Quantum physics is labelled 'strange' almost as its natural epithet; but this can only mean that it is strange by reference to our experience of the large composite objects with which we have daily dealings.) The principle that there are no necessary connexions between distinct existences is under threat of tight circularity indeed: the very notion of a distinct existence has a strong claim to being *defined* in terms of the modal property of merely contingent connectedness with any other existences.

The 'left' and 'right' halves of a proton certainly exist (provided only that protons are 'extended in space, which we are told they are, and of course that we settle on some perspective from which to do our labelling). They can be thought about separately, they can be named separately, they can be situated on different sides of a boundary-line (or boundary-surface); it might even make sense to divide ownership of a proton, so that you own the left half and I the right half. In the case of macro-level objects, all of this would and should convince us (on standard accounts – to be impugned later) that the two halves are separable so that each half is capable of independent existence in some world. But remember: protons are not macroscopic objects, and macroscopic assumptions regarding discernibility have not been demonstrated to be straightforwardly applicable at other levels.

Now, if all of this can be said even about *halves* of protons, it certainly applies *a fortiori* to unextended, point-sized parts of protons at least equally well.

Beyond these considerations, we have some evidence from physics (JS Bell's experiments; see Leslie, 1989, p. 87) that even distinct particles (or at least, distributed particle-like phenomena that we would normally want to count as distinct) come in pairs, though they are spatiotemporally distant. They arise together, and have surprisingly tightly linked fates thereafter. Again, such spatiotemporal separation in the macro-level world is ample warrant for talk of distinct existences, with a wealth of experiential precedent, but what justification is there for applying such a precedent in the case of these distant-twin particles? They are of a completely different order, for which we lack experiential evidence, except some scientific evidence that they do *not* behave as distinct existences. The physical law seeming to yoke their fates together may, for all we have been able to ascertain, have a foundation in some utterly inviolable but hidden principle of necessitation à la Philo, just as there were additional hidden but relatively easily discovered restrictions in the case of the ideal gas law.

In the case of whole atoms of any of the chemical elements we have ample warrant, from the very concepts of *atom* and *element*, as these have evolved, to say that they could exist separately. If they could not, after all, what would be the point of calling them *atoms* – and discrete tokens of *elements*? The very idea carries with it a certain basicness and independence. Lewis writes:

If you said that wherever *carbon* is instantiated, *bromine* must necessarily be instantiated next to it, that would make good enough sense as a matter of nomological necessity. There is no such law of nature, but there could have been. But suppose you said that it was a matter of necessity *simpliciter* – absolute 'logical' or 'metaphysical' necessity. Then what you say is not only false: it is entirely unintelligible how it could be true. Why couldn't anything over here coexist with anything else over there, and in particular why couldn't the presence of an instance of *carbon* over here coexist with the absence of any instance of *bromine* over there? (1986b, p. 41)

And this seems right enough, except when we consider the underlined assertion. If we were discussing the sides of proton, it would be *intelligible* how it could be true, though spatial extension (or for that matter temporal extension) may not, *pace* Lewis, strictly entail separability. We could, quite *intelligibly*, assert that with carbon and bromine there had been a dreadful mistake: they are not elements after all, and there are no atoms of either of them; we had been misconstruing the evidence. There is only one entity here to speak of – a unique kind of extended atom of a type best called, perhaps, *bromocarbon*, and it is even in principle not separable into atoms of any other chemical element, but only into the usual protons, neutrons, and electrons. Wildly counter to Mendeleevian orthodoxy? False? Impossible? Yes, all of these perhaps. Not *unintelligible*, though. But the parallel story with proton-halves? For all we know, the halves of a proton cannot exist independently – as a matter of the broadest necessity. There is no warrant for their separability parallel to the warrant we have in chemistry concerning atoms and elements.

It is also worth asking in this context whether Lewis's suggestion, raised earlier, that there can be 'occupancy by different kinds of matter' (1986a, p. 90) is backed with sufficient warrant. The point is related to my observations above concerning stipulatedly different sorts of substance, whether material or immaterial. We simply don't know what should count as different fundamental kinds of matter. All we have as models are largish, spatially extended particles like protons, much smaller electrons, and the rest of the zoo of non-point particles so far provisionally inventoried by physics. We simply do not know how many *kinds* of more basic matter constitute these extended structures, or what truly fundamental non-structural properties, if any, could serve to distinguish different species of minimissima, assuming *their* possibility. Nor can we determine, from the available actual precedents, whether any different fundamental types of matter

could exist at the same spatiotemporal point (in the same world); Lewis himself seems to waver on this (see well-informed discussion in Pruss, 2001a, in respect of named physical particles). If there are infinitely many kinds of minimissima, there must be infinitely many single-minimissimum worlds, one (at least) corresponding to each kind. But it is especially difficult to make sense of such a class of descriptively indistinguishable worlds.

Anomalous macro-object worlds

Pulling back the focus from minimissima to ordinary large objects, and from the Mendeleevian to the Mendelian, we may still find room for doubt about Lewis's worldmaking principles. Before Mendel, it was thought biologically possible that offspring should express inherited characteristics mixed in various proportions from both parents. Mendel's great discovery was that this is *not* biologically possible: the norm was for offspring to have all, none, or some inexplicably constrained mix of inheritable characteristics. Whatever unknown micro-level mechanisms were at work, they issued in a surprisingly coarse *graininess* in genetic inheritance. We now know also that drosophila with certain mutations may have more or fewer whole segments to their bodies, but not just any imaginable *continuous* reshaping. And, to give a more general case, we are convinced that the total anatomical and functional complexity in a mature human body is drastically more than the total biological-level complexity in a zygote; we think, therefore, that the whole ontogeny of the mature organism is not infinitely variable, but grainily limited by the initial biological state of the zygote from which the mature organism develops (though of course ontogeny is also essentially environmentally influenced). In short, a biological version of LR is unlikely to be true; analogously, why should we think that just any *notionally* constructible large object – world or part of world – is possible *simpliciter*? There may be some imaginable time-slice of universes that are possible given some history, but not given some *other* history. There may be other time-slices that are not possible by *any* account of the history of a universe. We have no well-founded reason to expunge these caveats from the modal charter. We were wrong about biological variability and the limits of biological possibility; we may be analogously wrong about broad modal variability and the limits of possibility *simpliciter*. Even Lewis's eminently commonsensical talking donkeys

ought now to be suspect: and so should mermaids and centaurs.

At least to fix ideas, let us see how Descartes and Lucretius are more circumspect, in their diverse and fanciful ways:

For, in truth, painters themselves, even when they study to represent sirens and satyrs by forms the most fantastic and extraordinary, cannot bestow upon them natures absolutely new, but can only make a certain medley of the members of different animals; or if they chance to imagine something so novel that nothing at all similar has ever been seen before, and such as is, therefore, purely fictitious and absolutely false, it is at least certain that the colours of which this is composed are real. And on the same principle, although these general objects, viz. [a body], eyes, a head, hands, and the like, be imaginary, we are nevertheless absolutely necessitated to admit the reality at least of some other objects still more simple and universal than these, of which, just as of certain real colours, all those images of things, whether true and real, or false and fantastic, that are found in our consciousness are formed. (Descartes, *Meditations*, I)

If we take this imaginative activity Descartes describes as a kind of generation of Lego-worlds, he seems to allow that it is legitimate because it ultimately depends on recombination of items in our experience – a restriction to actual properties. Contrast the more severe Lucretius:

But Centaurs ne'er have been, nor can there be
Creatures of twofold stock and double frame,
Compact of members alien in kind,
Yet formed with equal function, equal force
In every bodily part – a fact thou mayst,
 However dull thy wits, well learn from this:
 The horse, when his three years have rolled away,
 Flowers in his prime of vigour; but the boy
 Not so, for oft even then he gropes in sleep
 After the milky nipples of the breasts,
 An infant still. And later, when at last
 The lusty powers of horses and stout limbs,
 Now weak through lapsing life, do fail with age,
 Lo, only then doth youth with flowering years
 Begin for boys, and clothe their ruddy cheeks
 With the soft down. [...] (Lucretius, *De Rerum Natura*, Book 5)

In short, Lewis and other advocates of promiscuous recombination principles are typically *Cartesian* in their approach to macro-level objects, but a kind of higher *Lucretianism*, informed by the broader if so far unknown constraints on ontogeny suggested above, founded in necessity *simpliciter*, may have at least as much to recommend it.

So much for a critical survey of Lewisian worldmaking. Lewis invites trouble from Forrest–Armstrong type objections by this very profligacy; and rather than there being available some quick fix of a blanket, non-committal sort (see 2.2.3) in response to those general objections from set-theoretical cardinalities and the like, perhaps there is something deep in the very notion of wholesale recombination principles that is badly motivated. The Lewisian alliance of Lego-world and logo-world principles yields worlds that are open to question in numerous ways, some of which are rehearsed at length above. I have suggested in passing that there needs to be sufficient positive *warrant* for modal assertions; the standard purely logical and recombinatorial warrants have only negative force, and have been found wanting by a considerable weight of circumstantial evidence. We now explore this need for warrant in greater depth.

2.3.4 *Warrant for modal beliefs – the ‘antilogical’ turn*

In 2.3.1 we looked at ostension of actual entities or properties as providing a *prima facie* warrant for modal beliefs. The merely possible entities we construct (talking donkeys, unicorns) are to be formed from uncontroversial elements, possible because actual. Beyond this, *conceivability* has seemed to do the trick. It is worth remarking, though, that even manifest actuality of objects or phenomena has not always been sufficient to warrant general acceptance – so seriously entrenched and insidious are our intuitions about the way reality must be. This is evidence that we are by doxastic nature radical conservatives. How could Lewis hope to convince the majority, who are not even persuaded of the possibility of certain *actual* phenomena? Despite the evidence in actuality, and the fact that at least later physics makes use of the notion seemingly without effort, it was almost universally held in the seventeenth and early eighteenth centuries that action at a distance was *inconceivable*:

The next thing to be consider'd is, how *Bodies operate* one upon another, and that is manifestly by impulse, and nothing else. It being impossible to conceive, that Body should operate on what it does not touch, (which is all one as to imagine it can operate where it is not) or when it does touch, operate any other way than by Motion. (Locke, *Essay*, first edition, 1690, 2.8.11)

Interestingly, Locke later amends somewhat his own intuition regarding the uses of, if not the limitations on, conceivability:

It is true, I say, That bodies operate by impulse, and nothing else. And so I thought when I writ it, and can yet conceive no other way of their operation. But I am since convinced by the judicious Mr. Newton's incomparable book, that it is too bold a presumption to limit God's power in this point by my narrow conceptions. The gravitation of matter towards matter, by ways unconceivable to me, is not only a demonstration that God can, if he pleases, put into bodies powers, and ways of operation, above what can be derived from our idea of body, or can be explained by what we know of matter; but also an unquestionable, and every where visible, instance, that he has done so. And, therefore, in the next edition of my book, I will take care to have that passage rectified. (*Letters to the Bishop of Worcester*, p. 462)

Equally interestingly, Locke is not inclined toward any rumination concerning the previously unsuspected fallibility of such judgements, nor concerning what properties and powers might be essential to matter. It may be fair to consider Locke's protest that he 'can yet conceive no other way of their operation' as a neat escape from having to retract even more fully. He had clearly meant, as his original exact wording and his stated intention regarding the proposed next edition show, to say that matter *could only* act by impulse. It is a matter for further deliberation, not to be entered into here, to determine whether the revised statement, taken by itself, sufficiently or lucidly moderates the strong claim made in the first edition:

The next thing to be consider'd is, how *Bodies* produce *Ideas* in us, and that is manifestly by impulse, the only way which we can conceive bodies to operate in. (Locke, *Essay*, fourth edition, 1700, 2.8.11)

In any case, we have at least tentatively allowed the recombination of actual macroscopic objects, once they were indeed accepted as real. We have ample empirical examples of this kind of recombination from which to generalise a warrant in other unobserved or non-actual cases, though perhaps we should add that we need also an adequate *ontogenetic account* to confirm such a warrant. (See discussion of anomalous macro-object worlds in 2.3.3; perhaps a fall-back ontogenetic account can always be found by appeal to the possible but enormously unlikely random emission of large structured objects of arbitrary structure from black holes by Hawking radiation; Stephen Hawking, 1988, p. 110. But this puts enormous constraints on the kinds of contexts in which anomalous macro-objects could feature, and guarantees nothing about their extension in time – or, should we say more generally, the possibility of the sorts of continuant objects of which they may be conceived to be time-segments.) There will be more to say about this matter of ontogeny.

Lewis is unsatisfied with the limited warrant given by actual macroscopic objects and the well-founded presumption that they may be recombined. He goes much further, but as we have seen it is very likely that he goes too far, as evidenced by Forrest–Armstrong type objections and the particular problems I have detailed in 2.3.3. In 2.5 we shall consider a further class of reasons to constrain worldmaking; but all of these are general objections, to which Lewis can only react with generalities such as the ‘size-and-shape-permitting’ restriction mentioned in 2.2.3. Lewis claims as possible an enormously rich variety of configurations for worlds, and then challenges everyone else to find the logical consideration that will disqualify them. He *knows* that some will have to be disqualified, because of the set-theoretic objections that he concedes are compelling.

Consider again the example of Lewis’s *spirit worlds*. The only sample we have of a world by which we can form our intuitions as to how worlds can be – the actual world – gives no warrant for such a flight of worldmaking. (Alternatively, if we do have some model – say because matter and anti-matter, or positive, negative, and neutral particles, constitute different basic *kinds* of matter – why are we all so happy to count current materialism as a sort of monism? A point similar to this one is made by Galen Strawson, who suggests that some of the terminology concerning dualism and monism is purely conventional; see his 1994, p. 44. See also extended discussion in our Part 1.) No more have we such a warrant when equally implausible folk notions like spirithood *happen* to have evolved, and *happen* to have disposed us to regard certain states of affairs as possible, though they are utterly unexampled in the actual world.

Perhaps the onus ought rather to be on the proposer of possible worlds to show positively *how* the candidate world’s place in our picture of total reality is warranted. We might argue that it is insufficient to give a verbal description and to hold that, by default, it is to be presumed that there is a possibility corresponding to it. Unfortunately, procedures for arguing positively for possibilities seem limited. The two warrants derived so far deliver very little. They are both derived from experience, and it appears to be impossible to find any that are *not* founded in experience, or even any third warrant in the domain of experience beyond these two: actuality of entities and cautious generalisation from experienced macroscopic recombination. These do not depend on any deep understanding or insight into reality: it is all too easy to secure at least some

such possibilities deictically, even though we know not whereof we speak. (What is matter, that we can speak of it? Only... *this*! And we can rarely say much beyond that.) But in principle nothing like that is available to us beyond the domain of the actual. We *would* need an unattainable understanding or insight to secure our warrant for the existence of specific alien basic entities or specific basic properties. An ‘ambit claim’ of *prima facie* logical consistency is not, despite the modern temper of thought, a warrant for modal assertions; we need an ‘*antilogical*’ turn to match and abet Heil and Martin’s *ontological* turn (see 2.1.1).

And yet, for all our stiff-upper-lipped ontological seriousness, and determination to avoid mere linguistic-logical inquiry into modal truth, we must run up against the very same brick wall that prompted talk of ‘logical’ modalities in the first place. We have, as a matter of necessity, only one world to scrutinise. How to extrapolate from, or even make shrewd guesses from, a sample of one?

Well, oddly enough, though it is well-established lore in mainstream statistics that one can do very little inference on the strength of just one observation (to estimate population parameters *mean* and *standard deviation* – usual goals of statistical inference), this turns to be wrong. Finite, usable (though wide) confidence intervals can be derived for a normal – or even just a unimodal – distribution, solely from a single observation. It is only when we are restricted to using the usual time-honoured statistical tools that no such inferences can be made. (Anyone mildly literate in statistics ‘ought’ to be either astonished or incredulous when confronted with this claim; but such results, available in print since the 1960s, are reviewed for all to see in, for example, D Edelman; 1990. The claim must be hedged a little, but is well enough supported in the case of standard deviation, especially; at least there is enough to unsettle firmly held beliefs in the way mentioned.) The fact is that ordinary statistical inference theory developed on the assumption that samples of $n \geq 2$ are to be considered available, and moderately schooled intuitions tell us that such a sample size is required for any sane induction. So it is, I suggest, when we broaden the question to the nature of worlds themselves – not just the matter of single variables. We imagine that no sound inferences can be made on grounds other than the purely logical. But as I shall argue, we may be surprisingly mistaken about that, just as, analogously, conventional

statistical wisdom is demonstrably in error in the way mentioned above.

In the next two sections, 2.4 and 2.5, I consider two principles that harness a sort of statistical transcendental reasoning to progress to non-trivial and well-founded beliefs (though not deductively secured knowledge) about the range of possible worlds, by means other than the search for contradictions. The mutually countervailing tendencies promoted by the two ways of reasoning I discuss amount to new warrants for modal doxastics, and will jointly issue in what I call a *moderate necessitarianism*, to be set out in the final section of this part, 2.6. With that we may hope to navigate a course between the tortuously tight whirlpool of traditional Spinozism and the limitless ocean of spurious possibilia toward which Lewis pilots us. Nozick (1981, p. 651, note 4) once observed that to his knowledge no one had yet 'tried to answer a philosophical question by offering a statistical explanation within philosophy', excepting possibly his own work in his chapter headed 'Why is there something rather than nothing?' I think that in the two decades since Nozick wrote that there has indeed been more such work; and certainly I shall be offering some in what follows, informed by the same spirit as parts of Nozick's chapter, as I shall indicate where appropriate.

2.4 ANTHROPIC REASONING AND MODALITY

2.4.1 Transcendental reasoning: the transcendental and the rational

There is an analogy in sorts of meta-mathematical reasoning for the kind of inquiry we are engaged in here, which may at least be useful in the formation of sounder intuitions than the inadequate ones we start out with when we grapple with many-world hypotheses such as MR (and indeed ersatzism).

In April 1994 a hoax was perpetrated on and in various newspapers to the effect that the most famous of the transcendental numbers, π , was a rational number after all:

It's official: pi, the ratio of the circumference of a circle to its diameter, ends at the 2,075,932,542,102nd decimal place, according to a mathematical theory laboratory in the United States. Researchers were left open-mouthed last month when a computer given the onerous task of finding pi's last decimal place suddenly start[ed] printing 'millions of zeros' at the end of the 'irrational' number. (*The Australian*, 12 April 1994, p. 22)

The Australian soon published an ignominious retraction. They had been duped, fair and square: it has been firmly established that π is transcendental (and *a fortiori* irrational) at least since the publication of Lindemann's proof of 1882. But the hoax prompts interesting speculations. If the mathematical world *had* still been without such a proof when the hoax hit, what reasons could we nevertheless have to think it was fake news? Here are four such reasons (some others are omitted).

First, there is something fishy about the computer finding an unending series of *zeros* as the continuation of the decimal expansion. Why not *twos*, or *nines*? Such continuations would equally guarantee that π is rational; the choice of zero looks just a little too contrived.

Second, there is something far more fishy about there being any single repeated digit as the continuation, since a repeated *sequence* of digits, of any finite length, would do just as well to establish π as rational. A length of just *one* digit (zero, as it happened) for the sequence, out of the infinitely many lengths that we might have judged equally likely to turn up, any one of which would have done just as well, looks overwhelmingly likely to be contrived by some playful and moderately numerate journalist. (This is biased towards our accidentally favoured decimal system awfully suspiciously, too, since

expansions in number systems with bases other than ten would be much less elegant.)

Third, it is remarkable that π should jump from being thought *transcendental* to being thought *rational*, without there being any consideration given to the more cautious claim that it is *algebraic*. One would have thought it very likely that a proof would be found of this before the alleged demonstration of π 's rationality. (I pass over quibbles based on the equicardinality of the algebraics and the rationals.)

Fourth, if π had turned out (*per impossibile*) to be rational, what an amazingly unpredictable fact it would be that the decimal expansion of it should settle into a cyclic repetition so ludicrously *early*. A mere two trillion or so decimal places! If a number were to be selected from the infinite stock of rational numbers at random (a curious notion in its own right), the probability of its being so easily within the capabilities of humans to *name* it, or even for them to name the order of magnitude of the decimal place at which it starts its cyclic repetition, is for all practical purposes equal to zero.

If it is impossible for us to imagine being in the epistemic position of not knowing that π has been securely established as transcendental, consider Euler's constant, or γ , which is defined like this:

$$\gamma = \lim_{n \rightarrow \infty} 1 + 1/2 + 1/3 + \dots + 1/n - \log_e n.$$

This important constant, approximately equal to 0.5772, has never been proven irrational, let alone transcendental. Now, how should we bet about its decimal expansion? Never cycling? Starting to cycle at or before the $10^{1,000,000,000,000,000}$ th decimal place? Or starting to cycle after the $10^{1,000,000,000,000,000}$ th decimal place? I think the worst of these to bet on would be the second option, since that would mean that γ is rational, yet among the infinitesimally small proportion of rational numbers that we are capable of naming by anything like standard means. (In fact I would bet very confidently on the first option, for reasons that we need not go into; and again I pass over problems with cardinalities, which will be addressed later on in Part 2.)

The moral of this extended and somewhat cavalier excursion into mathematical epistemology is that even in the absence of deductively secured *knowledge*, we may have surprisingly many sound ways of establishing worthwhile suspicions, or bases for bets, concerning many hypotheses presented to us. The reasoning may be called

transcendental, and specifically *anthropic*, because we draw conclusions based on the likelihood of certain evidence given our doxastic and evidential position, rather than merely on the internal characteristics of the evidence itself. Another application of pretty well the same principle goes like this:

Is it possible for the values of the broad parameters found by physicists to determine the nature of our universe to be different from their actual, observed values? How could we know? Not by a mere declaration and a search for contradictions (as I have argued in general terms); and we have, perforce, only one sample value for, say, G (the universal gravitational constant) on which to base any standard inductive reasoning. But consider: the case is rather like the π hoax – something *looks* suspicious. Given that we would not have been here to observe the world if G had differed slightly from its extremely unlikely and 'considerate' value, and given that nothing seems likely to be able to guarantee that it *should* have this value, just as nothing in mathematics would 'determine' π or γ to have values well-adapted to humans, it seems a good bet that G takes other values as well, and that we exist where we *can* exist – at a world with G adapted to our existence.

So it is in general worth noting, I think, that speculation about other worlds is not as bereft of parallels in other domains of inquiry as we might have feared. Note also in passing that the 'selection' of a 'random' real value 'by' π or γ – or by any other 'device' – is deeply problematic; compared to it, the 'random' choice of 'our' world from an infinite range of worlds may be a trifling affair – at least on the assumption of MR.

The well-known but often misunderstood and abused anthropic principle, which I have introduced in an unorthodox fashion just now, is a kind of transcendental approach to founding beliefs about the variety and reach of reality. It may reasonably be extended beyond our world to establishing beliefs about other parts of an ensemble of worlds, and the place of our world in such an ensemble. Such an extension is often conducted in the context of comparing many-world hypotheses with theological design hypotheses. But we shall ignore theistic hypotheses; we shall proceed on the assumption that if there were a deity outside of the closed causal domain we call the *world* that somehow nevertheless caused this world or anything in it, all bets would be off. At least, *our* bets

would be off, and *our* inquiry would be futile. I should not be taken as suggesting that this assumption is deeply warranted, but in the present work only so much can be done. To proceed on any assumption other than the one I propose would encumber our endeavour needlessly, and extend this piece to unmanageable length.

With that settled, we can take anthropic reasoning to have a kind of *Ptolemaic* tendency, in the following completely non-pejorative sense: without anthropic reasoning we might assume that all regions of total reality (spatial, temporal, spatiotemporal, or modal) are much the same. For its part this assumption is often called, quite aptly, the *Copernican principle* (a *principle of mediocrity* that is also sometimes called, perhaps obtusely, the *cosmological principle*; noted by Leslie, 1989, p. 131), since the result that our place in the cosmos is not special was the general outcome of Copernicus's heliocentric cosmogony, supplanting the Ptolemaic geocentric scheme. Anthropic reasoning tends against such Copernican results, as can be seen by Carter's famous early formulation of the so-called *weak anthropic principle* (WAP):

[...] our location in the universe is *necessarily* privileged to the extent of being compatible with our existence as observers. (Carter, 1974 p. 293)

Our location *is* special – a broadly Ptolemaic conclusion. But immediately in this brief excerpt from Carter it is possible to discern two of the seeds of the abundantly proliferating confusion in the quarter century since he wrote.

First, the ambiguity of *necessarily*. What Carter clearly intended is that, *given* that we are here, it follows that this location is such as to allow our presence. It is not that some location had certain properties as a matter of some kind of *prior* necessity.

Second, the ambiguity of *universe*. The general idea would have been far more perspicuous if Carter had used the term *total reality*, or something equally non-committal. If he had done so, it may not have been necessary to formulate the notorious so-called *strong anthropic principle* (SAP):

[...] the Universe (and hence the fundamental parameters on which it depends) must be such as to admit the creation of observers within it at some stage. (*ibid.*, p. 294)

SAP has been subject to even more egregious misinterpretation than WAP, largely on the strength of the uncertainty attaching to its modal *must*. The teleology often read into

Carter's formulation has spawned a brood of ever more outlandish proposals, culminating in vatic deliverances like the *final* anthropic principle:

Intelligent information-processing must come into existence in the Universe, and, once it comes into existence, it will never die out. (Barrow and Tipler, p. 23)

I wish to stress, then, that I shall be using *only* a widened version of WAP; and I shall *only* be using it non-teleologically, with the corollary that I shall be ignoring design interpretations, as mentioned above. My preferred version of WAP:

WAP: Reflective observers can only find themselves located in those parts of total reality whose ensemble of properties is sufficiently conducive to the existence of reflective observers.

I say 'reflective observers' because that is what we relevantly are – the observers need not be *human* persons to engage in 'anthropic' reasoning, and they need not be 'carbon-based life' (*pace* Barrow and Tipler in their strangely careless version of WAP; 1986, p. 16). I say 'can only find themselves' to obviate the attribution of any sort of independent necessity to the locations at which we find ourselves. And I say 'sufficiently conducive' because the permissiveness in question is a matter of degree: some locations may be less and some more conducive to the existence of reflective observers than for example our own location is.

I have mentioned the relations between the anthropic principle and a Ptolemaic tendency in our understanding of reality and our place in it. It has to be admitted that these relations are complex. JR Gott (1993) gives the anthropic principle a different application from the usual, and derives a *Copernican* anthropic principle according to which one should think that one's position among this world's total population of observers is no special position (cf. the Doomsday Argument; Leslie, 1996). This alternative is not relevant to our purposes, and I shall stay resolutely with a *primarily* Ptolemaic understanding of our version of WAP. There are important *secondary* consequences of WAP-style reasoning to note that are best characterised as Copernican, but I wish to apply the property of being *primarily* Copernican to a principle that opposes the basic tendency of WAP, in the next section (2.5) – what I shall call the *paranthropic* principle. Here is a passage to illustrate the need for such clarifications and such disavowals of pejorative intent, and incidentally to illustrate stubbornly recurring readings of anthropic principles as teleological. On the heels of a

verificationist objection to SAP's support of many unobservable universes:

A second objection to the strong anthropic principle is that it runs against the tide of the whole history of science. We have developed from the geocentric cosmologies of Ptolemy and his forebears, through the heliocentric cosmology of Copernicus and Galileo, to the modern picture in which the earth is a medium-sized planet orbiting around an average star in the outer suburbs of an ordinary spiral galaxy, which is itself only one of about a million million galaxies in the observable universe. Yet the strong anthropic principle would claim that this whole vast construction exist simply for our sake. This is very hard to believe. (Hawking, 1988, pp. 132-3)

Indeed. Let us not believe it. The confusion stems from the least conservative reading of Carter's original vague formulation, grown into physical-science folklore. (Roger Penrose also, after giving an impeccably non-teleological sketch of WAP and SAP, inexplicably adds that with either of these 'one might try to show that consciousness was *inevitable* [...]', and then expresses the opinion that such an argument is 'technically' correct; Penrose, 1990, pp. 561-2.) As for undoing all the good campaign-work of Copernicus and company, Hawking need have no fear! Total reality will remain what it is, and our place in it will still seem diminutive.

On the other hand, many have criticised the anthropic principle in its moderate guises as trivial. It may well be that, but so are many valuable principles, arguably enough – including perhaps many of central importance in mathematics. If it is only a way of reminding us of something so obvious that it is frequently neglected, it certainly is useful.

In the context of our discussion of modal metaphysics and modal doxastics, WAP is a broadly *anti-Spinozist* principle. As Lewis himself notes in one of his few (and cursory) treatments of the principle:

A modal realist can appeal to the 'anthropic principle': we ought to find it not at all remarkable that the physical constants and boundary conditions turn out to permit the evolution of intelligent life, no matter how exceptional the required values may be. For there are many worlds, with all different values of the constant and boundary conditions. [...] Of course, any inhabitant of a world will find that his world is a habitable one. (Lewis, 1986a, p. 132)

As we shall see the relations between WAP, MR, and the ersatzist alternative are a good deal more complex than this suggests, but it does give the essential idea. Lewis goes on to point out that you do not have to be a modal realist to invoke WAP: it is primarily

used as a support for many-world quantum-theoretic hypotheses, and non-quantum cosmological many-world conjectures. In fact most treatments of the principle make only passing mention of MR, or are entirely innocent of it. But the basic mechanics of the argument in application to MR are the same as in other applications.

Lewis then has a little to say about the nature of *explanation*, and whether, given MR, the anthropic principle has anything at all to explain. But he does *not* go on to give typical Lewisian cardinality-based objections to the principle's application in supporting MR. We shall address such objections below. It is, all may agree, not possible to deduce the existence or natures of other worlds than the actual with apodeictic certainty (nor to form our modal beliefs reliably by any standard inductive means), any more than we are so far able to determine the status of Euler's γ discussed above. But to call for such certainty is to go beyond the reasonable standard we ought to require for such beliefs.

A cautionary note about 'rationality'

There are many accounts of how rational beliefs are formed or ought to be formed. For the remainder of this work I shall be using inductive arguments in quite unfamiliar contexts, so I should settle first what can be relatively easily settled. I shall be speaking as if rational beliefs are to be arrived at completely independently of desires, preferences, loose principles of informal 'ecologically valid' reasoning that we use *faute de mieux*, and generally all of those intuitions of doubtful provenance, mentioned earlier. If anyone disputes this way of speaking, I want no quarrel with them. I would be happy for anyone else to call what I call *rationality* by some other name instead. I shall be implicitly assuming a set of deliberative norms for arriving at a maximal stock of true beliefs, supervenient on the available evidence, and I shall be assuming that it is proper to deem such a set constitutive of *rationality*. In my understanding of rationality, it can only ever be rational to give exactly 0.001 credence to the proposition 'I hold the winning ticket of one of 1,000 tickets issued to 1,000 ticket-holders in this so-far-undrawn fair lottery' (disgusting gerrymandered cases aside!); or in the Wittgenstein anecdote presented earlier, and assuming that the only evidence is what is given in the phenomena, it is irrational to reject the hypothesis that the earth moves. This will be important to remember when we consider use of, say, Occam's razor reducing more or less to a way of expressing a mere *preference* for simplicity. In the contexts we are

working in, such principles may have good rational application for a world construed as the sum of all there is, but not so straightforwardly for some world (even our world) as part of the putative ensemble of worlds – or total reality.

One philosopher highly proficient in theory of rationality, decision theory, and Bayesianism confided to me: 'The truth is, it makes me feel a little strange to think of there being empirical evidence for modal realism.' I'm sure that most who are anything like as competent in the relevant areas as this philosopher will be quite uncomfortable with the idea: but the evidence *will* be presented! And all this is just a 'natural' extension of our 'naturalist' presumptions, when we are at our most ontologically serious.

2.4.2 *Ersatzism and 'equiprobability' of worlds*

In what follows we ought for the most part to be assuming that either MR or the extreme Spinozist alternative that there is only one world (a single causal domain) holds, and judging how each fares when confronted with anthropic reasoning conjoined with the fact that the observed world appears *fine-tuned* with respect to the basic parameters discovered by physical science. But it will be useful as a preliminary to examine how ersatzism fits into the story, since we shall want an account of 'equiprobability' for all abstract worlds of possessing the property *actuality* and its associated concreteness (possessible by only one world *sub specie aeternitatis*, for ersatzism) whenever we step aside from notional acceptance of MR.

(We shall sometimes drop the scare quotes for this species of 'equiprobability', despite its being – as I believe, at least – a categorially improper concept to apply to worlds and their prospective actuality. This impropriety is obvious and tantamount to incoherence in the MR perspective. But ersatzism certainly needs some notion of, or akin to, such a probability applied to 'candidate' worlds' being actual. In the end, that is the ersatzists' problem, not mine. But I still have to find suitable words to capture what can be captured of the opposing view!)

So for MR there can be no good question concerning the prior probability of any world having the *property* of actuality, any more than there can be a good question concerning the prior probability of any given person having the property of being able to use 'I'

legitimately to refer to herself. But such a question of probabilities applying to worlds can meaningfully be raised – or so many seem to have assumed, at any rate – for ersatzists. I shall argue that the question is answerable, to the limited extent of its being a reasonable question, only by the suggestion that every ersatz world ‘starts out’ equally likely to be the actual world. To do so I shall examine van Inwagen’s treatment of the question ‘Why is there anything at all?’ in his article bearing that question as its title (van Inwagen, 1996).

Van Inwagen considers this ‘most profound and difficult of questions’ explicitly on the twin assumptions that MR is wrong, and that this investigation is one of the few domains in which it makes a difference which of MR and some form of ersatzism is right (p. 95, note 1). (He does not spell out how things would go if we were to work instead with the assumption that MR is right, but the answer is straightforward enough. By a proper reading of MR there is, first of all, in an absolute sense something rather than nothing because it is a fact underwritten by the broadest necessity – our Ur-necessity – that all the worlds exist; or, as we have been putting it, that total reality be as it is. And secondly, in a world-relativised sense there is something rather than nothing because every world, being an isolated concrete particular, must be *something*.)

Van Inwagen’s main argument is to the conclusion that, while there is a possible state of affairs whereby there is nothing at all (no concrete entity), the prior probability of this state of affairs obtaining is 0. Here is that main argument, beginning with four premises:

- (1) There are some beings;
- (2) If there is more than one possible world, there are infinitely many;
- (3) There is at most one possible world in which there are no beings;
- (4) For any two possible worlds, the probability of their being actual is equal.

Now let *Spinozism* be the thesis that there is just one possible world. We proceed by cases.

If Spinozism is true, then, by premise (1), it is a necessary truth that there are some beings, and the probability of there being no beings is 0.

If Spinozism is false, then, by premise (2), logical space comprises infinitely many possible worlds. If logical space comprises infinitely many possible worlds, and if any two worlds are equiprobable – premise (4) – then the probability of every

world is 0. If a proposition is true in at most one world, and if the probability of every world is 0, then the probability of that proposition is 0. But then, by premise (3), the probability of there being no beings is 0.

Hence, the probability of there being no beings is 0. (Van Inwagen, 1996, pp. 99-100)

Van Inwagen proceeds to support the premises, finding that premise (4) is by far the most vulnerable. For our purposes, it is that premise that is all-important, since we are not so much concerned with the starting question as with the very 'equiprobability' suggested in that premise. It will be worth our while, then, to explore van Inwagen's detailed treatment of it.

Premise (4) is intuitively attractive to van Inwagen, though he says he has great difficulty articulating why this should be so. But perhaps at the root of this difficulty is the fact that probability is never normally applied to worlds themselves, as we observed earlier, but to events and objects *at* worlds, conjoined to the fundamental difficulty – attaching to all non-MR theories – of there being one absolutely privileged world that we call actual. If actuality is a property of just one world, we look for reasons for its having that property, and are automatically at a loss, since there can be no causal mechanism *beyond* a world to bring such a thing about. As van Inwagen points out (p. 107), and in accord with the decision we took above, we are not to imagine anything like Leibniz's Creator God contemplating a range of possible universes from which to choose one to actualise. Such a scenario makes God a part of a unified causal and contingent domain (in something like the usual understanding of 'contingent', but made a primitive) that we would have to call a world.

Writing explicitly outside the context of MR, van Inwagen postulates a fictitious entity he calls *Reality*, which is whatever it is that the states constituting 'logical' space are states *of*. Each is a *maximally specific* state that Reality could be in (p. 102). (Note that this is therefore not the same sort of entity as Leslie's *Reality*, or my *total reality*; I presume, as I think Leslie would if he were to deal more in more detail with the MR version of the multiverse, that the entity he and I refer to comprehends all the 'possible' states, as well as whatever is to be *in* those states.) After some worries and provisos, van Inwagen puts forward a suggestion:

I propose: for any system of objects (that has maximal states) the maximal states of the system should be regarded as equally probable, provided that the system is isolated. (p. 104)

He gives the example of a computer being approximately equally likely to have an Urdu novel as a French (etc.) novel on its hard disc by chance – not the computer on his desk, but rather one arising from a black hole (cf. 2.3.4). This causal mechanism is entirely indifferent concerning the configuration of the bits registered on the hard disc, so this is a reasonable approximate model for the configuration of a whole, uncaused world (p. 105), and so for the complete causal isolation that van Inwagen proposes. We may agree with van Inwagen that this all works well on the intuition that the world is like his Reality, which takes on any one of an infinite range of possible states indifferently and so with equal probability. But we may also agree with him that there is a serious difficulty, in that worlds are not normally construed that way. A state, on this 'Tractarian' account, is some vast combinatorial characteristic of a plenum of atomic possibilities, every one of which is either *turned on* or *turned off*, somewhat reminiscent of Lewis's point-by-point Quine worlds (see 2.3.3). (Van Inwagen himself notes the similarity, though he refers to Lewis's 1973 formulation, which is for ersatz worlds; and we should add that there is the separate question of the configuration of a spacetime manifold, for Lewis.) Among the problems with this are that it would make of the single world at which there is nothing an infinite array of atomic possibilities that are all 'turned off', and this does not match our intuitions well (pp. 109-110). Van Inwagen wonders whether his principle of equiprobability – and so of probability 0 for each of an infinite array of possible worlds – is also applicable when we discard the Tractarian systematisation. He is inclined to think so, but closes with an admission that he can take things no further: 'I am unable to convince myself that this inclination is trustworthy' (p. 110).

It is understandable that van Inwagen appeals to such a reduction to atomic possibilities; we may read the appeal as a way of keeping things uncluttered by the *principle of indifference*. It is timely that we digress a little to examine this principle in its relevance to the topic to hand. Paul Castell presents it this way:

Pol: Each member of a set of propositions should be assigned the same probability (of truth) in the absence of any reason to assign them different probabilities. (Castell, 1998, p. 387)

By some authors Pol is disparaged almost as readily as mentioned (hence 'the notorious principle of indifference', 'this noxious principle': Hacking, 1965, p. 147). Castell is at pains to present a restriction of it that will be acceptable: but his variant is a version of a principle of *irrelevance*, enjoining us to make matching sets of probability assignments in parallel *circumstances* – circumstances in which there is no relevant difference – rather than assign the same probabilities to propositions in *one* set of circumstances. But an example he gives of a wrong application of Pol itself is useful for our purposes here:

Consider a particle that can occupy [only] the following possible energy levels $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{100}\}$, where for all $i < 100$, $\varepsilon_i < \varepsilon_{i+1}$. Its actual energy level, E , is unknown. Applying Pol to the set of propositions $\{E = \varepsilon_1, E = \varepsilon_2, \dots, E = \varepsilon_{100}\}$ yields the uniform distribution over the possible energy values. (Castell, 1998, p. 392)

Now, the imputation is that such an assignment of (subjective) probabilities is improper. I suppose that van Inwagen would readily allow such a uniform assignment only if E were found to be atomic in the sense he suggests. However, while it is a matter for physical science to regulate our speculations about the variable E , we can go a little further on our own. There are two broad options for E :

- O1: E is atomic.
- O2: E is a composite – or a resultant – of hidden variables that are themselves atomic.

The obvious analogy here is to dice. Under O1, E behaves like a fair 100-sided die; at least, that would seem to be the Inwagenian view. Assuming O1 and either MR or ersatzism à la van Inwagen, the uniform probability assignment is justified (proof omitted!). But under O2, there are further questions, for example: How many atomic variables enter into the composition of E ? Are there intermediate composite variables entering into the composition of E ? What is the precise range of values for each atomic variable? The obvious problem is that we often have no way of settling which of a pair like O1 or O2 holds, or of answering the obvious further questions if O2 holds. But perhaps Castell's comment concerning assignments to E are not entirely warranted:

Note that, apart from yielding inconsistencies, Pol also has the curious feature of apparently producing knowledge from ignorance, without empirical investigation: e.g. $p(E > \varepsilon_1) = 0.99$. (*loc. cit.*)

The inconsistencies are presumably the usual suspects in imputations against Pol: we

could assign equal probabilities to the propositions $\{E = \varepsilon_1, E = \varepsilon_2 \text{ or } \varepsilon_3 \text{ or } \varepsilon_4, E = \varepsilon_5, \dots\}$ or to any other suite of propositions based on different partitionings of the set $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{100}\}$, yielding inconsistent but equally warrantable assignments. But this move seems unfair in the present case, since there is a 'naturalness' to the presumptively maximal 100-way partitioning that makes preferring it seem unimpeachable. Similarly for the allegation concerning knowledge from ignorance. In this case, the maximal 100-way partitioning is given as *available* (assuming the restriction with 'only' that I have supplied in Castell's passage above, and which certainly seems to be understood there). Such a maximal partitioning may not always be available, but where it is there may be a reasonable *a priori* presumption in favour of Pol's equal assignment. Anyway, the 'knowledge from ignorance' in this case is only really a *modest* claim to subjective probabilities, from *incomplete* ignorance.

Even in the classic Pol problem case of Bertrand's paradox (well explicated in van Fraassen, 1989, pp. 302 ff., despite a misleading misprint and an omission in labelling a diagram on p. 306), the three ways of deriving certain probabilities concerning the length of a 'random' chord through a circle yield only the three values 1/2, 1/3, and 1/4. Even assuming the paradox unresolved by Poincaré and ET Jaynes (see *ibid.*, p. 307), this variation in probabilities is not so wide as to render all inferences about the length of the chord worthless. In many practical circumstances an assignment of probability given by $1/4 \leq P \leq 1/2$ may be extremely informative. So even a clear demonstration that the principle of indifference has technically been abused is not always sufficient to dispatch the argument that employs it; a lesser claim may often be substituted for the one that had seemed to be required. While injudicious reliance on Pol is obviously a hazard, and we shall need to be vigilant as we proceed, we shall develop a way of circumventing the major difficulties that is hard to flaw.

To return now from the digression: can we advance things without sharing van Inwagen's reluctant, problematic, and provisional commitment to Tractarian atoms? I have already hinted at a solution (2.1.2): reasons we could have for thinking an ersatz world relatively 'unlikely' must also count as reasons for thinking it impossible. Note that this is not a matter of weighing the preponderance of *species* of worlds, in the manner of Bigelow (1976; see discussion in Lewis, 1986a, p. 122); that is an important

and pervasive problem, to which we must return. But the present suggestion is simply to take seriously, once more, the idea that probability assignments are properly made only to objects *at* worlds, not to individual token worlds themselves. Our intuitions about whole worlds, whether subjugated to MR or to ersatzism, are poorly developed, and it is a fundamental error to think of how worlds themselves may vary modally (see 2.2.4); they are non-contingent entities, if they are entities at all, and as such cannot vary. It simply seems to be a corollary of this fact that they are not apt for assignment of probabilities, if we take the view that *all* of our thinking concerning probabilities is rooted in the modal situation of concrete objects *across* worlds. Questions of probability only legitimately arise if questions of contingency arise beforehand.

Lowe, in a piece to which we have already referred and which follows van Inwagen's in the same journal as a response to van Inwagen, attacks his title question 'Why is there anything at all?' from a different angle, which it is safe for us to ignore for the most part, since it deals with abstract objects in a way not relevant to our present interests. But he also comments on van Inwagen's piece:

[...] on van Inwagen's principles, if 'Spinozism' is false then it is as improbable as anything could be that we exist in the world as we actually find it. (Lowe, 1996, p. 113)

But this is not articulated with strict propriety, even if we are careful to specify this improbability as a *prior* improbability, because we would 'find' the world as we actually do if the world had all manner of configurations. It would, arguably, present exactly the same appearance to us while infinitely many unperceived details remained unspecified. *This* world is a member of a whole class of worlds equivalent under *appearance to us*, a consideration that may seem relevant when we return to the vexed matter of cardinalities. Lowe continues:

This is something which devotees of the 'Anthropic Principle' in certain of its forms would surely want to challenge very forcefully, and it is not just obvious that they are mistaken. (pp. 113-4)

But perhaps if we take all supportable versions of the anthropic principle to be non-teleological and therefore far less powerful and mysterious than Lowe allows, this objection to van Inwagen becomes otiose.

Both van Inwagen (p. 103) and Lowe (p. 114) find a difficulty for van Inwagen's

account in its application of something like the much-deprecated principle of indifference discussed above. Lowe's version of the concern leads us to something akin to my own suggestion in 2.2.5:

[...] if the 'space' of logical-cum-metaphysical possibility cannot be unambiguously divided up into a unique set of exhaustive and mutually exclusive alternatives, then it makes no sense to assign objective degrees of probability or chance to worlds, conceived as such alternatives. (p. 120)

If this 'slicing' problem works as an objection to van Inwagen's contention, though, it is a significant problem for everyone, including Lowe, since as I have suggested it strikes at the foundation of all modal discourse. An example: with his rather idiosyncratic view of the relation between concrete and abstract entities, Lowe wants it to be the case that a set only exists at worlds at which all of its elements exist also (p. 117). But if the point he makes against van Inwagen is fair, what sort of sense can be made of this clear-cut world-based restriction on sets?

To summarise, I take the case for ersatz token worlds being 'equiprobable' (difficult term!) as candidates for actuality as intuitively well based (when we attune ourselves for long enough to the changed context, and when it is compared to alternatives) and argumentatively well made out, if not absolutely clinched. Objections typically depend on a misapplication of some principle or other, perhaps involving a category error, in a domain where we are not automatically equipped with reliable intuitions. In applying anthropic reasoning to the investigation of the range of possible worlds, we do not need to concern ourselves with such arcana as a world devoid of all content; we need first the *pre-anthropic*, *Copernican* supposition (later to be superseded) that the actual world, like any unbiased sample of a single world, be 'randomly selected' for scrutiny from the whole population of worlds. This is automatically secured for MR; we shall now proceed on the basis that it is equally secured for ersatzism.

2.4.3 *An anthropic line articulated*

To keep things general, call a part of total reality characterised by a certain uniform nature (to be explained below) a *location*. One basic application of anthropic reasoning in the modal domain may be set out as an argument for preferring the hypothesis that there are infinitely many concrete locations (the MR hypothesis, roughly) to the

hypothesis that there is only one concrete location (the ersatz hypothesis, roughly). This application assumes fixed laws of physics, but that the *parameters*, which together with initial conditions and these laws determine all the causal details of a world, are not fixed. Here is the argument, which we shall sometimes call a kind of fine-tuning argument (FTA):

- P1: Reflective observers can only find themselves at a *location* (that is, one of the parts of total reality, each of which has some determinate nature) whose *nature* (that is, whose relevant, independent, and uniform global properties, expressed as values on an ensemble of *parameters*) is *friendly* (that is, sufficiently conducive to the existence of reflective observers). [Adapted from WAP to suit the terminology of this argument.]
- P2: There are infinitely many distinct possible and equiprobable natures for locations to have, some of which are friendly. [P1, our own existence as reflective observers, and two common assumptions in MR and ersatzism argued for in 2.4.2, but modified as discussed below.]
- P3: There are vastly more unfriendly natures than friendly natures. [The result delivered by *fine-tuning* arguments, to be discussed below.]
- P4: If there are locations corresponding to each of the infinitely many natures locations can have, the prior probability of at least one location's nature turning out to be friendly is exactly 1. [P2, P3]
- P5: If there is only one location, the prior probability of at least one location's nature turning out to be friendly is extremely low. [P2, P3]
- P6: On the hypothesis that there are many locations corresponding to each of the infinitely many natures locations can have, reflective observers should find it *a priori* completely unsurprising if their location turns out to be friendly. [P1, P2, P4]
- P7: On the hypothesis that there is only one location, reflective observers should find it *a priori* extremely surprising if their location turns out to be friendly. [P2, P5]
- P8: Our location turns out to be friendly. [P1, and our existence as reflective observers.]
- P9: Of two hypotheses to account for the same observation, the one that occasions less *a priori* surprise is to be preferred, other things being equal. [Taken to be a general principle of rational belief.]
- C: We, as reflective observers at a location with a friendly nature, should prefer the hypothesis that there are infinitely many locations to the hypothesis that there is only one location. [P6, P7, P8, P9]

Now, if ersatzism is true there is only one concrete world. Therefore, if *locations in reality* are equated with *concrete worlds*, the argument above supports the conclusion that we should prefer the MR hypothesis – that there is a real concrete world corresponding to each of the infinitely many fully detailed ways worlds could be – to the ersatz hypothesis that there is only one concrete world. We may fairly take these two extremes to exhaust the range of viable hypotheses concerning the number of concrete worlds. The argument is based on *natures*, rather than on fully detailed 'maximal states', so it directly secures only the overwhelming likelihood of there being *some* concrete worlds having various natures compared to there being *only one* concrete world. But the position that *more than one* but still *not all* of the detailed ways worlds might be are represented by concrete worlds was considered and found wanting in 2.2.5. So there are infinitely many equally real concrete worlds, representing among them all of the possible *natures* – and exactly one of them (from the argument against indiscernible worlds at 2.2.4) corresponds to every fully detailed way a world could be.

If on the other hand *locations in reality* are not automatically to be equated with *real worlds*, there is further work to do, which will be done below. Of course, in the meantime there are some complications to negotiate in the argument itself. Let's take them in order of the premises in which they rear their heads.

P1 makes use of a notion of a *nature* for locations, dependent on their 'relevant, independent, and uniform global properties'. I wanted a maximal class of such natures, each one of which is shared by many locations, and each one of which is specified by values taken by a set of parameters. These parameters will be discussed below, but for now let us fix ideas with something like this: *if there are parameters of such a sort as to hold uniformly throughout a location, and each such parameter must have values within some range if there are to be reflective observers, then each possible combination of values taken by these parameters determines exactly one nature that locations might have.*

In P2 we have assumed something from the discussion of ersatzism in the preceding subsection: that there are infinitely many 'ways', and that the ways are all equiprobable. (Both of these assumptions may be disputed by theorists who reject world-based accounts of modality altogether. In 2.1.3 I proposed to set these wayward theories aside

as inherently implausible and unlikely to furnish adequate foundations for modal beliefs, and I should stay with that now. I still maintain that all that I say to make ersatzism less plausible works *a fortiori* against fictionalism.) But here there is a stronger condition stipulated: that the equiprobable 'ways' in question are not *maximal states*, as van Inwagen calls them, but rather much broader *natures* of a certain sort, determined by values on a suite of parameters. This is a serious restriction in the scope and plausibility of the argument, but one I find convenient in working towards any more general treatment. We shall examine the propriety of this restriction in the next subsection.

In glossing P3 there was further mention of fine-tuning arguments, which will need the discussion I give them below. There we shall also touch on difficulties with the principle of indifference and problems arising from sets with equal cardinalities.

The probability assignment of exactly 1 in P4 is unexceptionable. The failure to assign an exact value in P5 is a consequence of our ignorance of the extent (if any) to which unfriendly natures dominate friendly natures; this will also be addressed when we discuss cardinalities.

The *a priori surprise* mentioned in P6 and P7 is supposed to capture informally just how surprising it should be for reflective observers who ponder the probability of their circumstances obtaining while prescind from the datum of their own existence. (Compare the total lack of *a posteriori* surprise they should have if they did work from their own existence as a datum; given that they exist, by P1, all parties should expect a friendly location, on either of the two hypotheses canvassed.) For P6, reflection purely on the conjunction of P1, P2, P4, the hypothesis of infinitely many locations, and the fact of their own existence as a mere supporting premise leaves no room for surprise. For P7, reflection purely on the conjunction of P2, P5, and the hypothesis of only one location motivates extreme surprise at that single location being friendly, even without taking the reflective observers using their existence as any sort of a premise.

P9 includes a *ceteris paribus* condition. I am assuming that other things are equal: that we are not, for example, significantly predisposed to prefer either MR (profligate with its real concrete worlds) or ersatzism (profligate with primitive contingency and with real abstract worlds).

The fine-tuning of parameters needs discussion. It is taken as a fairly boring truism by most physicists and most philosophers (on the word of the physicists) that certain parameters that have a role in determining the behaviour of the universe *could* have taken other constant values than they have taken. (The force of the *could* here is given more fully by 'could, while all of the actual laws mooted by physicists remain the same'. Remember that we are not here considering worlds at which 'our' physics, as a system of laws, fails to apply.) A rejected candidate for such a parameter is Chalmers' *K*, which instead turns out to be a constant fixed for all relevant worlds (see 2.3.2), but there are many worthy candidates. These parameters are supposed to have an arbitrary value that is constant and uniform across at least a vast region of spacetime; whether they must be taken as holding for an entire *universe* or *world* is a question we return to soon. It is widely believed that there are dozens of such parameters that are relevant to the sustainability of life, let alone the presence of reflective observers (see for example Barrow and Tipler, 1986; Hawking, 1988; Leslie, 1989). But the fact that each of these parameters is presumed capable of continuum many different values puts us once more in peril of appealing to the treacherous principle of indifference first brushed against in 2.4.2. What division of the range of values that each parameter may take is 'natural' and 'apt' for the determination of subjective relative probabilities? We can conveniently annex this to a cardinality problem mentioned earlier: although for each parameter there is a finite, specifiable band of values that it may take that are *friendly* (in the sense used in our argument, above), there are continuum many values in that band and continuum many outside that band. By what right then do we claim that the unfriendly values *dominate* the friendly ones? This sort of consideration appears often in Lewis (see 1986a, pp. 119 ff.) against all comers – but here it appears an impediment to securing belief in Lewis's own MR. The combination of these three thorny issues – fine-tuning, indifference, and cardinalities – will be tackled in detail in the next subsection (2.4.4).

As for the suggestion that the locations featuring in our argument may not be equivalent to worlds, where worlds are spatiotemporally and causally disjoint domains, that may not be as serious a worry as it has appeared to be. We have observed that anthropic reasoning is standardly used to support theories of quantum-theoretic multiverses, or cosmological many-world theories of one sort or another. Not all of these are well-aligned with a system of Lewisian worlds, by any means. There are many physicists

who take it for granted that the parameters of interest to us are indeed fixed for a whole universe (or world, let us say). The values of the parameters are, remember, spoken of as both universal and constant. So far we are accepting the edicts of physicists – our ‘field experts’, after all – for the sake of argument, and there may be nothing much wrong with doing that consistently for longer. The matter will be returned to towards the end of the next subsection (2.4.4); we may find that it makes little difference in the end if we speak simply and generally of *sections* of that total reality Lewis at least implicitly proffers for us to contemplate, rather than those sections which Lewis – and most of us, normally – would single out as worlds.

2.4.4 *Objections to the anthropic line*

McGrew, McGrew, and Vestrup (2001) provide subtle, representative, and up-to-date argument against the style of anthropic reasoning presented above (a kind of fine-tuning argument), so I shall focus my own discussion to their paper.

Against Leslie (1989) and others they contend that the probabilities involved in the fine tuning argument (FTA) are not *normalisable*, as will be explained below. While this is just one difficulty among many for any FTA, their treatment will enable us to expose the other problems as well. Added to the fact that it is impossible to ascertain just what ranges of the fundamental constants are compatible with intelligent life of some form or other, their objection is very likely fatal, they say, for FTA. While they take FTA as primarily an argument for design, most of what they say can be converted so that the import is that there are indefinitely many concrete worlds, and we shall continue to ignore design and keep many-world hypotheses as our focus.

McGrew *et al.* set up a K-dimensional Euclidean space of universe-types, R_+^K , each of the K dimensions representing one of our life-relevant parameters (p. 1028). So it is assumed that each parameter can take any positive real value. (We need not quibble with this blanket exclusion of negative or zero reals, though it is interesting to speculate about such an extension, for some parameters.) So each universe-type (which we shall continue to align with a *nature*, consistent with the terms of our argument above) belongs at one well-defined point in R_+^K , all by itself (and we should add, to be very explicit concerning a possible confusion, very many more specifically configured *token*

universes are therefore associated with each point). They then consider a K-dimensional ball, composed of and enclosing points representing what we have called *friendly* natures (universe-types each of which can support life). This leads the authors to speculate about the geometry and topology of the ball. May it be just a single point, so that there is only one 'setting' of the ensemble of parameters that can support life? We may all agree that this is extremely implausible; and agree also that the shape of the ball may be far from (hyper)spherical. McGrew *et al.* survey other configurations, including a foam-like distribution. It is inexplicable that they prefer to restrict the foam to 'a continuous region of universes favourable to life but sensitive to alteration of individual parameters' (p. 1029, note), since we are unable to rule out *discontinuous* distributions that have friendly regions completely topologically isolated from each other (some of which also perhaps 'trap' entire foams within their own confines). It may be, for example, that there are several quite different broad forms of life (see Leslie, 1996), and that natures friendly to one form are utterly inimical to others. Then there might at least be several well-defined but scattered balls, and so on. But we shall follow the simplification that McGrew *et al.* themselves adopt, and consider just one continuous and hyperspheroid region, along with their simplification that ignores friendliness being a matter of degree. Finally, we remark that all talk of the geometry of such a single ball is in fact dependent on there being suitable metrics for the parameters, and some specified unit for each parameter. These conditions are not easily met, as we shall see.

Now the authors use this model to give FTA a rigorous formulation. The first move here is to:

assume for the sake of argument that no particular range of values for the parameters is more likely than any other – an application of the Principle of Indifference designed to reflect our lack of information regarding universes. (pp. 1029-30)

This is well, but we should remember (see 2.4.3) that, while we may work from an assumption of equiprobability for world *tokens* by MR, and also by ersatzism as argued in the previous subsection, here the relevant lack of information concerns not worlds (or universes), but types or natures of worlds. It may turn out that some natures are indeed more 'likely' than others because there are more token worlds corresponding to some natures than to others (cf. the general import of Bigelow, 1976). We have been

proceeding explicitly on the assumption that this is not so, and that natures are also equally probable; and we can continue to do so for now. But we see here and there that McGrew *et al.* are a little incautious with this distinction themselves. (So: 'a "ball" – a set that includes the points representing our universe and those universes sufficiently similar to ours to meet the criterion we have adopted'; p. 1029. Read 'universe type' for 'universe'.)

The authors go on as expected to exhibit FTA as holding that the ball, finite as it is in all directions, has a volume that is vanishingly small as a proportion of the infinite volume of R^K , so that the prior probability of our universe having friendly settings for the parameters is minuscule.

Their first objection to this is that there is no proper *measure* possible on the volumes in question: 'the Euclidean measure function described above is not normalizable,' as they put it. This means that:

if we carve an infinite space up into equal finite-sized regions, we have infinitely many of them; and if we try to assign them each some fixed positive probability, however small, the sum of these is infinite. (p. 1031)

This in turn has the implication that all point-sized parts of these equal finite-sized regions suffer under the same defect: they too cannot have probabilities that sum to 1, as we require in any rational discussion of probabilities. The problem is not, they say, 'simply that there are infinitely many points both within the ball and outside of it: there are mathematical techniques for coming to terms with that problem.' But let us note that the problem so lightly dismissed is of a kind with the very cardinality objection that Lewis levels against Forrest and others (see below), and that Pruss deploys extensively (as mentioned in 2.2.3, for example). It seems, in fact, that McGrew *et al.* themselves stray from this normalisability problem to what amounts to a cardinality problem. For many of the specifics they move on to, assignments of 0 to individual worlds would seem to be harmless enough (cf. van Inwagen's practice, discussed in 2.4.2). But to pursue their exposition a little further:

The difficulty lies in the fact that there is no way to establish ratios of regions in a non-normalizable space. As a result, there is no meaningful way in such a space to represent the claim that one sort of universe is more probable than another. Put in non-mathematical language, treating all types of universes even-handedly does not provide a *probabilistic* representation of our ignorance regarding the ways that

possible universes vary among themselves – whatever that means. (p. 1032)

They address the rejoinder that one can get around the need to assign specific probabilities to regions in a non-normalisable space, first articulating it like this:

A small area in a larger finite region, like the bull's-eye at the centre of a target, is relatively less likely to be hit (at random) as its size diminishes in relation to the rest of the target. If we think of an *infinitely* large target with a finite bull's-eye, we seem to have the limiting case of low probabilities. Even if we cannot represent this as a ratio of areas in a strict probabilistic sense, are we not entitled to take the 'ratio' of a finite to an infinite measure as a basis for the FTA? (*loc. cit.*)

Significantly, their *only* response is that this would achieve far too much: it would be an argument in favour of *coarse* tuning, as much as it is in favour of fine tuning. The passage to set this response up is somewhat obscurely worded:

Suppose that the open set of life-friendly [universe-types] contained a ball in which the various parameters, rather than being constrained to within tiny intervals around those that characterize our own universe, could take any values within a few billion orders of magnitude of our values. (*loc. cit.*)

Given that the openness or otherwise of the set seems irrelevant, I take the underlined material to mean in effect: 'Suppose that the parameters could take values differing by up to a few billion orders of magnitude from "our" values and still be friendly.' The authors continue:

It is hard to imagine anyone's being surprised at the existence of a life-friendly universe under such circumstances. Yet the 'ball' in this case is isomorphic to the ball in the FTA: both of them have measure zero in R^K . In consequence, any inference we can draw from fine-tuning is not only paralleled by a coarse-tuning argument; it also has precisely the same probabilistic force. So if we are determined to invoke the Principle of Indifference regarding possible universes, we are confronted with an unhappy conditional: if the FTA is a good argument, so is the CTA. And conversely, if the CTA is not a good argument, neither is the FTA. (*loc. cit.*)

Well now, let us analyse this CTA-based objection to FTA, bearing in mind that the authors are very strict about setting aside mere intuition as a basis for argument: we are to do the hard maths here, and that's all there is to it.

The authors take FTA to be concerned with friendly values for parameters restricted to 'tiny intervals around those that characterize our own universe'; but in setting up CTA, they want the friendly values to be 'within a few billion orders of magnitude of our values'. Now, the latter is a matter of ratios, not intervals. But to make FTA and CTA

comparable, we would need them to be formulated *both* in terms of intervals from or *both* in terms of ratios to the observed values at our world. Work, then, with the first option, and restrict discussion to just one parameter, which we shall call P_1 , and say that P_1 happens to take the friendly value V_1 at our world. Let us further suppose that the differences between FTA and CTA are, as McGrew *et al.* suggest, somehow a matter of billions – say $10^{1,000,000,000,000}$ (call this number c) to be on the safe side, given that there was talk of orders of magnitude. So let the interval on P_1 that is friendly (within which V_1 falls) be I_1 according to FTA, and $c \times I_1$ according to CTA. (I shall use *interval* a little loosely to denote both some *range* on the real line between two fixed points that define that range and also the *length* of such a range; this is harmless enough.) Now, what units have we assumed for these measures on P_1 ? We must have chosen *some* units, otherwise any specification of I_1 is indeterminate, and so also is our value c quite meaningless. If our arbitrary choice of units had gone a certain way, then the absolute values bounding the friendly ranges on P_1 may happen to go like this:

$$\text{By FTA:} \quad 10 < V_1 < 11 \quad [I_1 = 1]$$

$$\text{By CTA:} \quad 10 < V_1 < c + 10 \quad [I_1 = c]$$

(For CTA, the lower bound might have differed from that for FTA as well as the upper bound; but I simplify harmlessly.) Now, this looks intuitively like an enormous difference between FTA and CTA. But suppose we happen to have chosen very much larger units. We might well have, after all: there is the whole infinite range of positive real numbers to choose from, for a parameter P_1 that is itself presumed capable of taking *any* positive real value with equal prior probability. Suppose our units were larger than the units assumed above by a factor of, say, $100 \times c$. Then:

$$\text{By FTA:} \quad 0.1/c < V_1 < 0.11/c \quad [I_1 = 0.01/c]$$

$$\text{By CTA:} \quad 0.1/c < V_1 < 0.01 + 0.1/c \quad [I_1 = 0.01]$$

How do things look intuitively now? By FTA, V_1 still looks admirably fine-tuned; but it also looks fine-tuned by CTA. Let's look at the numbers differently, spelling out c 's actual value:

$$\text{By CTA:} \quad 10^{-10,000,000,000,000} < V_1 < 0.0100000000001$$

And *surely*, we might reasonably say, given the precarious nature of the reasoning we perforce had used in deriving bounds for the friendly interval I_1 , we can round all this out, so that:

$$\text{By CTA:} \quad 0 < V_1 < 0.01 \quad [I_1 = 0.01]$$

'What coarse-tuning argument?' we might fairly ask.

McGrew *et al.* had ruled out intuitions as unreliable in this very speculative domain, then they appealed to the intuition that CTA was absurd; as if that weren't impropriety enough, we have just seen that a mere change of units makes CTA look eminently reasonable by the very intuition the authors have illicitly appealed to. So, by their own ruling, FTA should look reasonable to us also (*a fortiori*, in fact) – if indeed there is any real distinction between these two arguments.

It may be objected that some of the parameters are dimensionless – mere matters of pure ratio, for example (see discussion of K , in 2.3.2); in such cases it is arguable that change of units may not do the trick to convert CTAs into FTAs. But there are four replies to this. First, and rather incidentally, some such ratios may turn out to be matters of necessity after all (just like K did), and I would be happy with this, because it tends toward the restriction of recombination principles, which I want to argue for. Second, we can simply remove such difficult parameters from the set that is of interest to us; we don't need them all for our argument. Third, even without conversion by change of units, the authors' appeal to an intuition as to what are wide and what are short intervals is both against the rules they have set (as mentioned above) and answerable simply by pointing out that *all* finite intervals on an infinite line are 'short', anyway. (In the hard-core cosmological literature one comes across examples of physicists taking confinement to several orders of magnitude for some ratios between variables to be surprising, which supports the contention I here make in the abstract.) And fourth, even dimensionless constants have the values of their numeric expressions settled by the choice of units for the parameters from which they are derived, and among which they express ratios. (K , for example, was found to be dimensionless, but pressure, volume, and temperature were measured in units that we could have chosen differently.)

A slightly stronger objection might be that our choice of units is never strictly *a priori*.

While we acknowledge an infinite range of possible values for the parameters, in practice our choice of units is conditioned by observed values. If this were not so, we might measure the mass of everyday macroscopic objects in femtograms, for example. But this objections still misses the mark, by the third point above concerning our intuitions and the infinite; and because the *idealised* appeal to *a priori* choice of units is not shown to be unreasonable.

We can obtain the same sort of result if we choose to work uniformly with ratios instead of intervals in distinguishing FTA and CTA; and the arbitrary (because unit-dependent) lower bound equal to 10 that we started with above can be replaced with any other positive value without affecting the transformability of a CTA into an FTA by a choice of appropriately large units.

Nor need we be ruffled by the fact that FTAs can be made to look like CTAs – by a switch to much *smaller* units. The point is that, under the assumptions McGrew *et al.* set up for us, there is really only a *tuning* argument – and it may be a good argument in its context. If we take seriously the assumption that the range of values on P_1 to which we assign *a priori* equal probability densities is between 0 and $+\infty$, and choose our units before setting out to measure V_1 , finding any nameable finite value for V_1 should astonish us! This is a lesson we can learn from the story of π and γ , in 2.4.1; and this even before we speculate concerning the range of friendly values into which we know V_1 *must* fall (since we are here to measure it). And quibbles based on the fact that we always select units in a way conditioned by experience are answerable as above.

It is arguable that in most of what we have just examined normalisability is not the real issue, since there has been little question of an attempt to assign finite probabilities to finite intervals for comparison, nor of any attempt to compare such probabilities. In any case, McGrew *et al.* go on to consider whether the principle of indifference is the real culprit:

If we do not insist on treating all equal intervals for all parameters as equally probable, then we can perfectly well speak of the probability that a particular parameter falls within a given interval by invoking density functions that integrate (or can be scaled so as to integrate) to unity. (p. 1033)

They then observe that there is no way to choose among the infinitely many candidate

density functions that will achieve this, or indeed to exclude even those density functions that work *against* the desired conclusion. In certain limiting cases the observed values taken by the parameters turn out to have prior probability 1, which is equivalent to there being no tuning at all: the observed values are simply necessary.

This may appear to be a satisfactory reduction of any move away from assignment of equal probabilities. But is it? Perhaps there are principled ways of limiting the range of candidate values for some parameters, while retaining grounds for the presumption of a flat probability distribution within that limited range. This is ultimately a matter of physical theory, not philosophical investigation. We can, however, look at the consequences of such a move.

Again, let us restrict our analysis to the single parameter P_1 . The presumed friendly interval I_1 will then be smaller by a definite proportion than a larger interval, say J_1 , that physical theory may settle for us as the range of values possible for P_1 .

In this highly idealised situation, in which by now we have availed ourselves of a large number of auxiliary assumptions, how do we fare in assessing the *a priori* probability of V_1 – our observed value of P_1 – falling within I_1 ? Not well. A Lewis-style cardinality objection lies in wait, and it is curious that McGrew *et al.* do not invoke it specifically and explicitly. The cardinality of the infinite set of points in I_1 is the same as the cardinality in J_1 . And there is no available metric on the parameters we are considering, which are after all not lines in empirical space on which we can do measurements and comparisons of lengths. All we have to work with are sets of points that make up intervals. Here is where the principle of indifference really does become worrisome: anything we can do to assign equal probabilities to sets of points as intervals on P_1 is viciously arbitrary. And so our generous helping of assumptions has all been to no avail.

There is a way forward when confronted with this common impasse, though it scarcely ever seems to be taken.

Everyone wants to insist on the difficulties that arise from our parameters being continuous real variables; and then the defects associated with the principle of indifference can always be raised to vitiate our reasoning. But it is not compulsory to work with the parameters that way. In statistical inference we have the option, where we have reason to suspect some feature of the scaling of a variable, or the accuracy of

interval measurement in gathering the data, of working *non-parametrically*.

(I have not seen anything in the literature that closely resembles my line of reasoning in what follows. I have come across only one other explicit and sustained treatment involving anything like the multiplicative strategy that I am about to present: by Stephen Coleman, 2001. The context there is the question whether the apparent fine-tuning of the universe requires explaining – a question resolved in the affirmative. While Coleman applies his strategy, as I do mine, to probabilities concerning the ‘constants’, there is no discussion of the principle of indifference, of non-parametric analysis, or of anything that brings us into the territory marked out by McGrew *et al.* But it is a stimulating and original piece, in which the author also offers useful discussion of the question what should and what should not be considered surprising. I also find something bearing a family resemblance to my analysis in Nozick, concerning binary partitioning of a probability space and safe conservative use of the principle of indifference; but the application is different, and there is nothing like my multiplicative move; Nozick, 1981, pp. 127-8, and note 9 on that passage at pp. 669-670.)

‘Non-parametrically’ means this: We might, for example, ignore (or might not even have) any information beyond the mere rank of each score within the data set. We might take as data only the rank order of heights of a group of people, say, without using the specific absolute heights. And we can do something like this with the parameter P_1 . We can go as far as to treat it merely as a discrete binary variable. (The *parametric* element in *non-parametric* is not to be confused with the term *parameters* as we have been using it; the similarity is an unfortunate coincidence.)

First the preliminaries, using just one parameter, as before. There is a perfectly basic, conservative, and minimally committal assignment of prior probabilities to the set of values on P_1 : it is 0.5 to the set of friendly values and 0.5 to the set of unfriendly values. These assignments sum to 1 as they should. But what about the objection concerning our use of the principle of indifference? We have assigned equal probabilities out of ignorance. There are two responses to make to this charge. First, while we have assigned probabilities out of ignorance, our assignment is *maximally* conservative for our support of FTA. We have not sought to exploit some size difference in the sets of points in question. Despite the *appearance* and the strong *intuition* that there is such a

difference to appeal to, and that this will favour a much lower prior probability assignment to the friendly values, we have accepted the full force of any cardinality objections and ceded the ground claimed by McGrew *et al.*; we have not taken any unearned advantage. Second, we can even afford (as we shall see) to concede that in the case of some parameters there may be reasons to assign a *higher* value to the set of friendly points than to the unfriendly set – though it is hard to see how that such a concession could be forced from us. Any strictly *a priori* argument – which must, especially, rule out of court the evidence of the observed values of parameters – could work at least equally to decrease as to increase the *a priori* probability of friendly values.

Now, it will be objected that none of this is of any use. With a prior probability of 0.5 concerning P_1 all we can say is that there is an even chance that the world should turn out to be friendly to life. Hardly support for a many-world hypothesis! But remember that there are supposed to be ‘several dozen’ independent parameters that are relevant. We can assign our probabilities non-parametrically to those as well, ignoring all our inclinations to compute probabilities according to suspected tight intervals on a line presumed infinitely long. We may well be justified in assigning 0.5 as the probability of a friendly value in the case of each one of these several dozen parameters.

As we proceed, the numerical details will differ according to all manner of further assumptions; but let us do one or two sample calculations of the net effect of such assignments.

First, conservatively take *several dozen* to mean just *three dozen*. Of these 36 parameters, conservatively remove from consideration 12 parameters suspected of not being genuinely independent of some of the remaining 24, or of not truly being subject to variation. For each of 12 of these remaining 24, assign 0.5 as a prior probability of their taking friendly values, as justified above. For each of the remaining 12 parameters, conservatively assign 0.67, making a more-than-fair concession to anyone arguing on *a priori* grounds that the assignment should weigh towards the friendly values. Here is the resulting calculation for the joint *a priori* probability of the whole suite of 36 relevant parameters taking values that are friendly to life:

$$\begin{aligned}
P(\text{life-friendly nature}) &= 1^{12} \times 0.5^{12} \times 0.67^{12} \\
&= 1 \times 0.000244 \times 0.00818 \\
&= 0.00000200
\end{aligned}$$

So, striving in every detail to calculate conservatively from the barest assumptions, we still end up with an *a priori* probability of a mere 1/500,000 for a life-friendly world. If we started with equal credence for our two hypotheses – one concrete world or many – then this solid evidence must surely effect some change in our credence.

The figure we get if we start with an assignment of 0.5 for each of 36 parameters, by the way, is 0.0000000000146, or 14.6 trillionths.

One might be inclined to analyse *semi-parametrically*, distinguishing three sets for each parameter: values friendly to life, values too low to be friendly, and values too high to be friendly. Assigning probabilities equally to each such set (which is still quite conservative), a calculation with 36 parameters each assigned 0.333 chance of taking a life-friendly value yields an overall probability of 0.0000000000000000643, which is 6.43 quintillionths. But allowing any parametric contamination of the argument, even if it can be theoretically justified (which is not certain, though there is much precedent in theory of statistical inference) weakens a very powerful analytical gambit; so perhaps it is best to rest content with the probabilities yielded by the pure non-parametric approach.

It is hardly required of us, then, to play the customary cardinalities-cum-non-normalisability game; we have a worthy alternative. But let us recall that all of this has assumed equiprobability for *natures*, which still needs to be examined. And beyond that, we need to think that we have worked only with worlds that have a physics like our own, with only the parameters varying. Further still, we have not yet considered the vast range of putatively 'logically possible' worlds at which things are so unruly as to defy any systematisation at all, on the supposed wild and lawless outer reaches of modal space.

In fact, all these frayed loose ends are neatly singed using the non-parametric gambit. Even if natures are not equiprobable, we do not strictly require that assumption when all the parameters that determine natures are thought of as mere binary variables. We need

only the far less assailable assumption that each parameter takes some friendly value or some unfriendly value with equal probability – and we can afford to make concessions even then, as we have seen. This minimal assumption may reasonably be taken to quash the objection that it is only token worlds that should be presumed equiprobable, with there being, for some reason, a higher cardinality assigned to the infinite set of token friendly-natured worlds and the infinite set of token unfriendly-natured worlds. But we shall return to consider the relative probability of natures below.

A similar move annuls the objection that we have assumed worlds with 'our' physics, examining changes only on certain life-relevant parameters. If the domain of possible, equiprobable token worlds is extended to include worlds with other physical laws (for the moment presuming such worlds broadly possible), this can only help our support of FTA. Working with the most conservative of the results derived above, and conservatively reducing the matter of having 'our' physics exactly or any of the incalculably many other putatively possible physics to a mere binary variable, we may reasonably use the following calculation for the *a priori* probability of a world having a friendly nature:

$$P(\text{life-friendly nature}) = 1/500,000 \times 0.5 = 1/1,000,000$$

And so also with the extension to include utterly unruly but 'logically' possible token worlds.

One could be worried here, and also earlier in the discussion, that the malignant principle of indifference has still succeeded in casting its glamour over us. The set of all token worlds has been partitioned according to possession of life-friendly natures, which is an arbitrary principle. By other partitioning principles, the probabilities might have worked out differently. But, we may reply, this could only be a worry if the parameters we worked with were derivative rather than basic. Without descending to the specifics of physical theory, on the assurance of a consensus of physicists the parameters we have dealt with *are* independent and fundamental. And the fact that we have selectively spoken only of life-relevant parameters is no problem, since any expansion beyond those would simply be like adding extra sides to a die; and our argument is only impugned if we were caught illegitimately making life-friendliness *less* probable, not *more* probable by some such restrictions as we have adopted.

Perhaps what plausibility such an arbitrariness objection might have is derived from models such as we find in McGrew *et al.*: R_+^K is a Euclidean space, and its coordinates are K quite separate parameters. Perhaps when we imagine R_+^K we are picturing it as like the real-world empirical space that we inhabit, in which a set of three coordinates must be imposed arbitrarily. Once we have chosen coordinates for such a space, they are not final; we can rotate or translate them at will to make new coordinates better suited to our needs or whims. But R_+^K is not like that.

We are near the end of the defence of the specific anthropic line of argument introduced in the previous subsection (2.4.3) – a version of FTA for preferring a many-concrete-world hypothesis to a one-concrete-world hypothesis. We could also have pursued a line of argument based on worlds not as tokens of world-types determined by *natures* (defined in terms of life-friendly parameters) but as tokens each of its own *infima species*. In that case we could then have constructed a space of vastly many dimensions (cf. the general procedure in Penrose, 1990, pp. 228 ff.), though we could have only an indistinct notion of what those dimensions would be, and attempted to partition it into a set of types of worlds with reflective observers and a set without them, to explore the *a priori* probabilities. There would be a mapping of all token broadly possible worlds onto such a space (though, I say, with no guarantee that every point would correspond to some world). The articulation of such an argument would be a hugely more complex affair than what we have just been through. An intermediate kind of argument would have involved only world-types each of which is regular enough to be thought to have its own physics, susceptible to some sort of modelling by assigning each physics to its own point in a suitable space. But the theoretical difficulties would again be huge, and confronting them would be a waste of effort; as I have argued above, we can do all that we need to support FTA with the limitation to *natures*.

We now must address just one pair of related problems signalled above: first, that some natures are arguably much more *a priori* probable than others – perhaps too quickly dismissed above; and second, a version of the ‘slicing’ problem that we have encountered, which the problem just mentioned neatly leads to – that FTA may not yield a result unequivocally in favour of MR, since total reality may be partitioned in other, non-world-like ways.

... first of this troublesome pair, then: some natures may turn out to be far more likely than others. In particular, the very same settings on the parameters that are friendly to life are, very plausibly, just those that issue in the greatest intricacy and variety among token worlds. By many unfriendly natures, the parameters will determine worlds with extremely short histories, or small spaces, or only tiny amounts of relatively unstructured matter, etc. But then, our argument in terms of equiprobabilities of natures is called into question. Here are four observations about this problem: one against our FTA and three for it.

First, if it works at all the objection does indeed hit hardest against an MR interpretation of FTA. (I shall speak loosely, in this context, of worlds as if they were universes.) If FTA is interpreted as favouring some cosmological, many-big-bang style of many-world hypothesis, the objection is weak because on such a theory the parameters are presumably set *by, at the time of, or along with* (take your pick!) the big-bang event itself. We should recall Carter's wording of SAP, from 2.4.1: '...the Universe (and hence the fundamental parameters on which it depends)...'. And the *a priori* probability of each such setting is, we should think, independent of the consequent variety engendered in the world in question, so no argument from greater diversity among token worlds can be effective. By contrast, MR is not about generative principles and their operations, but about the necessary existence of every possible token world without any thought for the manner of generation of the variety in that world; and the natures of which we have spoken, correlated with parameters in which we happen to be interested, are a sort of abstraction from token worlds taken as primary.

Second, much of the supposed greater variety in friendly worlds may be spurious. While structures supporting life are *interestingly* complex for us, who are alive, it is not so certain that in more pure informational terms we should expect a 'randomly' selected lifeless world to be less complex than a 'randomly' selected world with life. In fact, by close analogy, of two pictures with equal numbers of pixels the one whose computer file is harder to subject to lossless compression is the less structured, more noisy one, since it has more pure information in it. (Or think of life as like crystallisation, as we are sometimes urged to do. A crystal is a simpler structure, analysed on many of the available levels, than an equal-volumed uncrystallised consignment of the same stuff.

So might the presence of life – or better, the ubiquity of the sorts of structure that are a precursor to life – reduce the informational complexity of a world.)

Third, since in the course of articulating our version of FTA we had to make strong concessions because of cardinality objections, we ought to be able to take advantage of the very same reasoning in defending our FTA against the present attack. If it seems that one nature corresponds to greater variety in its token worlds than another nature does to its token worlds, this must be a matter of different *numbers* of token worlds for each nature if the objection is to have any force. But since it will be granted that there are infinitely many token worlds for each nature (practically without exception, at least), it would take further argument to establish that the cardinalities themselves are different, and further argument again that their differences are such as to work against our FTA. On the face of it a case can as easily be made for any such cardinality differences favouring FTA.

Fourth, perhaps we have done enough, in the course of exhibiting the full detail of the nature-based FTA, to show that a token-based FTA – immune to the objection we are considering – would also be fully robust. It is only for expository convenience, as discussed above, that we have avoided such a version.

The second and more nuanced of our pair of final problems for FTA follows straightforwardly, since we now need to look again at the sorts of *locations* (see P1 of our argument, in 2.4.3) that are the possessors of the natures we have spoken of. Though we may have done enough to defend the argument if it is granted that total reality divides ‘naturally’ into uniform-parametered worlds better than into any other sorts of locations, this has not been finally settled. Indeed, I have suggested that it is improper to assume such a ‘natural’ partitioning, which is simply one of many that are feasible, and one that we are very much inclined to choose because of our confinement to one world and our impulse to speculate beyond the observed course of things at that world. A partitioning into worlds does all the modal work for us that we wanted a model for in the first place. But total reality need not divide in accord with such accidental urgencies as ours.

I have hinted that it may not be a matter of great consequence if other sorts of divisions are taken to be the locations that our anthropic FTA is concerned with, and I stick with

that. In terms of the natures version of FTA that we have articulated in detail, it could for all we know turn out that MR is false. There might be just one vast universe, perhaps even with spatiotemporally and causally separate regions permitted, and there might be an enormous number of spatiotemporal locations, each with one or other of the natures that we have assumed broadly possible. And total reality may not, for all we know, be as all-encompassing as to instantiate every possible nature. We considered only two competing hypotheses: one friendly concrete location versus a plenum of all possible concrete locations. But now we must consider messy intermediates between these extremes.

As I have suggested, in these intermediate positions we must accept either ersatzism or Spinozism (since we have ruled out other, non-worlds-based accounts quite early on as unproductive for our inquiry). Now, if Spinozism is right, we have secured a result, which we can take into the third and final part of this thesis and harness to all the work that we want to do there. It is a position counter to the deepest intuitions, but we have passed well beyond any easy acceptance of intuitions regarding modality. There may be nothing we can do to defeat this interpretation of the intermediate position regarding total reality.

If however ersatzism is true (despite the objections raised against it in 2.1.2), then total reality as a whole is subject to a first, dichotomous, partitioning: into a single, concrete real world and a huge ensemble of abstract worlds – as many, in fact, as there are if MR turns out to be true, as argued early on concerning the independence of recombination principles and ontologies for worlds. But we have accepted it as highly plausible, by van Inwagen’s arguments examined in 2.4.2, that under ersatzism every possible world is ‘equiprobable’ as a bearer of the uniquely possessed property actuality. If this is so, the domain relevant to our FTA is all of total reality – all locations in the concrete part and all in the abstract part. We know that we are at one of the locations in the concrete part, of course; but in our judgments concerning the range of other locations than our own we must consider ours to be selected from all possible locations – concrete or abstract. This is a direct consequence of the equiprobability of all possible worlds. So if we are unsure whether MR or ersatzism is right, we might reason that the alternative hypotheses are as presented in Figure 1.

Four hypotheses:

H1: MR is true, and each world has only one location: some locations are friendly.

H2: MR is true, and some worlds have many locations: some locations are friendly.

H3: Ersatzism is true, and the single concrete world has only one location, with a friendly nature.

H4: Ersatzism is true, and the single concrete world has many locations, at least one of which is friendly.

*Some symbols:**Abstract worlds:**Concrete worlds:**Abstract locations:*

← Friendly locations →

Concrete locations:*Our concrete location:**Four hypothetical views from a window opening onto modal space:*

	<i>One location per world:</i>	<i>Some worlds have many locations:</i>
<i>MR:</i>	<p>H1:</p>	<p>H2:</p>
<i>Ersatzism:</i>	<p>H3:</p>	<p>H4:</p>

Figure 1 Four hypotheses, and diagrams to aid visualisation.

H3 can readily be ruled out by FTA. Remembering the conservative and cautious way in which the details were filled in, and bearing in mind that teleological explanations for actualisation of a concrete world have been peremptorily ruled out, it is wildly implausible that the only concrete world is among the tiny fraction of worlds that are friendly to reflective observers. (We may perhaps allow ourselves this licence to speak of tiny infinite fractions, ignoring cardinality objections; we have done the required hard work already with the alternative non-parametric analysis above.)

It seems to be a fair assumption that the likelihood of any randomly selected location being friendly is unaffected by whether it constitutes a whole world by itself or is a region of a many-location world. This is a reasonable extension of the standard assumption we adopted earlier, that parameters determining the nature of a location take their values with equal probability; and it is in tune with LR. We proceed on this basis.

Anyone committed to ersatzism, then, is committed to either H4 or the implausible H3. Let us now examine the implications of H4. For a start, there may be an argument to show that the cardinality of the set of many-location worlds is greater than the cardinality of one-location worlds, under H4. Once we allow that many-location worlds are possible, we may not be able to restrict in any principled way the number of locations worlds can have. And then, to put it crudely: there are vastly more many-location worlds than one-location worlds that are candidates for actualisation; worlds are 'selected' to be actual with equal probability; so by H4 it is overwhelmingly likely that the actual world is a many-location world. (But if this reasoning fails, and the actual world under H4 has just one location, the scenario is still as implausible as any delivered by H3, since it is directly ruled that way by FTA.)

For this reason I have interpreted H4 in Figure 1 so that the actual world includes 'many' locations. Printing limitations preclude showing all but 2 of the 346,573 locations that our most severely conservative estimate above shows would be needed to give a 0.5 chance of the actual world including at least one friendly location! The proof of this result:

$$\text{For any location, } P(\text{location unfriendly}) = (500,000 - 1) / 500,000$$

$$= 0.999998$$

$$P(\text{all of 346,573 locations unfriendly}) = 0.999998^{346,573} \approx 0.5$$

$$P(\text{at least 1 location in 346,573 friendly}) = 1 - P(\text{all of 346,573 locations unfriendly})$$

$$\approx 0.5$$

And it can be shown by similar elementary means that the single concrete world would need 1,000 locations to give even 1 chance in 500 of including a friendly location, still by our very conservative assumptions. And by the usual far less limiting assumptions, working *parametrically* with the parameters (which may yet be justifiable), the number of locations required to make it at all likely that a concrete world could support our existence is vastly higher. So the ersatzist is virtually committed to the actual, concrete world having a very large number of locations, each with its own randomly set 'universal constants' – values on the parameters relevant to life. Meanwhile, on the modal realist alternatives – either H1 or H2, using precisely the same principles of worldmaking (as opposed to ontology) as H3 and H4 respectively – the result is simply *certainty*. Prior probability 1, that some concrete world be such as to support our existence; and then the indexicality of actuality does the rest.

If this is not sufficient 'transcendental' evidence in favour of MR and against ersatzism, to add to the general difficulties for ersatzism rehearsed early in this part, consider the implications of what amounts an ersatzist commitment to a non-modal multiverse. There seem to be three alternatives: the locations may be part of a spatiotemporal whole, each being a different epoch or a different spatial region of that whole; the locations themselves diverge according to some variant of principles outlined in the Everett interpretation of quantum mechanics; or the locations are entirely spatiotemporally unconnected. The first two of these are surprising macro- and micro-physical commitments for an ersatzist to be hobbled with, to put it mildly. The third invites these questions: If you already allow some very large number of completely spatiotemporally and so, may I suggest, *causally* separate, actual, concrete quasi-worlds, please remind us what your objection was to an *infinite* number of concrete worlds? Isn't your position hopelessly arbitrary? These questions are fair if we are to take ontology seriously, as we set out to do at the start. MR and ersatzism postulate ontologies, and once they do that they can legitimately be compared on the basis of these ontologies, freed from both intuitively and linguistically motivated concerns –

and, most importantly, even from *modal* concerns. At the very least, in fact, ersatzism may reasonably appear no less counterintuitive and profligate than MR is often alleged to be, and it is shown to be hostage to cosmological theory in a way that MR is not.

That concludes our main anthropic defence of Lewis's modal realist ontology. But we now turn to a transcendental consideration of the independent question of recombination and other worldmaking strategies, in order to settle more precisely than earlier what limitations are proper on these.

2.5 FEWER WORLDS THAN ARE DREAMT OF

2.5.1 *Introducing paranthropic reasoning*

Anthropic reasoning, conservatively construed and applied, supports the contention that reality is larger and more varied than had been thought, by showing that our situation is ('Ptolemaically') special: so other places are not like this one, and variety is greater than had been thought. A good anthropic conclusion is one suggesting that reality is less uniform than one might have supposed. May there not also be some sort of transcendental reasoning that aims at delivering a converse result? There is – though it is normally neither specifically distinguished nor named. Such a converse I call *paranthropic* reasoning (Greek prefix *para-* + *anthropic*). An *anthropic* argument proceeds from evidence that certain observed conditions are what provide our very opportunity to observe them to the conclusion that we can only observe such conditions as those (or, on a more moderate formulation, that we are *a priori* more likely to observe certain conditions obtaining than their not obtaining), regardless of the nature of reality as a whole. A good *paranthropic* conclusion is one suggesting that reality is more uniform than one might have expected, for some or other *a priori* reason. A *paranthropic* argument proceeds from evidence that certain observed prevailing conditions are *not* positively correlated with the prior probability of our observing them to the conclusion that such conditions are more pervasive than we should have supposed if they *had* turned out to be positively correlated.

To illustrate with diverse examples: if you are already inclined toward the 'Ptolemaic' theory that most stars in the universe are unlike our sun, anthropic arguments lend plausibility to your theory (most stars *may well* lack planets, for example, since a planeted star is far more conducive to the existence observers than an unplaneted one). Similarly for your theory that most things in the world do not have a constant temperature (since your having a constant temperature seems to be highly conducive to your subsistence as an observer, most things in the world are not observers, and we have independent reason to think that temperatures can change). Paranthropic arguments support the 'Copernican' contention that animals in general have a cellular structure (the types and the individual animals that I observe all do, and this evidence seems uncorrelated with my selection of them to observe, or their selection of me to be

observed by: if reality were less constrained, I would observe greater variety): and they support Newton's contention that gravity is universal (*on the assumption that its obtaining where we are seems uncorrelated with our presence as observers*).

Here is a formulation of the principle, worded to conform to our version of WAP, shown here also for comparison:

PP: Features that are no more conducive to the existence of reflective observers than the absence of them would be are representative of the whole of total reality, not just the observers' own location.

WAP: Reflective observers can only find themselves located in those parts of total reality whose ensemble of properties is sufficiently conducive to the existence of reflective observers.

Perhaps PP is no more than a standard presumption that one's sample of reality is unbiased – a common enough basis for sound statistical reasoning. But I have placed that common assumption on an elevated footing similar to that of the anthropic principle, to do the same general sort of work that we expect from that more established principle. In 2.4.1 we noted that many think the anthropic principle trivial, but that it is useful despite this appearance. I claim the same for my paranthropic principle. We now deal with two preliminaries, on our way to a pure application of PP.

2.5.2 *First preliminary: the 'epiphenomenal' objection to MR*

We saw in 2.2.6 Forrest's epiphenomenal objection to MR, which was the suggestion that Lewis alone is unable to use Occam's razor to exclude the presence of 'epiphenomenal rubbish' at the actual world. We now examine it more closely:

[Consider] someone who knows all about the actual world except whether there exist some totally redundant entities which I call epiphenomenalons. Hence, on Lewis' theory (s)he knows that she lives in one of an infinite set of worlds which differ only in their epiphenomenalons. Furthermore (s)he knows that of those worlds precisely one is lacking in epiphenomenalons. So by a proportional syllogism (s)he should infer that she lives in one of the worlds with epiphenomenalons. So Occam's Razor could not rationally be used by such a person to excise epiphenomenalons, which is absurd. (1986, p. 22; see also Forrest, 1982)

It is, for a start, not clear how this could be accepted as a reasonable reductio of what it is intended to be a reductio of. Forrest is drawing attention to a very natural consequence of modal realism annexed to Lewis's readiness to consider possible just

about anything that seems coherently describable. It can be made, I think, to work as a *reductio* of those positions with the *further* annexation of Lewis's assertion that ours is a 'contingently materialist' world.

Consider Lewis's response (1986a, pp. 120-1; the response focuses on Forrest, 1982, in fact). He addresses Forrest's contention that the deceptive worlds (those with epiphenomenalons) predominate over the non-deceptive worlds (those without). He argues in effect that the cardinality of the set of 'rubbishy' token worlds is the same as the cardinality of the set of 'clean' token worlds, so that we should not expect our world to be clean any more or any less than we should expect it to be rubbishy; and that this is simply a problem of induction shared by advocates of all theories of modality.

Now, while I certainly agree that the problem hereabouts is everyone's equally, I shall argue that it is not quite the problem that Lewis thinks it is.

Epiphenomenal rubbish may have spatial characteristics but be very little extended in space. (Lewis does not contest this; and in Forrest's original 1982 formulation the epiphenomenalons are spatial subatomic particles.) Consider some possible but non-actual world w_b with a spatiotemporal manifold very much like our world's. Partition the space of that world into as many and whichever largish divisions as you like – say 100. Now formulate 100 parallel hypotheses, each concerning just one of the spatial regions R_1 to R_{100} . Consider the hypothesis associated with R_1 , which we shall call H_1 :

H_1 : R_1 is free of epiphenomenalons

What probability should we assign to H_1 ? The Forrest-style answer (in respect of the *non*-actual world w_b), is to say that we should assign a probability of practically 0, since the number of microscopic epiphenomenalons in R_1 is no more likely to be 0 than 1, 2, $10^{1,000}$ or any other small or indefinitely large number, allowing \aleph_0 as a limit. Lewis would presumably agree, but apply some variant of his cardinality objection: there are \aleph_n worlds (the value of n depending on details to be fixed somehow), varying in all sorts of ways, each world having a counterpart of R_1 , for each possible number of epiphenomenalons at R_1 . The cardinality of the set of worlds having some (natural) number of epiphenomenalons greater than 0 is $\aleph_n \times \aleph_0$, which is equal to \aleph_n . So the cardinalities are such that we can assign no definite probability to H_1 . Very well; but let

us instead work *non-parametrically* with H_1 . On the relevant assumptions, there is no reason to prefer 0 as the number of epiphenomenalons at R_1 , and let us grant also that there is no reason to prefer that the number be greater than 0. Let us therefore feel free to assign a probability of 0.5 to H_1 , on the grounds that any argument against this will favour a lower probability (for fairly obvious reasons, provided the assumption of indifference in the distribution of epiphenomenalon-natures for worlds and regions, granted by all, is respected), and fortified also by the arguments for careful and conservative uses of the principle of indifference presented in the previous subsection (2.4.4). Now, remember that R_1 is just one of 100 regions of w_b ; so the joint probability of there being no epiphenomenalons at all of the regions and so at w_b itself, is given by the joint probability of H_1 to H_{100} being true. This is worked out like this:

$$\begin{aligned} P(w_b \text{ is free of epiphenomenalons}) &= \prod_{i=1 \text{ to } 100} P(H_i \text{ is true}) \\ &= 0.5^{100} \\ &\approx 10^{-30} \end{aligned}$$

Lewis's contention that we have no reason to favour either of the hypotheses he considers – epiphenomenalons versus no epiphenomenalons – is defeated, since under MR the argument regarding w_b must apply equally to w_a . (For any world large enough to contain at least 100 epiphenomenalons, our credence concerning the hypothesis that there are epiphenomenalons should be no less than $1 - 10^{-30}$. I omit the elementary proof regarding the size provision here.) But so also is Forrest's contention that only Lewis is under an obligation to believe the *actual* world infested with epiphenomenalons; by the independence of worldmaking principles and ontologies for worlds, and by the equiprobability of all worlds as candidates for actuality (both amply argued for earlier), the actual world w_a is in no way relevantly different from our arbitrarily selected world w_b under ersatzism any more than under MR.

We consider just one objection to this line of argument (other objections may be answered by means parallel to those employed in 2.4.4). The number of divisions into which we have split w_b is arbitrary, so probabilities derived from it are arbitrarily determined also. Quite so – but the arbitrariness was extremely kind to the opposition (a tone set in 2.4.4). A less arbitrary and quite a reasonable partitioning for present

purposes would be into regions the size that we presume the smallest epiphenomenalons to be. Anything larger, and we have been very generous.

Lewis is mistaken: he is not entitled to use Occam's Razor to support his belief that the actual world is free of epiphenomenalons, and it is a mysterious fact that he should want to do so, except to save his often-stressed but equally mysterious commitment to the actual world's being one of only a 'few' possible worlds at which materialism is true, along with many other received opinions.

Forrest answers Lewis's reply to him (Lewis, 1986a) as follows:

It might seem as if a similar inference could be drawn on *any* theory of possibility. However we conceive of them, the possibilities with epiphenomenalons outnumber those without. [...] [W]e can avoid the commitment to redundant entities, but at the cost of claiming just to know, presumably *a priori*, that redundant entities are unlikely to exist. And anyone, including Lewis, could make that claim. But it is an *ad hoc* claim for Lewis to make, since in his case it would be an *a priori* knowledge claim about one world in many. However if we claim that there is a difference in category between the actual and the merely possible, then this *a priori* knowledge claim is about all there is of a certain kind, and so is not *ad hoc*. (Forrest, 1986, pp. 22-23)

Lewis's claim is indeed *ad hoc*, but Forrest's would be similarly vitiated. In what sense are the possible epiphenomenalon-ridden worlds even *merely* possible, if it can be known *a priori* that the actual world has no epiphenomenalons? This is, in the nature of the case, the only way for those who believe epiphenomenalons are possible to arrive at a ruling concerning whether there are actually any epiphenomenalons! To know *purely by a priori means* (as it were) that the actual world has some nature *N* is to be in a position to rule out any genuine possibility of worlds with natures other than *N*, so Forrest's avowal is tantamount to claiming knowledge that epiphenomenalons are impossible. This is far from his intended result, and counter to the central premise he relies on in his argument against modal realism.

By a standard Copernican default assumption, *this* location is no special location with respect to epiphenomenalons, so it cannot be presumed that the number it has of them is the most 'special' number of all: 0. But my own view of all of this is that we have no warrant for a belief in epiphenomenalons at *any* world, using the notion of warrant established in 2.3.4. They are a spurious product of mere linguistic and imaginative extravagance. But the case has been elaborated so that we can now move on to an

argument that involves what is *observable* at our world, and that enables us to move closer to a use of that sharpening of the Copernican assumption that I have called the paranthropic principle, or PP.

2.5.3 *Second preliminary: the future irregularity objection to MR*

The future irregularity objection (FIO) to MR, like the epiphenomenon objection, uses a form of argument alleging the failure of induction under MR, and MR alone. Its origins are obscure; Lewis notes the objection, mentioning some names that may be associated with it, but then subsumes it under a more general objection concerning induction (1986a, pp. 116-7). Following Lewis's hints we find a quite workable version of it by George Schlesinger (1984, p. 11 ff.); and it is plausible that it was mentioned to Lewis by Robert Adams or Jack Smart. It has sometimes been attributed to Forrest, but appears not to be in anything published by him. In Forrest (1986, p. 22, note 17), in the context of introducing the epiphenomenon objection, we are directed to Lewis (1982, p. 23, note 5), and again there is mention of Adams and Smart. But at the Lewis location cited there is again little more than mere mention of Adams, Smart – and again Forrest. Leslie presents the objection (finding it 'decisive' against MR) in rudimentary form, without shedding new light on its sources (1989, pp. 97-98); and only latterly in more developed form (2001, pp. 26-30). In both of these analyses Leslie relies on what may be deemed an under-explicated (or at least problematic) notion of the 'ranges' of points in sets differing, to overcome Lewis's typical rejoinder based on equal cardinalities of sets. In the later piece Leslie also resorts to specific pantheistic features of his own view of reality to deal with what he sees as a parallel problem in his own commitment to a kind of multiverse (2001, p. 30); but we cannot follow him there, given the terms of reference of our present inquiry.

We also find versions in such places as Koons' defence of the cosmological argument for the existence of a god, again without attribution (Koons, 1997), discussed earlier. Regrettably, though, it seems not to be possible to locate a version of FIO fully articulated as a formal argument, for Lewis to have responded to specifically; for too long it was the stuff of recurrent independent invention, personal communications, remembered asides, and gnomic footnotes. But Pruss has very recently articulated rigorous and developed forms of it (for example in 2001a and 2001b), strongly

reminiscent of Schlesinger's less formal version (so much so that I suggest that my treatment of Pruss could readily be converted as a treatment of Schlesinger). Here is one of Pruss's versions, appealing to just one example of a way in which the future may fail to resemble the past, and to the vulnerability of MR at one of its crucial points – the indexicality of actuality:

- (1) Let D be a complete non-indexical description of the actual world up to the present (t_0) in non-future-involving terms. (Definition.)
- (2) D contains the claim that gravity has always held prior to t_0 . (Premise.)
- (3) Conclusions about the actual world reached by reasoning in accordance with the canons of inductive reasoning are justified, and in particular knowing that gravity has always actually held prior to t_0 justifies one in believing it will continue to hold after t_0 . (Premise.)
- (4) There are at least as many worlds satisfying D in which the law of gravitation fails after t_0 as there are worlds in which it continues to hold. (Premise.)
- (5) * Therefore, knowing that an entity w is a world satisfying D does not by itself epistemically justify inferring that w is a world at which gravity holds after t_0 . (Premise, justified intuitively by appeal to (4).)
- (6) * Theoretical reason is impartial with respect to merely indexical facts: If knowing that x is F (where F is purely non-indexical and x is a definite description or proper name) does not epistemically justify inferring that x is G (where G is purely non-indexical), then neither does knowing x is F and that x is I (now, here, etc.: any pure indexical will do) justify inferring that x is G . (Premise.)
- (7) * Actuality is indexical. (Premise.)
- (8) Therefore, knowing that an entity w is a world satisfying D and w is actual does not epistemically justify inferring that w is a world at which gravity holds after t_0 . (By (5)-(7).)
- (9) * But knowing that the actual world satisfies D and w is actual epistemically justifies inferring that gravity holds in w after t_0 . (By (2) and (3).)
- (10) Therefore, knowing that the actual world satisfies D and w is actual both does and does not epistemically justify inferring that gravity holds in w after t_0 , which is absurd. (By (8) and (9).) (Pruss, 2001b)

Pruss goes on to suggest which premise ought to be rejected in this reductio:

The premises marked with an asterisk form an inconsistent quadruple. All of them, except (7), are highly plausible, and hence we need to reject the premise (7) that

actuality is indexical. Another way to look at this argument is to see it as showing that if actuality is indexical, then inductive reasoning violates (6) and hence is guilty of the fallacy of partiality. But in fact we take inductive scientific reasoning to be a paradigm of impartial reason, and hence actuality is not indexical. (*loc. cit.*)

Now, Pruss's argument can be taken as a sort of *ad hominem* against Lewis, since it draws premises from both MR proper and LR (which of course I distinguish). Lewis's response would almost certainly have been a version of his general reply to such arguments:

I have no intention of becoming a sceptic. What we call 'inductive reason' is rightly so called; and I, as a modal realist, have no more reason to foresake [*sic*] inductive reason than anyone else has. I *do* have the reason that everyone else has; And I agree with common opinion that this reason is insufficient. (Lewis, 1986a, p. 117)

I agree with Lewis that the problem this style of objection raises may be everyone's if it is his: because of the equiprobability of all worlds being actualised that we can force on ersatzism, to align it with MR in the crucial respect. But I think that a combination of MR and LR – along with ersatzism and LR – does indeed have a more serious problem here than Lewis allows. Presumably Lewis could use some cardinality-based reply to weaken, though not to dispatch, the objection, like the one he deployed against the epiphenomenal objection (see 2.5.2). But I have already illustrated at length how such Lewisian replies can fail. Without going through the details a third time, I simply sketch how the moves might go for the present case.

Lewis says, perhaps, that there are *exactly* as many worlds at which gravity fails after t_0 as there are worlds at which it does not fail – the cardinalities are the same. This is just to get tough with Pruss's premise (4), and to speak of cardinalities where Pruss does not. So there is no reason to prefer the hypothesis that gravity will fail after t_0 to the hypothesis that it will not. At least there is no reason to believe that the future will be utterly irregular and unlike the past, which would be a *really* serious worry for induction. And the same reply can be made to arguments in which the manner of the future's varying differs from that given in the present case.

To this, I would answer that one can take some large number N of features that the physicists tell us are independent, like gravity holding as it always has *or not*, protons having the positive charge they do *or not*, etc., and apply my style of non-parametric

analysis to them. Though each individual hypothesis can have no more firm support than a probability of 0.5, the ensemble of independent hypotheses supporting future regularity has a probability of at most 0.5^N . In fact, if we believe in either MR or ersatzism, combined with some profligate 'Humean' principle of worldmaking like LR, the very low value 0.5^N should also be an upper limit of our level of credence that all of the N matters used in the argument will stay after *any* future time t_i as they stood before t_i . This is the style of response I used in 2.4.4.

Alternatively, I could answer just in terms of gravitation, and partition our location into 100 spatial regions for separate consideration, and then combine hypotheses concerning gravity continuing to hold or not at each of these regions individually (cf. the style of response in 2.5.2). This is justified against anyone believing in the sort of worldmaking principles assumed by Lewis and Pruss, since by those principles spatial regions, along with their contents and intrinsic properties, are as recombinable as anything else is. Again, the joint probability of all the hypotheses concerning all of the 100 regions supporting future regularity is 0.5^{100} . In both these ways of preceding, the principle of indifference is used in a harmless form, because it is used maximally conservatively.

(We should record here that these two strategies can be combined with perfect propriety, to make $N \times M$ hypotheses about N features continuing as they are now after t_1 at each of M regions. If we assign our usual conservative $P = 0.5$ to each hypothesis, we then achieve a much lower assignment of *a priori* probability of $0.5^{N \times M}$, for the whole suite of features being uniform and stable at our entire location simply by chance, assuming the usual permissive Lewisian recombination principles. Many other *higher-level multiplicative* strategies along similar lines are also available, and could have been used in earlier implementations of non-parametric arguments.)

Since I am convinced that it works *ad hominem* against Lewis (given Lewis's specific mix of commitments), I believe that Pruss's argument, at least in one or other of the strengthened forms suggested just now, works well enough as a *reductio* of something *interesting*. Let us suppose it to have been strengthened from now on, and take its premise (4) as much more potent than it is in the form Pruss gives it.

I disagree that it is his premise (7) that is the least plausible among (5), (6), (7), and (9). If we choose to work first with this quadruple that Pruss invites us to examine, I suggest

that (5) and (9) are less plausible. Let us examine these two, in turn:

- (5) * Therefore, knowing that an entity w is a world satisfying D does not by itself epistemically justify inferring that w is a world at which gravity holds after t_0 . (Premise, justified intuitively by appeal to (4).)

(4), invoked in support of (5), is an appeal to something like LR because we 'construct' the worlds at which the law of gravitation fails by recombining elements of the actual world (at which gravity has never been observed to fail) with elements of apparently similar worlds presumed broadly possible at which there is no gravity at all. Taken this way, we can see how (4) may 'intuitively' support (5). We may conveniently say, equivalently enough, that it is LR that supports (5). But if (5) is acceptable, it would take further argument to establish that (3) – the contention that induction works at the actual world – is true. It would be equally acceptable to argue against (3), using (5) as a premise. This is, after all, how the standard arguments against inductive success proceed. Our *only* evidence in the actual world for how the future will unfold is such as we have in (5), in which such evidence is said to be insufficient. All we know about our world (apart from the facts that it is concrete and ours) is that it is 'a world satisfying D ', yet we hold that this *does* 'by itself epistemically justify inferring' that it is 'a world at which gravity holds after t_0 '. If this is not so, we must indeed be deriving some extra knowledge from our world's being concrete, or from its being *ours*. But what could such extra knowledge be?

(4), if we believe it, might make us wonder why we have this confidence in future regularity in (3). This is, once again, the standard worry! But (4) ought, exactly equally, to have us worried about accepting (5). If we weigh the plausibility of (5) – which has it that there are many worlds at which induction fails – against the *competing* plausibility of (3), we have a choice to make. It is *actuality* that is supposed to make the difference in (3): but this is completely unargued. It is a mere assertion. Why could not a similar mere assertion be made against the counter-inductive claim in (5)?

Now, let us look again at (9):

- (9) * But knowing that the actual world satisfies D and w is actual epistemically justifies inferring that gravity holds in w after t_0 . (By (2) and (3).)

Here the assertion that it is actuality that makes the induction-supporting difference is

made more explicitly, but this assertion is just as unargued as before. (9) relies on (3), which has been brought into question above, and it is this reliance on (3) that vitiates (9).

(7) is a tenet of MR, but it could have been replaced by the following to be true under ersatzism:

- (7') The actual world is a selection for unique concreteness, with 'equiprobability', from all abstract worlds satisfying *D*.

This follows directly from three plausible sub-premises: the actual world does satisfy *D*, by definition; that this is all we relevantly know about the intrinsic nature of the actual world; and, by ersatzism, the actual world is the only concrete world, selected with equiprobability from all the abstract-world 'candidates' for actuality (argued for in 2.4.2).

(7) could alternatively have been replaced by the following to be true under both MR and ersatzism:

- (7'') The actual world is *either* indexically actual *or* it is a selection, for unique concreteness with 'equiprobability', from all abstract worlds satisfying *D*.

So if Pruss's argument is thought somehow to succeed as a reductio of (7), modified versions using either (7') or (7'') should be thought to succeed against these also. The modifications would have to compass a few *mutanda* in other premises; with (7'), for example, we would need this instead of (6):

- (6') * Theoretical reason is impartial with respect to *facts merely to do with concreteness and random selection from worlds with indiscernible histories* ... [etc.]

(6') is just another way of saying that the ways we should theorise about the histories of worlds is independent of their actuality or otherwise; it is a way of saying it that is adapted to ersatzism, as (6) is adapted to MR. If a generalised principle such as (6') or some (6'') were not true, very little use could be made of modal reasoning. We have to be able to consider how things would be if *only* certain limited changes were made in the world, without also being automatically burdened with induction failing at the changed world that we are contemplating!

It should be clear, then, that arguments formed using Pruss's template can be adapted to

work against ersatzism as readily as against MR. But what is really under attack, in fact, is LR itself, which is given expression in Pruss's (4), since it is invoked as the sole support for (5), which those of us who agree with Pruss that induction works in the actual world should, I think, want to reject.

2.5.4 *A pure paranthropic argument*

LR is very widely affirmed without fuss as an expression of a certain class of putative 'logical' possibilities. But we are taking a firm stand for *ontologically* based inquiry, and may not be impressed by such time-honoured asseverations concerning recombination. Let us examine the empirical evidence we have – the same empirical evidence as Lewis, Forrest, Pruss, and everyone has – from this new point of view. And, assuming MR for the moment, let us consider two competing hypotheses:

H1: Total reality is subject to (or well described using) LR.

H2: Total reality is constrained to be *much* more regular than LR permits.

And then let us look around at our world. *This* very token world, one of very many token worlds at which there are beings very similar to us – in fact, worlds sharing an indistinguishable history with us so far (or more correctly, up until a time corresponding to *this* time; there is no transworld time under MR). Now, under H1, we should not be surprised to see patches of sheer irregularity in various directions as we survey our surrounds. From moment to moment, we should positively *expect* to see odd areas of complete chaos. Such areas are neither conducive nor unconducive to our subsistence as reflective observers. Let's face it: 'most' (the caution here concerns cardinalities!) of our counterparts, indistinguishable from us up till now, *do* see areas of total chaos – very soon after *now*. Those of our history-sharing counterparts that *survive* beyond now, that is. Most do not, since in the vast majority of them chaos supersedes order even in their own bodies and beings, immediately after *now*. It is no use appealing here to 'closeness' of worlds, or such standard shifts: here we are talking pure ontology and its consequences. The domain of interest to us here is *all* worlds sharing our history (or, strictly, the *appearance* of our history) until now, and which of these worlds ours is must therefore, by MR (and by any 'sanitised' ersatzism, as I shall argue), be utterly unknowable. But lo: the world is, second by second, confirmed as a veritable paragon of

regularity: at least vastly more regular than it needs to be or than we have any right to expect it to be, if LR is correct. How can this be squared with our paranthropic principle?

It cannot be. But let us now return to consider H2:

H2: Total reality is constrained to be *much* more regular than LR permits.

This, the alternative to H1, is in perfect accord with PP. All the evidence we see, wherever and whenever we look, supports the conclusion that in this sample of total reality we call the world (or in this *location* within the world – it makes no difference) there is a superabundance of ‘lawful’ regularity; and this surely counts against the wildly speculative recombination principles.

So our rigorous, pure paranthropic argument might go like this:

- P1: Features that are no more conducive to the existence of reflective observers than the absence of them would be are representative of the whole of total reality, not just the observers’ own location. [PP]
- P2: We, as reflective observers, discover very many regularities whose presence is no more conducive to our existence than their absence would be. [Well-established empirical fact.]
- P3: Very many regularities are representative of the whole of total reality. [P1, P2]
- P4: According to LR, there are no regularities representative of total reality. [Directly from LR, according to which, ‘Humeably’, anything can coexist without anything else, in any spatiotemporal order.]
- C: LR is wrong; total reality is constrained to have at least the great majority of those regularities that we consistently observe that are not required for the existence of us as reflective observers. [P3, P4]

Now, this may seem to work only in the context of MR, and indeed we have assumed MR so far in order to fix ideas. But it applies equally in the context of ersatzism, since as we have seen LR is independent of ontology – and there are no moves in our argument that *require* MR. Under ersatzism, our world is just one of many that are conducive to our existence up until now, just as under MR; the addition of the inexplicable and strikingly lucky fact that this world alone is concrete, enabling *us* to be

concrete, is neither salvation for ersatzism nor a means of evading the conclusion of our paranthropic argument.

So Lewisian recombination – along with any similarly permissive recombination principle or worldmaking principle – is shown to be wrong, under MR or ersatzism.

Challenges are inevitable. Let us look at ten of them:

1. You consider MR and ersatzism, but what about Spinozism?

What indeed? Strange that the unique world should be so friendly, isn’t it? And *necessarily* so! But, as I have pointed out, if a strict world-based Spinozism is true, this is a result that will suit our purposes as we continue our inquiry in Part 3. And in any case, our result concerns LR and related principles. Nothing like that has application under Spinozism, or even – if we are to be strict – under fictionalist or other ‘neo-Spinozistic’ accounts of modality.

2. Worlds, or locations as parts of worlds?

The distinction between locations and worlds, which had seemed so important before, is here ignored because it makes no difference to the ‘Copernican’ result of our paranthropic argument. When we extrapolate paranthropically beyond the location that we observe, it makes no difference whether we extrapolate first to a larger world that includes this location, or directly to the whole of total reality. Regularities, observed at this location and having no bearing on our existence as reflective observers, may be presumed representative of *all* reality beyond this location. But if we extrapolated instead just to a ‘containing’ world, we could then immediately apply the argument again with respect to the world’s relation to total reality, having secured the required regularities in the world at large by our first application of the argument.

3. Cardinalities, and the principle of indifference, again?

Among worlds sharing this history up until now, the cardinality of worlds showing this superabundance of future regularity may be the same as that of worlds not showing it, etc.; and there could be related difficulties with the principle of indifference. But as soon as these arise we can have recourse to much the same sort of *non-parametric* approach that we have used earlier, and defended at some length. We could form an enormous number of multiple independent hypotheses (perhaps dividing the observed

world into time slices for separate consideration), assign probabilities to them independently, multiply the probabilities, and show that it is almost certain that, if LR-style principles were true, we would not experience the world as we do. Such is the potency of our transcendental methods, and we could defend their application here along the lines demonstrated earlier.

4. *But don't we observe irregularities?*

We do observe what might reasonably be classified as irregularities, if we look carefully enough: quantum indeterminacy is the best manifestation of it, perhaps. But in the large, the world is very regular indeed. Even if there were irregularities readily observable by the naked senses, there must still be a sufficient abundance of gratuitous regularity to fire our paranthropic argument. The very fact of our persistence as reflective observers at all, if LR were true, should be enough to astonish us utterly! If LR is true, then in the vast majority of worlds (a notion rendered proper by our non-parametric re-working) with this history till now, there is scarcely *any* orderliness to speak of, right after *now*.

5. *Why should we accept PP, anyway?*

P1 in my argument, a statement of the paranthropic principle (PP), is simply wrong? It would be hard to justify this allegation. It is just a statement of the Copernican principle universally accepted by cosmologists and statisticians as a *sine qua non* of systematic empirical inquiry. The fact that it is applied in unfamiliar ontological context is neither here nor there; if total reality is as MR or ersatzism hold that it is (and they differ only in irrelevant ways, or ways favouring MR as less arbitrary and as more reductive), LR is as subject to investigation as any similar principle applied *within* a world.

6. *May not all observed regularities be conducive to, or required for, our existence?*

P2 may be wrong because for all we know most of the regularities we discover are necessary for our existence in some hidden way. But it would be extremely difficult to mount a plausible argument in favour of LR with that among the premises! If the parameters' constant values, the laws of physics, and all the other guarantors of regularity are bound together in some such hidden nexus, this is virtually a rejection of LR all by itself. In any case, our argument need not, in the end, have made the distinction between life-relevant and irrelevant regularities. If it had been concerned

simply with *all* observed regularities, and if it were assumed that all were required for our existence, then if LR were true it should still amaze us that this whole weft and warp of regularities is sustained, every time we check to see that we still exist.

7. *What's this vague 'very many regularities' in P3?*

Is P3 insufficient for the conclusion, since there may still be many ways in which total reality can fall short of full chaos, as implicitly acknowledged by Lewis (with his response to the Forrest-Armstrong argument; see 2.2.3)? Such an objection fails when we remember that we have always been concerned with the whole class of LR-style principles, supporting a *predominantly* irregular total reality. Our argument was never intended to *prove* that total reality is utterly regular, nor do we expect the opposition to claim that it is utterly *irregular*. But our paranthropic argument pushes overwhelmingly in the direction of regularity and invariability. It gives us a sort of transcendental-empirical *warrant* for belief in the general, 'unHumean', overall regularity of total reality.

8. *But why not conclude that all of the observed regularities are there in total reality?*

This question can be taken two ways. First, why should our argument not result in the conclusion that all observed regularities of all kinds rigorously constrain total reality? This threatens to work as a reductio of our argument against LR, almost to the point of annulling anthropic arguments favouring variation. But this would simply issue in an implausible Spinozism, which we shall be happy enough with anyway (see the first of these challenges, above). Second, the question might be only about the observed *life-irrelevant* regularities. In that case, there is no malign consequence for our conclusion regarding LR; it would just be inexplicable that these regularities were invariable across all of total reality while those that happened to support our existence were variable. To the question taken either of these two ways we can also reply that our argument is still *inductive*, and so prone to error in at least some of its detailed results. But the result that 'very many regularities' are pervasive features of total reality is itself enormously robust.

9. *Where and what is Karakalpakstan?*

Karakalpakstan, or (in Uzbek) *Qoraqalpoghistan*, is an autonomous region of

Uzbekistan, next to the Aral Sea: *Karakalpak* means *black hat* – the traditional headgear. But if you're not going to be serious, we might as well just stop right now.

10. Wait!... This world may be one of those at which certain laws of nature hold in some strong non-'Humean' fashion, which is logically possible for worlds, isn't it? Some other classes of worlds may be subject to rampant recombination, but not worlds such as ours. May not a kind of natural necessity apply at such worlds, to which other worlds are not subject?

No. I really do think it's time we stopped. It is objected that the observed 'life-irrelevant' regularities are simply a matter of 'natural' necessity? We have to ask what this is supposed to amount to. This world *has* to be regular, but not all worlds do? There are 'logically' possible worlds at which our observed regularities do not hold, even if *this* world could not be like that (or perhaps *no* world could really be like that)? Such objections trade on an allegedly non-trivial but totally unexplicated notion of natural necessity that is, in the end, simply ruled out of court by our ontologically serious approach. What could it mean, after all, to say that *this* world is constrained to be in such-and-such a way, though other worlds are not? It is as if worlds themselves could vary modally – *at* other worlds! Even if such an unruly modal claim were acceptable, we could ask how we are to *know* that we are not at a world of another kind, which is not constrained in the manner that this world is supposed to be. If there are such worlds, and they are indistinguishable from *this* one up until t_1 , we cannot know that *this* one is not one of *those* ones. All of this follows inevitably if we just remember our assumption of MR with LR (for a 'paranthropic reductio' concerning LR), or of ersatzism with LR (also for reductio concerning LR). But as indicated earlier, 'restricted' modalities – natural and the like – will be exposed to closer scrutiny in Part 3.

So ends our paranthropic cartography of the worlds. There are no world-dragons, any more than there are sea-dragons (see the very beginning of Part 2). But we now turn to some loose ends, before a general consideration of these results restricting worldmaking principles, and of how the assertion of MR (on anthropic grounds) with the denial of LR (on paranthropic grounds) supports a new modal outlook.

2.6 CONCLUSION: A MODERATE NECESSITARIANISM

2.6.1 Some addenda

Other treatments of MR

Lewis has offered us a startlingly original and challenging proposal, which few have been able to take seriously for long enough to apprehend all of its weighty consequences. I have in effect, though, suggested that Lewis himself does not go far enough. Having reduced modality with devastating assurance, he still wants to be able to write like this:

I claim once again that I am within my rights to call the other worlds possible, not actual. If so, they give us no cases of actual inductive error, so they give us no inductive reason to distrust induction (1986a, p. 118)

But, while it is surely no mere ontology of the *actual*, the new ontology, once slipped free from its tether, will not be re-kennelled. If the actual is indexical, it is still a sample of the real – a sample that is not known in all its details. Lewis wants to finesse out of uncomfortable conclusions regarding induction, ultimately arising as I have argued from his permissive principle of recombination. But in resiling from the full-blown meta-Spinozist modal reduction that is at his command, to save cherished theory and pre-theory, he plays into the hands of less revolutionary philosophers. Holly Thomas (1993) may be one: her piece can be read as a thorough and sophisticated exploitation of such fence-sitting passages in Lewis as we have just seen. We do not examine the details, because we are, as inquirers into reality, about as interested in the precise doxastic and theoretical commitments of Lewis as we are in those of Hume. Thomas's painstaking critique works *ad hominem* against Lewis (with his commitment to the combination of MR and LR), as a great deal of good argument that we have noted also does; but its failure to address the ontology completely separated from the manner of worldmaking distances it from our central interests. Much the same can be said concerning the ground-breaking recent work of Pruss (2001a, 2001b), which we have had some occasion to analyse. None of these philosophers has seriously enough doubted the credentials of 'logical' modality, with its attendant free recombination principles. Or, to put it another way, they have retained the usual core *characterisation* of the broadest

possibility as 'logical'. But such a questioning has been our program in this part, and it will inform Part 3 as well.

With even greater reluctance do we forgo very detailed consideration of Scott Shalkowski's penetrating treatment of the same matter as we address (Shalkowski, 1994). He examines ontological foundations for modality, as we do. He finds a circularity in Lewis's MR: the restriction of the multiplicity of concrete worlds to *possible* worlds involves a primitive modality, which negates any worth in the effort to reduce modality in the first place; and he identifies the *same* circularity in ersatzism. But we might answer this way: if there were *per impossibile* only all the abstract worlds that ersatzism proposes (including as equal among them the one corresponding to the actual world), there would be no primitive modality in specifying some single world as actual. The situation would be, by extension from my arguments earlier, entirely isomorphic to total reality under MR. Shalkowski seems to want to equate MR and this sort of notional ersatzism we have just set up: both fill modal space as we should want the worlds to do, yet do not overfill it with *impossible* worlds. That this should be a problem for ersatzism-*sans*-actual-world is clear; and if it is a problem for that modified ersatzism, it is surely an equal problem for standard ersatzism, to add to the problem of a primitive actuality. But Shalkowski ignores this crucial, saving difference for MR: that for MR there is no objective difference between what we quotidianly distinguish as the actual and the possible. Modality truly is reduced; it vanishes from the picture altogether.

So I do not agree with Shalkowski's overall conclusion that there is indeed a primitive, irreducible modality in both MR and ersatzism; I find it only in ersatzism's utterly mysterious treatment of the *actual*. It seems to me that the conclusion relies on the same drive to re-find modality once it *has* been reduced before our very eyes that we discerned in Bigelow's work (see 2.2.5; and see below). This passage is indicative:

The justification of primitive modality has two components. [... First:] Reductions are doomed to fail because they end either in subtle arbitrariness or circularity. The second component involves showing that there is some point in working with the hypothesis that modality is primitive. If there is no such point, then perhaps the irreducibility of modality is evidence of its dispensability [*sic*]. However, dispensing with modality is not a viable option, since an adequate philosophical account of other phenomena requires a modal framework. (Shalkowski, 1994, p. 687)

Regarding the first point I say that there is only 'subtle arbitrariness' if one fails to appreciate the new and unfamiliar context in which concreteness is now to be viewed (see 2.2.4). If we have no problems with some presumed arbitrariness in the abstract objects of mathematics, there should be no reason to find problems with those concrete necessary entities that are Lewisian worlds and their parts. This has been argued earlier, and if Shalkowski succeeds in sowing doubts *now*, perhaps this shows simply that Hume (and his Philo) are very likely right: no human algebra is equal to the task. We do better at mathematics!

To Shalkowski's second point I reply that modality is indeed dispensable, when we adjust our gaze to look (unaccustomed and unequipped as we are intuitively) *sub specie aeternitatis* – beyond our necessary causal confinement to just one world. And this is, *pace* Shalkowski, a 'viable option'; any difficulties occasioned for other areas of philosophical inquiry are regrettable, but not the serious ontologist's fault. If MR is true, there simply *is* the ontology it postulates (by broad necessity, if you insist – which is *characterisable* as logical, if you must; but it is not compulsory to talk that way!). And the ontology does all that could be done to ground what there is of *de re* modality, even if it does not have direct and causal involvement in modal epistemology. How we come to know modal truths is a separate mystery, similar to how we come to know mathematical truths. But with neither of these two mysteries can we draw ontological conclusions from the impossibility of our causal interaction with the postulated objects. In the case of modally relevant ontology, we have our transcendental means of exploration. And we do get results; no argument from the inconvenience of these results, or their surprising 'unHumean-ness', is of consequence. We can survey the extended geography of total reality by means borrowed from statistical theory and cosmology – and draw conclusions that are limited but of surprising strength when we compare them to what we had thought available by mere 'logical' means.

Necessity or not necessity?

In the course of Part 2 (in 2.2.5) I have suggested that total reality is as it is as a matter of *Ur-necessity*; I then snatched the idea away again: 'There is really no *Ur-necessity* because there is no *Ur-contingency*.' This was a way of saying that we really *can* reduce modal talk to ontological talk; and when we have done so, we had better *keep* it

reduced, just as we had better not speculate about the putative liquid qualities of a molecule of H_2O (in a consignment of water at temperature 35° Celsius, and pressure 1 atmosphere) after we have seen that talk of liquidity is perfectly reducible to talk of the forces among molecules. (Cf. some Greek atomists' presupposition that the nature of atoms echoes in miniature the gross nature of matter as we observe it: smooth atoms underwriting the smoothness of butter, for example.) Consider 'necessity' in mathematics: *we would have no occasion to introduce the term at all if we stayed in the mathematical domain*. Necessity of worlds and their worldbound contents is as unbending as necessity of mathematics; in its proper context of MR or sane alternatives, it is simply concerned with the concrete rather than the abstract. *In fact, there seems to be no good reason to resist simply identifying the concrete with whatever is causally involved, and with whatever is worldbound*. And, since we are concrete, worldbound, causally involved objects ourselves, we fabricate contingency to deal epistemically and practically with our confinement to one world. If we did not do that, the notion of necessity would be completely vacuous for us, having nothing to oppose it. The concrete, viewed in the way I suggest we can view it, is quite as fully and broadly necessitated as the abstract is.

But the habit of mind that insists on modal talk as appropriate in all contexts of discourse, even 'reduced' and non-worldbound contexts, dies hard indeed. It therefore serves incidentally as an instructive parallel to another equally entrenched and pervasive habit of mind: just as we insist on manufacturing modal variation anew, even when it has been theoretically dispatched, many insist on *confabulating consciousness* as something escaping the net of physics, even though it be demonstrated that commitment to such a separate reality is unsustainable, given the causal closure of physics as we have argued that the term 'physics' should be understood. It is as if we were back with Locke denying that physics could brook action at a distance, in the face of experience of the behaviour of actualia.

But the task of exhibiting, analysing, and superseding the habit of confabulating consciousness belongs to what follows in Part 3 – where we shall illustrate a potent use of our newly regulated notions of causal, concrete reality.

An ethical sweetener

In 2.2.1 we considered a straightforward but intuition-disregarding response to those who object to MR because of its worrisome ethical implications: 'So much the worse for ethical theory! Metaphysics trumps ethics.' Well, it still does, I say. But now, having considered the matter of ontogeny (at 2.3.3) and how it may well be a far more modally restricted affair than is normally assumed, we can re-visit the ethical objection.

For the modal realist not committed to Lewisian recombination, but schooled instead by our paranthropic extrapolations from actuality, many scenarios that seem at first to be broadly possible (perhaps because 'physically' possible in some conventional sense) are open to serious doubt. Imagine a glass-blower, producing a smooth hot bubble of glass. Imagine the process involved. Next, set aside the whole process, and simply imagine a hot bubble of glass with surfaces that, say, bristle with sharply pointed and serrated surfaces. Is this a *possible* configuration for a consignment of hot glass? Well, any configuration of atoms is *possible*, you might say. But wait: is there a possible *history* that leads to the imagined serrated bubble? Is there a completely and coherently imaginable process by which such a thing could have been manufactured? Or – to take it right back – are you confident that you are imagining a *universe* (and so effectively a *world*), with *possible* initial conditions and *law-like* evolution, that will include in its history some such a serrated bubble? To take seriously our paranthropic reasoning requires being open to doubts about such things. At best we might only be able to say that we are unable to rule on the matter, or that we cannot see why such a thing should not evolve. But this is crucially different from saying that the thing is broadly possible!

Now, just as some alleged imaginings of creatures are, as a modern Lucretius might have put it, really no more than spuriously imagined time-slices of impossible whole creatures – so that there could not be a normally conceived, born, and raised centaur, for all we know – some imagined *social* scenarios may be mere plausible-seeming time-slices of impossible whole histories. There may be no corresponding history that could bring them about. We may assume that there are laws fixing the development of animals, even if they be probabilistic and approximate laws by the standards of hard physical science; they nevertheless are dependent on those lower-level laws. And, by parity, we may assume that there are laws fixing the evolution of societies, similarly

probabilistic and approximate, but equally grounded in lower-level laws. If we accept that the low-level causal laws are constrained in the way argued above, we must accept that animals and social situations are modally constrained also – perhaps in ways too complex for manageable investigation. At the very least modal realists of the moderate necessitarian denomination can be blissfully agnostic on this issue: such modal realists are not rationally compelled to believe that any proposed non-actual evil scenario could have evolved, so they are not rationally compelled to believe that they are possible – or, equivalently for them (and that's the rub), real. In fact, if we are to take our paranthropic reasoning seriously enough, we might conclude that most possible societies are much like our actual societies. Ours have evolved according to certain laws, and – bad as things may be – we could reasonably surmise that societies are pretty well constrained to be more-or-less as moderately brutish as ours are. So although every ethical situation that we are *genuinely* able to bring about does indeed arise in some world, and we are stuck with that fact, there are far fewer such situations and therefore far fewer such worlds to worry futilely about.

2.6.2 Summing up

In the preceding two sections I argued for two main results. The first was that MR (or alternatively – for those who refuse to countenance such a thing – an ersatzism rationalised to the extent that any ersatzism can be) was strongly supported by an application of anthropic reasoning (best construed as founded on the original WAP – Carter's *weak anthropic principle*). The second result was that LR and similar worldmaking moves are far too permissive, as shown by the application of a sharpened Copernican assumption borrowed from standard cosmology and from standard practice in statistical inference. That sharpening I called the *paranthropic principle* (PP), to elevate the Copernican assumption to equality with WAP. This second stage took momentum from the first, since our paranthropic argument worked best assuming the truth of MR (or a rationalised ersatzism, for the dissenters); but once that booster stage had been burned up, it could be jettisoned – if that's what we wanted to do with it. The result that we sought to launch into Part 3 is that *the range of genuine possibilities is far more restricted than LR, poorly construed 'logical' possibility, or anything much like these, would suggest*. That said, while not harbouring hopes of altering anyone's

personal canon of Moorean truths. I hope that a strong case has been made in MR's favour – at least as a theory to be taken more seriously than the famous and still-persistent incredulous stares suggest that it should be.

Setting aside their bearing on our reception of MR, reasoning from WAP and reasoning from PP work in mutually opposing directions. WAP tends toward acceptance of real, concrete, possibilities, and of great *variety* among real possibilities (the honorific 'real' may simply be assumed!); and PP puts a brake on this expansive tendency, suggesting far greater regularity and therefore far less variety. Earlier (in 2.1) I held out the prospect that 'we may be able to define limits or outlines to discipline our modal, Ur-physical speculations'. We are now in a position to do something of this sort, at least provisionally and subject to revision by the physicists, schooled by our two opposing principles.

I allow myself a little polemical overstatement in what follows, taking as my model the ubiquitous over-confident assertions in favour of various 'logical' possibilities. Some of these may be taken as ambit claims, but serious enough claims nonetheless, and supported well enough in the preceding sections of Part 2.

Assuming principled distinctions between physical laws, physical parameters (the 'constants'), and initial conditions, we can affirm what the physicists have for the most part believed without question or comment. By our *anthropic* argument we affirm the standard fine-tuning result: that the 'constants' are constant only for this world, or for this spatiotemporal part of this world. And by our *paranthropic* argument we affirm that the laws are pretty well constrained to be as physicists make them out to be – as a matter of the broadest necessity. So Ur-physics is likely to have as contents *our* physical laws, with at least some subset of the parameters taking quite a different suite of values at different worlds – and perhaps also at different *parts* of worlds, subject to the physicists' further inquiry into the matter.

(A caveat is required here. A scenario not so far canvassed is that there are indeed possible alternative sets of laws to those holding at our world, but just a finite few – incommensurably few relative to the infinitely many possible sets standardly assumed by Lewis and others. This is compatible with all that I have been at pains to urge, and it is sufficient for our needs as we move on. More strictly speaking, what has been clearly

ruled out are all the many highly irregular 'rubbishy' and composite worlds allowed by Forrest, Pruss, Lewis, Chalmers – I presume: see Part 3 – and just about everyone else, including worlds at which induction regarding the future fails. Now, in this 'several-physics' scenario we could simply and quite legitimately redescribe, so that the 'small' ensemble of sets of laws reduces to just one set of more complex and rather disjunctive laws, and no harm is done. In any case, and under either description, even if the laws are more numerous or complex than *direct* experience at this world suggests, we have secured an important result that we can work with as we proceed.)

The true basic laws of physics, then, are necessarily true. *Tout court*. Whatever these turn out to be (and that whole vast unsettled frontier of inquiry is not our territory *qua* philosophers), we have made a case for their being the laws of Ur-physics itself.

But if we accept this strong and most unconventional conclusion, we are now in a position to conjoin it with the conclusion of Part 1: so that whatever Ur-physics is, it compasses the principles governing all possible causal vicissitudes – at this world and all others. The joint outcome is this: *all* of total reality is utterly constrained to be much more like what we experience at our world than we had suspected. *Ur-physics is our physics!*

Recall that in 2.2.3 we looked at this from Lewis, when he had accepted the Forrest–Armstrong argument, according to which Lewis's worldmaking principles made too many worlds:

The only limit on the extent to which a world can be filled with duplicates of possible individuals is that the parts of a world must be able to fit together within some possible size and shape of spacetime. Apart from that, anything can coexist with anything, and anything can fail to coexist with anything. (Lewis, 1986a, pp. 89–90)

Recall also that we observed how Lewis was not able to take this any further, to say what those limits on size and shape of spacetime might come down to, in detail. There must be some more *particular* way in which spacetime is constrained, though Lewis and we cannot imagine exactly what this constraint might be. So now, we find by another powerful argument that the range of possible worlds is *somehow* limited: far more limited than the Forrest–Armstrong argument suggests, and perhaps not by any *findable* logical constraint. But we can conduct our paranthropic investigation as often and as

searchingly as we like, to bring our credence in such a powerful constraint on modal variability arbitrarily close to 1. This is, if not to find a clearly defined and logically guaranteed constraint, at least to home in on one!

The modern philosophical temperament, at least when it speaks English, rebels against this rebarbative proposal. What has become of our separate, broader, autonomous and supreme 'logical' possibility, that will soar above such impediments? *Surely we can conceive of a world at which our physical laws do not apply...* or what you will. But this is simply to refuse to take the ontological turn, which is the arch-principle appealed to as we began; and to refuse to countenance the challenge from our further inquiry.

Logical means are of course paramount in summarily excluding certain candidate worlds; and they must always retain that crucially important *triage* role. But I have argued at length that we can do more than that to illuminate the darkest recesses of reality, hidden from *direct* investigation by empirical means. Modern cosmology has suggested powerful new tools; we have deployed them in a new domain, and we have our results. Some may prefer to stay with footnoting the remarkable and penetrating insights (though finally of doubtful purport) in early writings of an eighteenth-century master philosopher (insights strongly qualified on mature reflection, as we see in our epigraph): David Lewis's acknowledged *miglior fabbro*, David Hume. Or we may move on, along with the sciences as they advance, to *new* methods of investigation and *new* conclusions. Perhaps this is the proper next step in a serious philosophical naturalism.

The outcome of our deliberations I have called a *moderate necessitarianism*. *Necessitarianism*, because the range of modal variation is severely restricted – far more than standard 'human-algebraic' or 'logical' considerations suggest, so that we should not be 'Lewisian recombinationists'; *moderate*, because as I have argued there is compelling reason to believe that there are other concrete 'possible worlds', so that we should certainly not be traditional Spinozists. But we, along with Lewis, ought to be meta-Spinozists: the ensemble of all the worlds is as it is by necessity *simpliciter*, if such a modal judgement, so worded, may be applied at all to worlds themselves – which it may not, in strict propriety! We contingent beings have trouble finding words that are apt beyond a single world. Once we have reduced or otherwise accounted for modality, it is improper to reintroduce it, as we must remind ourselves again and again. So we are

now, at last, ready to encounter what I will call the *dualist temptation* properly armed, though there will be still more to say about modal foundations as we proceed.

Part 3



The Dualist
Temptation

PART 3

The Dualist Temptation

"As I was saying," resumed the visitor – "as I was observing a little while ago, there are some very *outré* notions in that book of yours Monsieur Bon Bon. What, for instance, do you mean by all that humbug about the soul? Pray, sir, what is the soul?"

"The – hiccup! – soul," replied the metaphysician, referring to his MS., "is undoubtedly –"

"No, sir!"

"Indubitably –"

"No, sir!"

"Indisputably –"

"No, sir!"

"Incontrovertibly –"

"Evidently –"

"No, sir!"

"No, sir!"

"Hiccup! –"

"No, sir!"

"And beyond all question, a –"

"No sir, the soul is no such thing!" (Here the philosopher, looking daggers, took occasion to make an end, upon the spot, of his third bottle of Chambertin.)

— Edgar Allan Poe
(*Bon Bon*, 1832)

3.0 INTRODUCTION: UNEASE WITH MATERIALIST MONISM

So far I have sought to establish sound credentials for two fundamental and far-reaching proposals. The first is that materialism is so obviously true that, properly understood, its denial as a necessary thesis is close to an incoherence (Part 1) – so we have a justified presumption of *necessary materialism*. The second is that the most plausible and useful account of modality is a certain modification of David Lewis's modal realism, with a commitment to far fewer kinds of worlds (and so of worlds themselves) than Lewis himself allows – a thoroughgoing and experientially supportable reduction of modality that I call *moderate necessitarianism* (and if exactly that reduction should be

unacceptable. I have urged on readers the closest 'clone version' of it that their doxastic budgets will permit). These two proposals are closely enmeshed, the second making available evidential support for the first, which had seemed to be founded at best on mere historical, linguistic, conceptual, and definitional argument. In fact the combination of the two proposals makes for a strong axis indeed; and I could have left my apologia for materialism right there.

But there are dissenters to reckon with. The facts of consciousness are before (or rather, behind) our very eyes; surely, they will protest, these facts and the properties they have to do with are simply of the wrong *kind* to be either physical themselves or determined by the physical. Say what you will, they will insist, you have so far left out the most salient part of the world – the whole experiential domain, which is so intimately bound with us that it almost *is* our world.

What is it, if it is anything at all – this ocean of subjectivity seeming to surge within us? The question yaps and nips at our heels even as we stride forward at our most scientifically confident. There is the besetting worry that the story we tell will never come close to addressing the central datum: our own experience of whatever *else* it is that we choose to investigate. Pierre Bon Eon, the philosopher, restaurateur, *bon vivant* and – let us imagine – *materialist* in Poe's story, has recourse to his stock of superb wines; but if he had kept his nerve and his sobriety, he might yet have had some hope against the devil, and might not have been struck by an iron lamp loosed by his own angrily thrown bottle. (He should also have cultivated the ability to argue blithely on, unperturbed by the interruptions of his opponent – a more mundane but no less cardinal requisite for successful philosophising.) So in this third and final part of my thesis I face up to the devil, and meet squarely the temptation to dualism that all our experience continually whispers at us. At least I intend to show that we have excellent grounds for finding the dualist's claims unsustainable, and to give reasons for advancing steadfastly under our materialist banner.

Since David Chalmers offers a formidably detailed attack on materialist monism, I choose him as the pre-eminent champion of the dualist cause. Chalmers says so much, and so well, that I cannot address all the argument he brings to bear; but I can certainly show how the apparatus so far developed tells against his central dualist doctrine. Along

the way I shall have observations to make about the classic qualia objections to materialist orthodoxy, and about some of Chalmers' own engaging variants of these.

We work primarily with Chalmers' virtuosic and major work *The Conscious Mind* (1996). Chalmers is very prolific, and there is much else that could be cited: where it is more convenient, or necessitated by a change or addition of content, we shall make use of later works as well, including incidental mention of his substantial web-based output. But the essential position of 1996 remains unchanged, so that is our basic source.

First we should note that Chalmers' rejection of materialism is only concerned with the *phenomenal* – the qualia, the what-is-it-like-ness, the raw feel of experience. Indeed he normally reserves the word *consciousness* for what others might distinguish as *phenomenal* consciousness. I shall usually follow him in this: that is our primary concern, also. He is happy to go with the current of contemporary materialism concerning what he calls the *psychological*, which is all of the features of mind that are not *phenomenal* (like cognition, belief, judgement – even awareness, where this is a matter of cognitive or straight informational content). As a reasonable approximation the broad orientation is like that of Thomas Nagel (1974), and of an earlier Frank Jackson (1982).

Chalmers is no substance dualist. He does not hold with Descartes's conception of mind as different in substance from matter, discussed in Part 1. While he is committed to almost all features of the world being 'physical' (the scare quotes are also amply justified in Part 1) as Descartes was, for Chalmers the irreducibly non-physical remainder is strictly a matter of phenomenal *properties*. Even more explicitly and uncompromisingly than Descartes, Chalmers is committed to the complete causal closure of the physical. Every physical effect is fully determined by its physical causes, if it is determined at all. The physical sciences, idealised of course, can exhaustively explain all the events constituting the physical history of the world; but they cannot, even in principle, account for the entire nature of the world, because the facts of phenomenal experience are beyond their brief and beyond even their idealised competence. That is not to say that there can be no *naturalistic account*, or *naturalistic science*, of phenomenal consciousness. There could be, though we are nowhere near having such a science; Chalmers devotes the latter part of his book to a speculative

prospectus for such a naturalistic science of the phenomenal. In accord with these observations, he calls his whole program *naturalistic dualism*.

Many points in the foregoing summary would raise eyebrows (and we may question the coherence of some): Chalmers qualifies and hedges on a few of the key elements. This he does at length and brilliantly. Again, we cannot examine all of these qualifications: a literature devoted to them is fanning out from 1996 into the foreseeable future. Our essential task is more circumscribed: to apply the results derived in Part 1 and Part 2 to the case of Chalmers' particular potent brand of dualism.

While Chalmers' findings are not Cartesian, there is much in his technique of argument that is Cartesian, and much in the structure of his system that can be mapped onto the structure of Descartes's system. So striking is this that I am happy to call him, to use the term adopted in Part 1, a *Cartesian sympathiser*. The two systems have parallel points of weakness. I shall be arguing that some of those weaknesses are fatal.

Chalmers' claims against materialism work at two levels. First, at the logical and linguistic level, he sets up a 2-dimensionalist system of primary and secondary intensions. We shall address this matter only briefly, in the course of a first sally on ontological grounds at the level of worlds, which itself will require some more preparatory and definitional work.

3.1 MODALITY AND NATURE IN CHALMERS' DUALISM

3.1.1 *The taxonomy of modalities*

Chalmers, in common with most modern theorists, distinguishes various levels or 'strengths' of possibility and necessity. Having devoted the whole long Part 2 to modality, I have nevertheless deferred serious discussion of such differentiation, to deal with it in the context of Chalmers' dualism. There is a good deal concerning possibility as applied to *statements* in Chalmers, and the relations between that and possible-world talk are not always made perfectly clear. In this section, though, I address only possible-world modality.

Logical possibility

This is the broadest possibility of all (though see below, concerning epistemic possibility). Indeed, I have preferred to call it simply *broad* possibility, possibility *simpliciter*, or possibility *tout court*. At the level of worlds, Chalmers holds that *conceptual* possibility and logical possibility are the same. The story is reasonably well known: a logically possible world is one whose complete description entails no contradiction. This is a common characterisation of the broadest level of possibility, but as I have suggested in Part 2 it is a characterisation that I deem seriously misleading.

Metaphysical possibility

At the level of worlds, the metaphysically possible worlds are simply taken to be the logically possible worlds, but Chalmers sketches a distinction on behalf of certain dissenting theorists. They have it that there is a difference: there is no *metaphysically* possible world at which Hesperus is not Phosphorus (both of the terms being famous for rigidly designating a single entity – the planet Venus), though the identity of Hesperus and Phosphorus may break down somehow at the level of statements, beliefs, and generally anything based on what we are to understand as *primary intensions*. This is enough for some theorists named by Chalmers to claim that there is a *logically* possible world at which Hesperus is not Phosphorus. But Chalmers will have none of that. For him there is no such world, and for him the domains of logically and metaphysically possible worlds are coextensive:

[...] the oft cited distinction between "logical" and "metaphysical" possibility stemming from the Kripkean cases – on which it is held to be logically possible but not metaphysically possible that water is XYZ – is not a distinction at the level of *worlds*, but at most at the level of *statements*. A statement is "logically" possible in this sense if it is true in some world when evaluated according to primary intensions; a statement is metaphysically possible if it is true when evaluated according to secondary intensions. The relevant space of worlds is the same in both cases. (1996, pp. 67-68)

Chalmers identifies some theorists as accepting, overtly or covertly, the idea of *strong necessities* – associated, again, with an allegedly separate domain of metaphysically possible worlds. This strong necessity is more restrictive than logical necessity: there are 'logically possible worlds' that are not metaphysically possible worlds, because some 'logically possible worlds' (which by the account in question are properly not 'genuine' worlds at all) are those at which, for example, primary intensions alone can determine some of the propositions true at them. More directly relevantly in our context, in Chalmers' terms some say that zombies (beings completely the same as us physically, but lacking phenomenal consciousness) are logically but not metaphysically possible; or equivalently, that beings physically the same as us all have consciousness as a matter of strong necessity. Chalmers brooks no such differentiation, as we shall discuss. He has it that zombies, as described, are conceptually, logically, and metaphysically possible (all three amounting to the same, for worlds); so there are indeed worlds at which there are zombies.

Natural, 'real', or empirical, possibility

For Chalmers zombies are not *naturally* possible, though. This is the most problematic of the modalities he appeals to. It is problematic for easily stated but varying reasons: for me, because it is distinct in its ontological implications from *broad* possibility, and for others because it is not simply equated with *physical* possibility (dealt with next). Chalmers also calls it *empirical*, and on the evidence of some of the passages cited below he would perhaps allow it to be called *real* possibility (though I itch to put this adjective in scare quotes!) and *practical* possibility. Others label it *nomic* or *nomological*. Here are some relevant passages concerning natural possibility, which I name for easy reference:

- NP1: In the actual world, whenever there is a mole of gas at a given temperature and volume, its pressure will be determined: it is empirically impossible that two distinct moles of gas could have the same temperature and volume, but different pressure. [...] It is *logically* possible that a mole of gas with a given temperature and volume might have a different pressure; imagine a world in which the gas constant *K* is larger or smaller, for example. Rather, it is just a fact about nature that there is this correlation. (p. 36; cited and dealt with at length in Part 2)
- NP2: A naturally possible situation is one that could actually occur in nature, without violating any natural laws. This is a much stronger constraint than mere logical possibility. The scenario with a different gas constant is logically possible, for example, but it could never occur in the real world, so it is not naturally possible. (*loc. cit.*)
- NP3: To be sure, fading qualia are *logically* possible. There is no contradiction in the description of a system that is so wrong about its experiences. But logical possibility and natural possibility are different things. We have no reason to believe that this sort of case could happen in practice, and every reason to believe otherwise. (p. 257)
- NP4: This "dancing qualia" scenario may be logically possible (although the case is so extreme that it seems *only just* logically possible), but that does not mean that it is plausible as an empirical possibility, any more than it is plausible that the world was created five minutes ago. (p. 269)
- NP5: Now my zombie twin is only a logical possibility, not an empirical one, and we should not get too worried about odd things that happen in logically possible worlds. Still, there is room to be perturbed by what is going on. After all, any explanation of my twin's behavior will equally count as an explanation of *my* behavior, as the processes inside his body are precisely mirrored by those inside mine. (p. 180)

These extracts are very revealing about Chalmers' general approach to modality at the level of worlds, and I shall return to them in due course to show the profound difficulties this approach introduces – when we get *ontologically serious* about worlds, if not before. For now let us characterise the *naturally possible* (a modality primarily applicable to worlds and things at worlds, not to statements) as *that which is consistent with the laws of nature* (which laws Chalmers believes there are). Fading qualia, dancing qualia, and zombies will all be discussed later.

Now we examine NP4 in some detail. Given the context, and Chalmers' practice elsewhere, the second underlined statement ('the case is so extreme that it seems *only just* logically possible') cannot be read as meaning simply that the case is naturally or empirically impossible, and merely logically possible (though he does allege that also

about it). The statement is to be read as meaning that the case is so outlandish as to be toward the *outer reaches* of logical possibility. (Cf. Armstrong's use of a similar phrase with apparently similar intent: 'just barely possible'; 1980, p. 152, noted in Part 2, where I accused both Armstrong and Chalmers of *crypto-Spinozism*.) But what can this mean? Very little, I say. It is as if one were to describe some mathematical statements as *only just true* – as if a trans-googolplexically high integer might be described as *barely* finite, or a triangle with two of its angles equal to $10^{-100,000,000}$ degrees as being *only just* a triangle, teetering on the brink of being a mere line segment. It seems to me that for the modal realist, the absurdity of such *grades* of possibility is quite obvious. It would be like talk of worlds that only just make it as real or distinct (see Bigelow and Pargetter, in criticism of Lewis's modal realism; Bigelow and Pargetter, 1987) – *borderline real*, as we might put it, or perhaps *just scraping in* as distinct from some other world. In fact every possible world is equally possible, and equally distinct from all others – and equally close to being impossible, just as every integer is equally securely an integer, yet each sits so *perilously* close to being a mere transcendental. None of this is anything to marvel at! But Chalmers and Armstrong are not guided by the modal-realism heuristic, let alone committed to the literal truth of it; so they make such statements as we have seen.

The implications of NP4 for our understanding of Chalmers' views concerning world-level modality are worth pursuing a little further. What is this 'plausibility' in the third underlined statement in the passage ('but that does not mean that it is plausible as an empirical possibility'), applied to empirical possibilities? A statement surely ought to be either naturally (empirically) possible or not, for him. On a standard analysis, it is empirically possible that my first great-grandson should turn out to have red hair. But this is not to be confused with a related *epistemic* possibility (discussed separately below), which is itself quite properly and directly associated with talk of plausibility. So, again on a standard analysis, it is epistemically possible that my first great-grandson (if I ever have one) actually does turn out to have red hair (I haven't seen him: it's consistent with everything I know). What's more, the *statement* that he will have red hair is quite *plausible*. I inherited red hair from a great-grandparent, so why shouldn't he have and express the same genetic predisposition? But it is, on standard analyses and presumably on Chalmers', in accord with the laws of nature that he should have red

hair, so it is uncontroversially naturally possible – so its natural possibility itself is not merely *plausible*! This may seem like carping (or barking!) pedantry, but since Chalmers' whole case rests on the drawing out of fine divergences in modality, such concern for detail is not out of place.

We shall return to the problem of natural possibility when we consider Chalmers' account of modal ontology, and apply to it the results gained in Part 2.

Physical possibility

Physical possibility, common enough in the usage of most theorists, turns up very little in Chalmers' writing. But we know what he must understand by the term: the physically possible is whatever is consistent with the laws of physics. We must stress that for Chalmers the physical laws are not the only natural laws, and this has an interesting consequence. Strangely enough, zombies should count for him as *physically* possible. This is because the physical laws are included among the natural laws for Chalmers, but do not exhaust them; and it is only when we hold fixed both the physical laws *and* the psychophysical laws (the other natural laws to which Chalmers is committed – see later) that zombies get ruled out as not 'naturally' possible.

I, of course, am compelled to adopt a different view of the physically possible. I expand on this below.

Epistemic possibility

The state of play with epistemic possibility is quite unsettled in the literature (see Humberstone, 1996, for remarks on this, which the author bases on discussion with me). Some confuse it with doxastic possibility, about which nothing need be said here. Some say that it amounts simply to what we may call truth-for-all-I-(can)-know, so that 'Goldbach's conjecture is true' and 'Goldbach's conjecture is false' are both epistemically possible; or for someone in a hurry and without a calculator it is epistemically possible that 5.166476 is the cube root of 137.89 correct to six decimal places. In this sense, it *was* epistemically possible for us that the value of *K* was not logically constrained to have its actual value (once the units of measurement are fixed) in the equation for the ideal gas law, though I hope it no longer is for my readers – after the extended discussion in Part 2! (Note the use of two distinct kinds of modality in this

last example, which is often a haven for confusion.) The better way to understand epistemic possibility, according to Lewis (1986a, p. 27), is as genuine broad possibility subjected to restrictions. So it is never for anyone epistemically possible that 5.166476 is the cube root of 137.89 correct to six decimal places, since this is not even a possibility *simpliciter*, nor is it ever epistemically possible for the value of *K* to vary as Chalmers thinks it could. This does have the virtue of bringing epistemic possibility into line with other 'qualified' possibilities (in this case with a highly relativised and mutable qualification). It is another question whether one should accept Lewis's account of the epistemically *necessary* as known, even if only implicitly. (I strongly prefer Lewis's way with epistemic possibility, but I lean towards understanding epistemically *necessary* propositions as merely those entailed by all that one knows – again, as I claim, in conformity with other uses of restricted possibility terms – without one *knowing* all those entailed propositions. So I take it that the proposition correctly ascribing the truth value of Goldbach's conjecture is epistemically necessary for all of us, but not known by any of us. But there are complications that take us too far off track.)

So as we have just seen, in this case Lewis is uncharacteristically happy to surrender a commonly accepted understanding of a term (see discussion of his general unwillingness to do this, in Part 2). But Chalmers shows himself not so ready, as scanty evidence in 1996 and copious evidence in his web-based material (as extant in 2002) suggests. He argues in support of epistemic possibility along with most of its usual poorly regimented associations, in a manner akin to Lewis's accommodation of conventional ethical thought when by his own modal realist lights it would have been an expected and effortless matter to have ethical thought go the way the true metaphysics will dictate (see Part 2).

Chalmers' way of doing this is characteristic (though to be fair he acknowledges that it is rough and provisional): epistemic possibility is worked out, at what may strike some as tortuous length but with analysis that is ingenious and close-grained, in 'logistic' terms: primary intensions, epistemic intensions, epistemic space, ideal epistemic content, subjunctive intensions, epistemic dependence, etc, etc. But here we are to stay resolutely at the level of worlds, which in the end Chalmers allows as a legitimate

move, since the modal facts settling the issue of consciousness and its relation to everything else must in the end be a matter of truth at worlds.

Relations among the modalities

Lewis normally speaks of possibility *simpliciter*, only occasionally making reference to different types of possibility such as those displayed above. Such references are in terms of limiting the domain of possible worlds in some way or other, as we have seen for Lewis's brand of epistemic possibility:

Sometimes one hears a short list of the restricted modalities: nomological, historical, epistemic, deontic, maybe one or two more. And sometimes one is expected to take a position, once and for all, about what is or isn't possible *de re* for an individual. I would suggest instead that the restricting of modalities by accessibility or counterpart relations, like the restricting of quantifiers generally, is a very fluid sort of affair: inconstant, somewhat indeterminate, and subject to instant change in response to contextual pressures. Not anything goes, but a great deal does. (1986a, p. 8)

The point is that modal realism automatically gives possibility as a matter of truth at or concerning some world – a world that is a real thing in just the same way as our world, along with *every* world, is a real thing. There is therefore strictly no need to talk of *logical* possibility, or of some more restricted *metaphysical* possibility. Each of these, properly construed, could only amount to *real* possibility (where 'real' is here used in what I consider its best sense, not importing anything to do with a restricted 'natural', 'empirical', possibility). Each possible world is possible *simpliciter*. Its similarity to our world or any other world has no bearing at all on its ontological status – which may be taken as a lemma in the over-arching argument of Part 2. I have endorsed this claim for my modified modal realism. I have argued also that any halfway coherent variety of ersatzism must endorse a version of it as well – except as regards the mysterious ontological difference supposed to hold between the single actual concrete world and the vastly numerous abstract merely possible worlds.

It is for this ontological reason that modal realists and their fellow travellers need not concern themselves a great deal with the different strata or nested sub-domains of possible worlds. The boundaries, as we have seen Lewis point out above, are, let it be stressed, 'a very fluid sort of affair: inconstant, somewhat indeterminate, and subject to instant change in response to contextual pressures.' For Chalmers, things are not as

Lewis describes them here – or at least they *had better not* be like this if Chalmers' arguments are to succeed, as we shall discuss shortly.

Before we go any further, it is timely to note that there is a certain alternative understanding of the term 'logical possibility' that does permit a distinction between metaphysical possibility (as genuine and unrestricted possibility) and a more permissive *merely* logical possibility. Such a view of logical possibility calls to mind the alternative view of epistemic possibility I mentioned before, according to which something that is not genuinely possible is nevertheless epistemically possible. This is akin to the manner in which someone who is *apparently* suitable for a job is not *genuinely* suitable for a job, or a *suspected* criminal is not a *genuine* criminal. I call 'apparently' and 'suspected' as they are used here *privative qualifiers*. Lewis does not, as we have seen, favour any such 'hyperbolically permissive' account of logical possibility; though from the examples I give of Lewis's profligate worldmaking in Part 2 he may be thought to slip into use of it occasionally, so vitiating the claims of some of his worlds to be *real*. I say that Chalmers is guilty of lapsing into acceptance of it very often, through his logicist, non-world-based characterisation of broad possibility. I find evidence in the excerpts NP1 to NP5, especially in NP5.

An interesting and important application of this idea to my own account of the relations between modalities as applied to worlds, shortly to be summarised and contrasted with Chalmers' account, is that unexpectedly many eminently plausible candidate possibilities are not genuine possibilities at all, though most people would take this judgement to be clearly wrong. One example considered in Part 2 (at 2.6.1) was the hot bubble of blown glass, with a shape for which we very likely could find no *complete* law-abiding history, which I on that account claimed may, for all we know, not be genuinely possible. Even I might sometimes be *tempted* to declare that it is physically possible (or indeed, for me – simply *possible*), because its description seems to fit with many reasonably specifiable 'physical laws'. But for me, physical laws as standardly given are just certain approximate ways of picking out kinds of regularities that are in the end determined as a matter of the broadest necessity. So such talk as this is loose and 'hyperbolic', on my account. In fact we cannot even say with certainty that *any* very specific configuration for objects or *any* very specific event that we speculate about

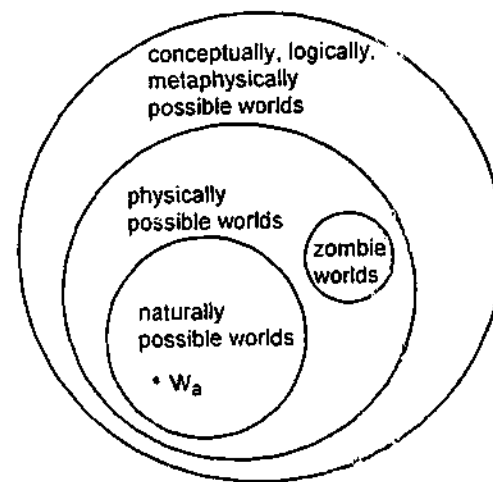
beyond the actual and observed is genuinely, broadly, possible. I cannot even say that some easy posture of the body that I can imagine myself adopting is one that I do adopt in some possible world: if I never actually adopt it, I do not know that there is any broadly possible complete history of a universe, arising out of some possible initial conditions, in which any counterpart of mine (to use the Lewisian dialect) ever comes to adopt that posture! This shows just how narrowed the range of possibilities becomes on my account of modality at the ontologically serious level -- at the base level of worlds. All of this is a clear consequence of the arguments of Part 2, just as a complete collapse into a classic Spinozism (which seemed to threaten just now) is clearly avoided by those arguments. Necessitarianism, sure enough: but a moderate, ontologically informed and experientially sustainable necessitarianism.

I say that these points are important partly because they give a way of drawing together the relations between various notions of the logically and the epistemically possible. The serrated hot glass bubble, the unadopted posture, all specific features that seem unimpeachably possible even though not actual -- it is epistemically possible in the 'hyperbolic' sense introduced above that there are worlds at which they feature. This is a sense that Lewis and I don't like. But they may not be 'logically' or broadly possible in the sense Lewis, I, and Chalmers' (when he is careful!) want. Some of them are 'logically' possible only in the hyperbolic, ungenuine, sense of *this* term, now: and Lewis and I reject that disreputable sense. My allegation against Chalmers is that he slips into acceptance of it.

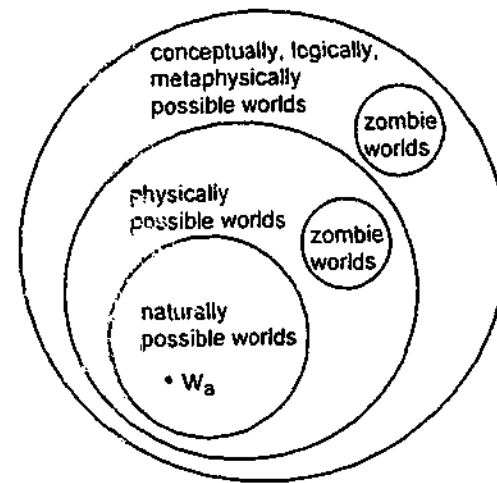
One last way of putting this matter. All sorts of things are hyperbolically epistemically possible with respect to acknowledged laws of physics. Consider the inverse square law. A certain conjunction is *plausible* (that is, at least hyperbolically epistemically possible -- we don't know, or perhaps can't know, that it is not true): 'the inverse square law holds and a big bang never actually happened' is plausible. From this plausibility we think we can conceive of worlds at which there is no big bang and the inverse square law holds. Ours might be one such world, for all we know. But the step that I would forbid is from this to saying that such worlds are epistemically possible in the *other* sense -- the well regimented sense in which the epistemically possible worlds must constitute a subclass of the broadly possible worlds. We simply cannot assume that the

inverse square law is *consistent* with the absence of a big bang (or *vice versa*), any more than we can assume that Goldbach's conjecture is consistent with the absence of a big bang.

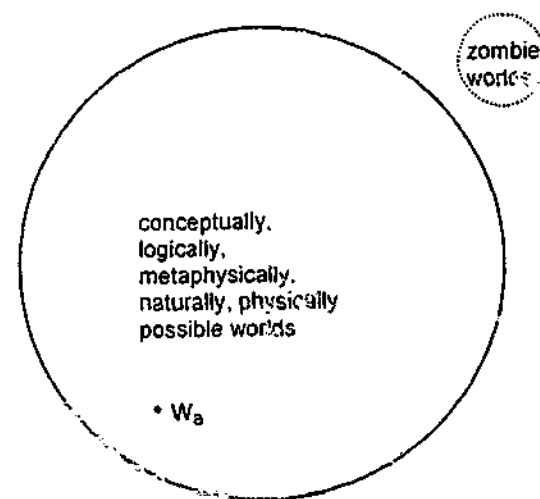
A: A way of dividing the space of worlds derived from Chalmers' account (locating the zombie worlds):



B: An alternative way for Chalmers (some zombie worlds not physically possible):



C: My preferred way (zombie worlds are not possible):



D: A way that I might also accept (zombie worlds are not possible):

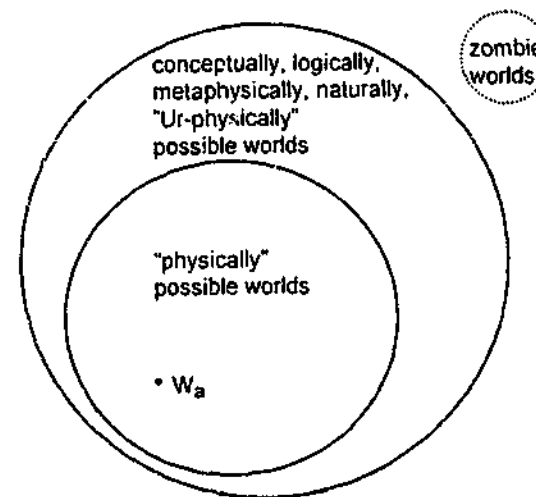


Figure 2 Four ways of dividing the space of worlds (showing our world as w_a).

(This last case is much easier than our example: Goldbach's conjecture may be a necessary falsehood, for all we know, and therefore may not be consistent with anything at all; but for our example, there may indeed be no worlds at which the universe does not start with a big bang – we just don't know! If it were so, then the absence of a big bang is inconsistent with whatever we would conjoin it with.) Chalmers is of course aware of a good deal of the problem here:

Arguing for a logical possibility is not entirely straightforward. How, for example, would one argue that a mile-high unicycle is logically possible? It just seems

obvious. Although no such thing exists in the real world, the description certainly seems to be coherent. If someone objects that it is not logically possible – it merely seems that way – there is little we can say, except to repeat the description and assert its obvious coherence. It seems quite clear that there is no hidden contradiction lurking in the description. (1996, p. 96)

Quite so: it just seems obvious. It seems that there is no hidden contradiction lurking in the description. But a *complete* description is needed to decide the matter, and that would have to be a *complete consistent causal history for a world* that includes a mile-high unicycle. Lewis and Chalmers surely are ready to believe that there are such consistent histories, and so that there are such worlds. This is because Lewis is committed to Lewisian recombination, and Chalmers is committed to a similar profligacy in worldmaking. For reasons made clear in Part 2, from my paranthropic arguments that severely constrain worldmaking, I must remain firmly agnostic about there being any such worlds. This is consistent with the proposition that people could reasonably plan to construct a mile-high unicycle; and it is consistent with the proposition that a mile-high unicycle is *plausibly* there as a feature of some worlds: it is surely 'epistemically possible' for us all that the worlds include such a thing, but only in the hyperbolic sense established above, which I, following Lewis, prefer not to use.

In Figure 2 I summarise two ways Chalmers might settle on for dividing the space of worlds, to show the relations between the types that are of interest; I also show (for later discussion) the location of zombie worlds. Below the two Chalmers ways (A and B) I show on the left my own preferred rather simple way (C) of understanding the worlds, using the same terms as Chalmers does, to enable comparison. On the right I show a way (D) of dividing the space that I would accept if it were to turn out that there is more than one possible set of standardly conceived 'physical laws', and we were to prefer this way of talking to the alternative that would use a single, more complex, set of disjunctive laws (see earlier, in 2.6.2). The two ways shown on the left (A and C) are to be taken as the more straightforward consequences of the two competing theories of modality; those on the right are shown merely for completeness.

(In both A and B I show *zombie worlds*, rather than just a putative unique zombie world. I take it that there are, in Chalmers' scheme, many worlds that meet his usual requirement for a zombie world: 'physically identical to ours, but in which there are no conscious experiences at all'; 1996, p. 94. From a good deal of what he writes and from

his general stance on modality it is clear that Chalmers accepts – and *must* accept, since he accepts zombies and the like – the ‘logical’ possibility of all manner of non-physical objects – say Forrest’s epiphenomenalons; see Part 2. And each of these may variously inhabit or not inhabit the zombie worlds, so ensuring the variety and distinctness of worlds that lack phenomena but that do not differ in any physical detail from our world. The difference in B is justified by this from Chalmers: ‘What is going on with my zombie twin? He is physically identical to me, and we may as well suppose that he is embedded in an identical environment [sc. a physically identical environment; otherwise there would be non-zombie persons in the environment of zombie Chalmers]’; 1996, p. 95. I am simply allowing for cases in which instead we do *not* hold the environment of zombie Chalmers identical, physically or *tout court*, to that of Chalmers, and furthermore allow parts of the environment that are effectively out of relevant causal connection with all the zombies not only to vary, but to violate the laws of our physics.)

3.1.2 *Modal ontology and Chalmers*

In the preceding subsection our focus was on the hierarchy of kinds of worlds, as characterised by the species of possibility associated with them. But now we need to look closely at the ontological basis for the worlds. We have seen that Lewis holds that all possible worlds have equal ontological status; all are equally real, regardless of similarity relations with our world, or among themselves. I am very much inclined toward this view, for the reasons presented extensively in Part 2. My issues with Lewis were not to do with ontological foundations, but with the *range* of kinds of worlds we should think there are.

Now we look at Chalmers’ ontological account of the worlds he uses in his argument for dualism. As we do so, there will be things to say about other matters of interest to us, such as the nature of scientific inquiry and of rational belief.

Chalmers in fact has little to say about the ontology of worlds. *Prima facie*, this is surprising. He gives a great deal of argument toward his central conclusion that phenomenal consciousness is a matter of non-physical properties – an ontological conclusion about phenomena and properties. He uses the possible-world idea in the

course of that argument, and holds that the extensive and intricate discussion regarding possibility in the domain of statements must be connected with underlying truths concerning worlds, as we saw before:

A statement is “logically” possible in this sense if it is true in some world when evaluated according to primary intensions; a statement is metaphysically possible if it is true when evaluated according to secondary intensions. The relevant space of worlds is the same in both cases. (1996, p. 68)

But the basic ontological standing of worlds themselves he deliberately leaves unexamined, as we see in one of his mentions of the topic. I quote the passage, first with some of its context:

A statement is logically necessary if and only if it is true in all logically possible worlds. Of course we have two varieties of logical necessity of statements, depending on whether we evaluate truth in a possible world according to primary and secondary intensions. We might call these varieties *1-necessity* and *2-necessity*, respectively.

We shall have little more to say about this matter of logical necessity of statements, in our ontologically based study. Chalmers continues:

This analysis explicates the logical necessity and possibility of a *statement* in terms of (a) the logical possibility of *worlds*, and (b) the intensions determined by the terms involved in the statement. I have already discussed the intensions. As for the notion of a logically possible world, this is something of a primitive: as before, we can intuitively think of a logically possible world as a world that God might have created (questions about God himself aside). I will not engage with the vexed question of the ontological status of these worlds, but simply take them for granted as a tool, in the same way as one mathematics for granted. As for the extent of the class, the most important feature is that every conceivable world is logically possible, [...] (1996, pp. 65-66)

And there is an important endnote annexed to this passage. Here is the relevant part of it for us:

Worlds should be seen prelinguistically, perhaps as distributions of basic qualities. Worlds are probably best not seen as collections of statements, as statements describe a world, and we have seen that they can do so in more than one way. To regard a world as a collection of statements would be to lose this distinction. Perhaps worlds can be regarded as collections of propositions (Adams 1974), if propositions are understood appropriately, or as maximal properties (Stalnaker 1976), or as states of affairs (Plantinga 1976), or as structural universals (Forrest 1986), or as concrete objects analogous to our own world (Lewis 1986a). In any case, talk of possible worlds is as well or poorly grounded as talk of possibility and necessity in general. As with mathematical notions, these modal notions can be

usefully deployed even preceding a satisfying ontological analysis. (*ibid.*, p. 336, note 30)

We turn now to an important later piece by Chalmers (1999), in which he answers objections from a number of critics. Stephen Yablo (1999), as Chalmers interprets him, had suggested that zombies are logically but not metaphysically possible, so that there are no 'logically possible zombie worlds' (Chalmers, 1999, p. 480). Chalmers, though, claims that 'if zombies are ideally conceivable, we need logically possible zombie worlds'. He invites us to consider a certain "strong necessity" view, on which the only possible worlds are those whose laws of nature are our laws'. (This is one characterisation of my own view, perhaps – but it is not a good one. I have given evidence and argument for the position that the true and broad 'laws of nature', if we must call anything that, are 'logically necessary', but I therefore think that there is *strictly* no need to speak of laws of nature at all.) Then, writes Chalmers: 'On such a view, counternomic scenarios, although conceivable, are reflected in no metaphysically possible worlds at all. But we still need *logically* possible worlds with different laws, for all sorts of reasons.' (And some of my arguments in Part 2 are intended to refute such a view as that: there are no real 'logically' possible worlds at which the regularities are different from what we may choose to call the 'true Ur-physical laws', which are really constrained by broad 'logical' necessity, even if we can't know the details, as shown by strong empirical, inductive, paranthropic arguments.) And those reasons? Chalmers continues:

Think of the reasons why "possible worlds" talk is introduced into philosophy in the first place. Possible worlds are introduced to deal with counterfactual thought, the semantics of counterfactual language, rational inference, and the contents of belief, among other reasons. A scientist can think counterfactually (and rationally) about scenarios with different laws, and can make true utterances about these scenarios. If we are to use possible-worlds talk to characterize the contents of her beliefs in discovering laws, we will need to appeal to counternomic worlds. Without counternomic worlds, we will not be able to use worlds to make sense of her inference processes. And so on. Ruling out counternomic worlds will make possible worlds useless for many or most standard purposes. Even if someone insists that such worlds are not metaphysically possible, we need logically possible counternomic worlds. (1999, p. 481)

Concerning the last underlined portion, I say that what we need is talk of plausible *candidate* worlds – worlds that we expect there to be, or that we cannot yet know not to be possible. Let's indeed call them *scenarios*, as Chalmers earlier does. Now, I would

say that progress in science is achieved by coming to know that some of these candidates – some of these scenarios – are indeed not represented in genuine possible worlds (or are not *likely* to be so represented, perhaps). This is how we 'discover laws'! And scientists may do this by starting with the only world we have epistemic access to – our own – and attempting to make parts of it match the scenarios they envisage. Failures to achieve such matches count as good evidence, like my paranthropic evidence (which is after all based on standard scientific practice) about the nature of worlds and therefore of 'laws', which is what scientists like to call the principled and robust regularities they discover. They do not need *counternomic worlds* at all; and I have argued that there are none. Chalmers continues:

Something similar applies to zombie worlds. Even on a type-B materialist view [*sc.* a view like Yablo's], we can think counterfactually (and rationally) about the possibility of a different distribution of phenomenal properties with the same physical properties. We need worlds corresponding to these possibilities to make sense of counterfactual thought, of the semantics of counterfactual utterances, of rational inferences involving consciousness, of the contents of rational beliefs about consciousness, and so on. We can write coherent science fiction about zombies, and speak coherently about the truth in such fictions. Talk of logically possible zombie worlds is justified in the usual way by their role in these uses. (*loc. cit.*)

Again I give short responses, justified by preceding argument (or in the case of zombies, argument yet to come). Yes, we can 'think counterfactually (and rationally) about the possibility of a different distribution of phenomenal properties with the same physical properties,' but when we do so we may not be thinking of ultimately coherent scenarios – of possible worlds. Just as we can argue *per impossibile* (and sometimes usefully, if we are very careful: see my arguments using *ersatzism-sans-actual-world*, in Part 2), so we can speculate about all manner of notional recombinations and dissociations. Our ability to do this does not warrant belief in the existence of corresponding 'logically' possible worlds: not if we are cautious and ontologically serious about worlds. Yes, we can write coherent science fiction about zombies: but the zombies of fiction are usually very different from Chalmers' 'philosophical' zombies (as he himself tells us); and where they are not, coherence breaks down – as it does occasionally in much good fiction – even if we are unable to spot this breakdown! I further suggest that the responses I have just offered show that we should not be convinced concerning the final underlined portion: it is not talk of 'logically possible zombie worlds' that is so

justified, but the continued use of zombie scenarios, to explore whether they do in fact represent possible worlds at all. We shall do some exploration of this sort later. But now, here at last are the ontologically interesting statements, which follow immediately upon the preceding quote (pp. 481-2). I interpolate my comments:

I suppose one might resist by holding an extreme modal-realist view (even stronger than David Lewis's) on which possible worlds are simply "out there" and on which their connection to the rational uses in question is neither here nor there.

I don't see that the position apparently envisaged here would be 'stronger' than Lewis's, unless he means that the concrete worlds are somehow more constrained than Lewis would have them. Not much hangs on this point, perhaps. But I do think that it would be a significantly mistaken characterisation of Lewis's view (or even of any view that is somehow 'stronger') to hold that his 'possible worlds are simply "out there"'. By this I assume Chalmers means that the Lewis worlds are simply *contingently*, simply *fortuitously* 'there' (talk that may be apt for things *at* worlds, but that cannot be for worlds themselves). Or perhaps they are to have their existence mysteriously guaranteed by something akin to an unexplicated 'natural' possibility for worlds, beyond which possibility there is more modal space that the worlds neither fill nor account for. Lewis's reduction of possibility is a full reduction of *all* genuine possibility. That is its whole motivation, and I have argued that, ontologically, it is a very acceptable reduction. I also pointed out at the end of Part 2 the need to remind ourselves of this! So it is difficult to be sympathetic, perhaps, to Chalmers' continuation:

But even on such a view, we would end up having to postulate worldlike objects ("ersatz" worlds, at the very least) for the rational purposes.

No, because we do not have direct causally based epistemic access to any world beyond our own world, anyway – not on Lewis's account, nor on any major theorist's. We work with plausibilities and scenarios, as discussed above. Anthropic and paranthropic evidence and argument are about all we can hope for, as a surrogate for evidential access to other worlds. Such reasoning is transcendental, and of course strongly inductive. But even it does not issue in epistemic access to the worlds in an interesting *causal* sense: all the evidence is still at our world.

And on the dominant view on which all counterfactual worlds are regarded as ersatz abstract objects, there seems no ground for resistance. It is easy to construct an ersatz object that behaves in just the way that a zombie world should.

(So much for the precise ontological account of worlds making no difference, as we have seen Chalmers suggesting!) No, the ersatz possible worlds, if they are maximal states, maximal sets of properties, maximally consistent sets of propositions, or what you will, are not easy to set up. To think so is to fail to take the qualifier 'maximal' seriously.

The obvious strategy is to use maximal consistent worldbooks, where "consistent" is understood in the *a priori* sense. One has to tread carefully in handling two-dimensional phenomena and centering, but the matter is straightforward.

The matter is not straightforward: however 'consistent' is to be understood, we have to aim at the broadest possibility, and always at the level of worlds. Nothing in language can deliver the final verdict on that. Logical consistency (a 'linguistic' matter, perhaps) and a commitment to a permissive recombination principle (extended to allow all manner of alien properties) were sufficient and sound principles for setting up Lewis's plenum of worlds, it had seemed. And Chalmers endorses at least that extent of prodigality. But I have adduced argument that Lewis's plenum overflows! Chalmers' plenum must also, *a fortiori*. It is not easy to set up genuinely possible worlds – concrete, abstract, or fictional. It is easy to imagine wild and interesting scenarios, which we may then explore for inconsistencies, if that is our interest. The fact that we do not *succeed* in finding inconsistencies in a scenario does not entail that it fits into any of the fully articulated histories that some genuinely possible world has.

Neo-Spinozism

We have surveyed a representative sample of Chalmers' comments bearing on ontology and worlds, and I have sniped using ammunition collected in Part 2. The interim summary may be presented like this: Chalmers seems to be committed to ontological 'class-distinctions' among worlds. It seems that the 'naturally' possible worlds (those that conform to laws that are supposedly contingently true, at our world) have deeper ontological roots than the other worlds. There *are* those other worlds – but in a sense there also are not. We see hard evidence for this neo-Spinozism or crypto-Spinozism (terms that I have explained in Part 2) in our next two sections. Chalmers does not, I

shall argue, univocally allow the genuine possibility of worlds that are not 'natural': and it will follow from this that he must believe they are not as 'real' as the 'natural' worlds. I shall attempt this first using Chalmers' preferred apparatus – supervenience theses.

3.2 SUPERVENIENCE: WORLD-TALK OR WORD-TALK?

3.2.1 *Some history*

Nothing characterises contemporary philosophy of mind so fittingly as its disputations over whether, when, how, and how completely facts at one level determine facts at some higher level. Even to speak of lower and higher levels suggests some concinnous sort of a hierarchy, and if this is not to be an arrangement according to supervenient dependency relations, we are entitled to ask the speaker for an alternative way of understanding it. The intricacies of supervenience-talk are legion, and cannot be dealt with comprehensively here. But some preliminary classification and criticism are necessary before we can explore how materialism is served handsomely by this broad integrating device – welcome in the thought of our time, confronted as it is with the accustomed 'cements of the universe' cracking and coming unstuck, at so many joints. Chalmers makes heavy use of versions of supervenience, but I shall suggest that his uses are ill-founded, since they are not founded in a sound ontological account of worlds, as suggested in the preceding section.

According to Jaegwon Kim, use of the notion of supervenience, but not of the term, began in ethical theory with GE Moore and others (Kim, 1993, pp. 136-8). Kim notes also that the term itself was occasionally used by the British emergentists in the 1920s, while its cognates are used in other senses by much earlier theorists, not only philosophers. He remarks particularly on a use by Leibniz (in Latin), and comments: 'Leibniz's use of "supervene" in this context seems not inappropriate in our light [...] I have not found any other [similar] occurrence of the term since then, until we come well into the present [sc. the twentieth] century' (*ibid.* p. 136). I cannot forbear to record here occurrences of 'supervening', deployed in strikingly modern fashion in the course of a discussion of the then young and exciting theory of epiphenomenalism, that seem to have eluded the historians:

But what then is the force of the 'epi,' and what becomes of the primacy of the materialistic terminology? The tables seem to be completely turned. What we see and feel, the facts of perception, become the real phenomena. Instead of states of consciousness supervening upon certain motions of mass-points or some peculiar complex of ethereal vortices, these motions, etc., prove to be but ideal conceptions superimposed upon phenomena by the mind that seeks to connect them in respect of their quantitative relations. (Ward, 1899, vol. II, lecture XIV, p. 103)

The assumed primacy of the physical as against the psychical is due, first, to the fact that in his absorption and interest in the objective attitude the naturalist has forgotten *himself*; and next, to the fact that he has mistaken his abstract conceptions for presented realities. The notion of an epiphenomenon supervening on physical phenomena is in flagrant contradiction with the mechanical conception of connected masses. From the standpoint of physics itself such a notion could never arise: while from the wider standpoint of psychology, to regard mind as the collateral product of its own external perceptions is simply to invert the facts. (*ibid.*, p. 106)

The Gifford Lectures on which Ward's book is based were delivered from 1896 to 1898, so Ward's uses are a little over one century old as I write. This appears to counter Kim's suggestion that the modern term, as opposed to the *notion* of supervenience, was used first in emergentist evolutionary theory and metaethics. Ward's uses are squarely located in philosophy of mind, and somewhat emancipated from evolutionary theory. We should note also that the term has been used earlier than Kim suggests, in a sense likely to be acceptable as 'modern', in translation from Aristotle and Plato. At least, I have found two such uses. The Plato is in Jowett's (nineteenth-century) translation of *Timaeus* (at 50e, p. 1177). Here is some of the Aristotle:

And so the formula of the circle does not include that of the segments, but the formula of the syllable includes that of the letters; for the letters are parts of the formula of the form, and not matter, but the segments are parts in the sense of matter on which the form supervenes [ἐπιγίγνεται: form of ἐπιγίγνομαι, equivalent to Latin *supervenire*]; yet they are nearer the form than the bronze is when roundness is produced in bronze. But in a sense not even every kind of letter will be present in the formula of the syllable, e.g. particular waxen letters or the letters as movements in the air; for in these also we have already something that is part of the syllable only in the sense that it is its perceptible matter. (Aristotle, *Metaphysics*, Book 7, Ch 10, 1035^a 13)

And finally here is a very natural-seeming occurrence in a translation of a paraphrase of Aristotle's *De Anima* by the somewhat Platonistic Themistius (c. 320 – c. 390 AD), perhaps adding to the impression that the idea of supervenience is of respectable age:

24,32 (407^b35–408^a3) The soul causes the body to move, whereas the attunement does not cause movement in what has been attuned, but supervenes on what has been attuned, while someone else prepares the attunement, as where the musician tunes the strings. This means that the soul will need another soul to provide its attunement. (*On Aristotle on the Soul*, p. 41; I am grateful to Dirk Baltzly for this example.)

Explicit and sustained supervenience-talk is, however, an expedient most characteristic of the philosophy of our own times. Such writers as Thomas Nagel, Jackson (pre-1995),

and Chalmers have attacked materialism by attempting to show its failure to prove the supervenience of mind on a material base – or on *any* base, for that matter. Not all have used the term itself, but this is the neatest way to capture their intentions.

3.2.2 *The varieties and uses of supervenience theses*

One of the foremost theorists in the area of supervenience has been Kim, whose articles dissecting all of the notion's types and turns are too many to enumerate here, let alone analyse. We shall mostly find it sufficient to look to his useful summary (Kim, 1994), and from that to select only what is central to our purpose.

Kim and those following his lead make a fundamental distinction between *weak*, *strong*, and *global* supervenience theses. Here are Kim's formulations modified for generality and to fit with Chalmers' usage (since we examine his variants next), each showing what it is for coinstantiations of a set of properties *B* to supervene on coinstantiations of a set of properties *A*, within some chosen domain of objects:

Kinds of supervenience

- Weak: Necessarily (that is, in every possible world), if any *x* and *y* (in the domain) are indiscernible in *A* they are *B*-indiscernible.
- Strong: For any individuals *x* and *y*, and any worlds *w_j* and *w_k*, if *x* in *w_j* is *A*-indiscernible from *y* in *w_k* (that is, *x* has in *w_j* exactly the same properties in *A* that *y* has in *w_k*), then *x* in *w_j* is *B*-indiscernible from *y* in *w_k*.
- Global: Any two worlds that are indiscernible with respect to *A* are indiscernible with respect to *B* (that is, they cannot differ in how *B* properties are distributed). (Adapted from Kim, 1994, pp. 577-8)

Strong supervenience is a kind of extension of weak supervenience to transworld cases; so strong supervenience implies weak supervenience. It also implies global supervenience, which itself, says Kim, *formally* implies neither of the other kinds, though it implies both when conjoined with 'plausible metaphysical premises' or with a restriction to intrinsic properties. The details are not important to us, since it is mainly global supervenience, as defined by Kim and assumed qualified in the manner suggested, that will be used in the discussion here. Accordingly, when I use the term I can generally be taken to intend global supervenience.

The links between supervenience and the modal notions of possibility and necessity are explicit and clear, and much that has been said earlier in terms of possible worlds can be applied (with minor shifts in terminology) to supervenience as readily as to those notions. It already has been so applied, informally; for example, in Part 1 we looked at Jackson's three candidate characterisations of materialism, which were given as supervenience claims, though we ignored any ramifications of this fact at that stage. Just as there are many uses to which standard modal distinctions and discourse can be put, there are many to which supervenience-talk can be put, including the investigation of ethical or aesthetic properties and their relations to underlying properties (*natural* or *physical* properties, for those recalcitrantly capable of making any distinctive sense of these!). But we of course are concerned to apply such talk to the relations between the physical and the phenomenal. Chalmers discusses this interestingly, suggesting significant ways in which these various employments of supervenience-talk differ. For example, he suggests that 'moral facts are not phenomena that force themselves on us' (p. 83), as he alleges that facts of phenomenal experience, being more or less Moorean in their non-negotiability, are.

3.2.3 *Natural supervenience: a restricted supervenience?*

Chalmers makes a crucial distinction of his own between *logical* and *natural* kinds of supervenience. This corresponds closely with his common distinction between logical (or conceptual) and natural (or, roughly equivalently, empirical, nomic or nomological) kinds of possibility. His logical supervenience corresponds to Kim's standard supervenience, and he normally takes it as global (as I have said we shall do, also). The logical-natural distinction is orthogonal to the global-local distinction discussed earlier. Natural supervenience is emphatically not the same, for Chalmers, as *physical* supervenience; though we can imagine use being made of such a variant, Chalmers seems not to use that notion explicitly. In alternative formulations of supervenience theses Kim has used the word *necessary*, and suggested that it can be left open whether the necessity in question is to be logical, metaphysical, analytic, nomological, physical, or what you will, to generate variant supervenience theses as they are. It will be useful if we can achieve a clear general notion of this matter of restricting (global and logical) supervenience of *B* on *A* relative to some restricted domain, following Kim's basic

pattern and his further indications regarding the interpretation of necessity. This way we can work methodically towards an appreciation of Chalmers' natural supervenience. Calling our restricted domain ' Δ ', and calling Δ -laws (along with Δ -constants, Δ -boundary conditions and any other relevant Δ -factors) Δ -constraints, let us define Δ -supervenience like this:

[Δ :] Any two worlds the same with respect to Δ -constraints that are indiscernible with respect to *A* are indiscernible with respect to *B* (that is, they cannot differ in how *B* properties are distributed).

An example of such a restricted type might be *kinship* supervenience, and a specific claim might be made concerning the *kinship supervenience of sisterhood properties on family-tree properties*, with the details as follows:

[Kinship:] Any two worlds the same with respect to *kinship* constraints that are indiscernible with respect to *family-tree* properties are indiscernible with respect to *sisterhood* properties (that is, they cannot differ in how *sisterhood* properties are distributed).

Kinship constraints might include who get counted as cousins, how many 'mothers' one can have (among the Mardudjara Aborigines one's biological mother is a scarcely distinguished member of a whole class of 'mothers' to whom one must pay equal respect, etc.; see Tonkinson, 1978). The kinship restriction on supervenience might usefully be applied because family-tree properties might be indeterminate without it. What might be useful about our example is that, supposing immaterial realisations of family-tree properties to be at all possible, we can engage in discourse about them without any serious metaphysical commitments getting in the way. Applying our ' Δ -template' to physical supervenience, we can make interesting specific claims, such as the *physical supervenience of computer-chip properties on electronic-circuit properties*, whose details go like this:

[Physical:] Any two worlds the same with respect to physical constraints that are indiscernible with respect to *electronic-circuit* properties are indiscernible with respect to *computer-chip* properties (that is, they cannot differ in how *computer-chip* properties are distributed).

This specific claim might be a controversial one, depending on precisely how one is to define electronic-circuit properties and computer-chip properties (and for me, of course, *physical* is not a *bona fide* restriction). But, again, at least one has set aside irrelevant

metaphysical concerns, and is in a position to focus on the controversy of interest.

The precise content of the *physical* constraint is something one would have to settle, even beyond my own reservations. Are broad, over-arching facts concerning the dimensionality or the particular shape of space-time included, or are these not matters to do with the domain but rather more specific matters, determined contingently but in accordance with the constraints imposed by the laws, constants, and whatever else *do* belong in the domain? There is nothing in this problem that is specific to the business of restricting supervenience theses, however; determining the general scope of physical law, for example, is already beset with a similar difficulty. (Note that only lesser difficulties arise for the advocate of necessary materialism, for whom at least all the coinstantiations of properties at a world are supervenient *simpliciter* on the coinstantiations of PNPs at a world, and these PNPs will all be physical by definition; see Part 1.)

The Δ -template seems to provide a reasonable way of construing restrictions in supervenience theses, but I am not committed to their propriety or usefulness. It was desirable to come up with *some* framework, and this is the best readily available sort.

We do encounter formidable problems when attempting to fit Chalmers' natural supervenience to our Δ -template, say with talk of the *natural supervenience of the experiential on the physical* (which is the main use Chalmers puts his natural supervenience to):

[Natural:] Any two worlds the same with respect to *natural* constraints that are indiscernible with respect to *physical* properties are indiscernible with respect to *experiential* properties (that is, they cannot differ in how *experiential* properties are distributed).

Chalmers writes: '[It] is hard to find cases of natural supervenience on the set of *physical* properties without logical supervenience, but consciousness itself can provide a useful illustration' (1996, p. 37). It would be valuable to have a non-prejudicial example to start with, rather than the controversial one Chalmers, and I following him, perforce, have used. But it is indeed exceedingly difficult to come up with anything half-way plausible that is independent of the contentious claim that the experiential supervenes only naturally, not logically, on the physical. Chalmers cannot do so, but seems

unperturbed by this. I would be perturbed! But I am only doing what I can to set up natural supervenience à la Chalmers, and am in no way committed to this particular kind of restricted supervenience myself.

Just what *are* the contents associated with this *natural* domain, standardly or for Chalmers? If it is just an ensemble of 'logically' necessary regularities and conditions determining the limits on all facts, then the central example just given is vacuous indeed: natural supervenience collapses into logical supervenience, which is not what Chalmers intends at all. The natural domain is by that account no restricted domain at all. If on the other hand the natural is to be taken as the physical plus some laws fixing the ways in which the physical is to determine the phenomenal, then again talk of the natural appears afflicted by vacuity: the specific claim of interest amounts to the assertion that the physical determines the experiential in the way in which it does so! Why not just say that the physical is *logically* (and so, equivalently remember, *metaphysically*) constrained to give rise to the experiential in the way it does? As we have seen, Chalmers believes that we should not do that for two reasons. First, we can (allegedly) *conceive* of ourselves, for example, being physically the same but lacking experience altogether. We could, conceivably, have been zombies, while remaining the same physically. Such shifts have already come under attack, but they needed to be mentioned here in respect of the concept of natural supervenience; they will more conveniently be examined, at some length, in our next section, which is dedicated to the 'classical' problems arising from the notion of qualia, some of Chalmers' variants, and especially to zombies. Second, Chalmers believes that:

nobody has an idea of how *any* physics could [entail and explain consciousness]. Indeed, given that physics ultimately deals in structural and dynamical properties, it seems that all physics will ever entail is more structure and dynamics, which [...] will never entail the existence of experience. (p. 163)

This opinion has been criticised earlier also, as prejudging the outcome of the inquiry into relations between the physical and the experiential at the very outset.

I hope that I have at least established that there is room for serious doubt concerning the propriety of *natural* supervenience as a variety distinct from both logical and physical supervenience. But let us explore more what Chalmers himself has to say:

The position we are left with is that almost all facts supervene logically on the physical facts (including [among such facts] physical laws), with possible exceptions for conscious experience, indexicality, and negative existential facts. To put the matter differently, we can say that the facts about the world are exhausted by (1) particular physical facts, (2) facts about conscious experience, (3) laws of nature, (4) a second-order "That's all" fact, and perhaps (5) an indexical fact about my location. (1996, p. 87)

There are some problems with this explanation (even apart from the very odd *indexical fact*, about which Chalmers himself expresses serious doubt). He sometimes appears to use the phrases *natural law* and *physical law* interchangeably, as suggested for example by the index to his book, in which discussions of physical and natural laws are all assembled under the head 'laws of nature'. But it emerges that the class of natural laws is formed by the union of the class of physical laws and the class of psychophysical laws (which classes are themselves disjoint), as we have seen:

Creation myth: Creating the world, all God had to do was fix the facts just mentioned [see last excerpt]. For maximum economy of effort, he first fixed the laws of nature – the laws of physics, and any laws relating physics to conscious experience. Next, he fixed the boundary conditions: perhaps a time-slice of physical facts, and maybe the values in a random-number generator. These combined with the laws to fix the remaining physical and phenomenal facts. Last, he decreed, "That's all". (*loc. cit.*)

(Note that according to this creation myth, the allegedly non-natural and totally unruly merely 'logically' possible worlds could, arguably, not be created – unless one thinks that all worlds have some regularities that will count as their laws of nature. But not too much should be made of this: we noted earlier that Chalmers characterises the logically possible worlds metaphorically as those that God could have created. But perhaps the formula rules out *Hume worlds* – worlds like the natural-law-regulated ones in every way except that there are no natural laws at them. Their alleged possibility, taken ontologically seriously, gives us good grounds for believing that there are no natural laws at all, but I do not present the paranthropic argument for this. It will be paralleled in my main argument against Chalmers' zombies.)

The quote shows an inconsistency. Even with this further clarification, there are two serious uncertainties. First, a fair thing to wonder is how conscious experience *itself* is supposed to have come into being. We can see that it is to be related to physical facts, but it is not settled for us whether conscious experience itself *could* have arisen entirely

unsupported by anything physical. If that is indeed to be considered one of God's options, then we might further wonder whether the phenomenal facts could have been created unsupported, and if not, why not. Generally, why should the *only* non-physical natural laws allowed in the logical space contemplated by God be laws relating some non-physical instantiations to some presupposed physical instantiations? Specifically, why was a Berkeleian world not an option? Or a Cartesian world? If these were options – and for Chalmers, given his Lewisian lavishness when it comes to logical (and so metaphysically) possible worlds, such worlds and other more complex and hybrid cases surely *were* options for God – there would have to be the prospect of instituting some independent nature for what we might call *pure* phenomenality. The restriction to merely fixing relations with the physical fits ill with the broad, free-ranging, modal speculation Chalmers needs to engage in to set up his notion of natural supervenience securely and comprehensibly. But then, if he were to modify the Myth to allow for these presumed possibilities, the property dualism he is at pains to develop looks very arbitrary and much more open to modal attacks.

The second uncertainty about the Creation Myth arises from the fact that the status of constants (like the gravitational constant *G* and the other parameters discussed thoroughly in Part 2; not, *pace* Chalmers, like his gas-law constant *K*, though we allow it for illustrative purposes below) is not made clear. We might have thought that the determination of their values was achieved by fixing the physical laws, or alternatively by fixing the remaining (non-physical) natural laws. Remember how Chalmers had earlier attempted to explain his position regarding nature and natural supervenience and necessity, using *K*:

It is *logically* possible that a mole of gas with a given temperature and volume might have a different pressure; imagine a world in which the gas constant *K* is larger or smaller, for example. Rather, it is just a fact about nature that there is this correlation. (1996, p. 36)

Setting aside my objection specifically in respect of *K*'s alleged contingency (we have to work here with Chalmers' own example), *K*'s determination cannot be a matter of *physical law*, otherwise it would not have been necessary for Chalmers to specify boundary conditions separately (see the Creation Myth excerpt, above), since these are, according to standard cosmological theory, what determine the constants (see for

example Barrow and Tipler, 1988, pp. 444-9). But then, they cannot be considered as determined by the *natural laws* either, since the scope of the natural laws, according to Chalmers' Creation Myth, is very restricted – the only natural laws are those relating consciousness to physics, so they must be supposed to have little to do with physical constants! The inclusion of God's decree *That's all* doesn't help: we want to know what else she *could* have added, if she had chosen to make the world differently; and the *separate* fixing of the boundary conditions in the Myth leaves their place in the scheme of things uncertain, and so the exact import of various supervenience theses remains uncertain, too.

As discussed, Chalmers appears to make assumptions about the nature of the 'nature' of the world in the absence of clear specifications which, given the very broad context in which his inquiry is conducted, we really do need. We have further evidence now of this. It seems now that we need to construe his notion of natural supervenience otherwise than the Δ -template way, though that had seemed to be the rational way.

We earlier noted conflation and uncertainties regarding logical, epistemic, and natural possibilities in Chalmers. Rather than rehearse these again here, I present below a striking passage from late in Chalmers' book which I think is quite revealing, and which we looked at for another purpose earlier (see the excerpt labelled 'NP5', in 3.1.1). He has been wrestling with the fact that his logically possible zombie twin, having precisely the same causal history in his world as Chalmers has in the actual world, his duplicate in every way (certainly all physical ways) except in lacking consciousness, will have exactly the same beliefs as he has (on a reasonable understanding of the notion of belief), including the unshakeable conviction that he is conscious, and that there is a 'hard problem' associated with that consciousness. Along with this, of course, we are to remember that Chalmers is explicitly committed to the causal closure of the physical; otherwise he could not so comfortably believe that he conceives coherently of his zombie twin. So, Chalmers comments:

Now my zombie twin is only a logical possibility, not an empirical one, and we should not get too worried about odd things that happen in logically possible worlds. Still, there is room to be perturbed by what is going on. After all, any explanation of my twin's behavior will equally count as an explanation of my behavior, as the processes inside his body are precisely mirrored by those inside mine. (p. 180)

The details in the last two sentences need not concern us now, though the obvious and much-worked zombie theme will be played again (I hope with some interesting variations) in the next section. To Chalmers' credit he gives the problem a sustained, chapter-length treatment, and more. What is fascinating is the first sentence in the excerpt. Why does Chalmers think that 'we should not get *too* worried' about what happens in [merely] logically possible worlds? These are, remember, to be accepted as *metaphysically* possible worlds. That is to say (or *should* be to say!), they still represent ways the world could be, even though these ways are more outlandish than the more obvious and 'normal' ways. These worlds are not 'close' to the actual world in logical space, in that whatever criteria are to be taken as specifying the nature of our world, they lack more of them than a great number of other worlds do. So a sensible and sustainable account seems to run, anyway. *If* it is a fact that the actual world is endowed with phenomenal properties somehow beyond its physical properties (however construed), and not logically necessitated by them, then the worlds entertained in the zombie-twin scenario share little of the nature of the actual world, by certain reasonable criteria. But of course, by other equally reasonable criteria, they share a great deal of that nature – they are *physically indistinguishable* from the actual world. If 'natural' were just another qualification, useful for restricting possibility and supervenience theses, we have strayed far from any proper or canonic use of the term if complete physical indistinguishability from actuality counts for so little in determining the 'naturalness' of worlds!

My diagnosis of the problem is this. Once again, as we saw in Part 2, it is dangerous not to adopt some framework like modal realism, or a decent simulation thereof, when contemplating how the world might have been different. Chalmers has not set himself up with any such framework, and things have indeed come unstuck. The whole point about the equivalence of logical, conceptual, metaphysical, *broad* possibility is that logical considerations should be a guide to *real* possibility. At least this is a reasoning standard understanding of these terms, and not mine alone; it seems to be Chalmers' understanding when it suits him (that is, when he invites us to accept that the facts of phenomenal consciousness genuinely *could* differ, even though all physical facts stayed the same). We need an articulated alternative if things are to be understood otherwise. For Lewis, all possible worlds are equally real; and for robust Ersatzers all merely

possible worlds enjoy the same ontological status (the same *measure of reality*, as it were) because they all could equally have been actual instead of the world that *is* actual. As we have seen Part 2, Forrest's epiphenomenon-ridden worlds, to be even *merely* possible, must be *a priori* as 'likely' to have been actual as each other (and as any other merely possible worlds). If there were differences in likelihood, so that we could judge *a priori* that some worlds are 'ruled out' as unlikely, excessively fantastic, outlandish, or an affront to norms of one sort or another, that ruling out could only serve to exclude those worlds from the domain of the possible altogether. At least, we are entitled to assume that this is so unless presented with an articulated account of how *a priori* differences in the 'probability' of worlds being actualised can be at all proper.

Also in Part 2 we saw that, even with the assistance of a belief in modal realism, Lewis himself sometimes fails, in his own way, to take the 'equiprobability' of worlds seriously. The actual world (indexically, *this* world) could equally turn out to be any particular one of the infinitely many epiphenomenon-infested worlds of a certain class of physically identical worlds, or the single world in that class that lacks epiphenomenons: and we could have no way of selecting among this feast of options, though Lewis is a firm believer in an epiphenomenon-free actual world.

I conclude that when Chalmers speaks of natural possibility, the only way we can begin to make sense of what he says that has any chance of making it plausible is to understand him as talking about real possibility – possibility *tout court*, which in the end must be equivalent to logical (and so metaphysical) possibility, since *they* are elsewhere taken as real. This is so despite his protests that logical-metaphysical possibility differs from natural possibility, and despite his heavy dependence on these differences to set up the proposal that phenomenal consciousness is both real and only contingently annexed to the physical. If the supposed 'merely' logically possible worlds do not represent full-blooded ways the world could be, they are not genuine possible worlds at all. They are ruled out by that very failure to represent, for all that we might have fleetingly entertained them as *candidate* possible worlds.

The same difficulty attends his customised natural version of *supervenience* – which can be taken as the linchpin of Chalmers' whole program. Natural supervenience can only coherently be understood as supervenience *simpliciter*, despite all the appeals that

Chalmers makes to the contrary. And understood that way, any claims of his that depend on some distinct *logical* supervenience being supervenience *simpliciter* fail, as surely as his views based on equivocations regarding *logical possibility* do.

Before turning to examine one of his key applications in the light of this assessment, I offer two more instances, from the many that are available, to confirm my interpretation. First, concerning a law that God might establish to ensure that certain B-facts get determined by certain A-facts, Chalmers writes: 'Once the law is in place, the relevant A-facts will automatically bring along the B-facts; but one could, in principle, have had a situation where they did not' (p. 38). The qualification 'in principle' provides further circumstantial evidence that we are here talking not about *real* possibilities but something merely and irremediably hypothetical. If this interpretation of the merely logical possible as the merely hypothetical still seems to be in order, and in happy accord with Chalmers' myth-making about the 'logical' possibility of things being otherwise, let us remember the crucial work that Chalmers allocates to this logical possibility, and how it sometimes cannot be taken as other than the most general sort of real possibility – when setting up a myth as a model of how things came to be as they are, for example.

The second instance arises in this observation (that we have also seen before, as NP3 in 3.1.1) concerning the natural but not logical supervenience of qualia on a physical base, and the consequent *natural* impossibility of fading qualia (for which see next section):

To be sure, fading qualia are *logically* possible. There is no contradiction in the description of a system that is so wrong about its experiences. But logical possibility and natural possibility are different things. We have no reason to believe that this sort of case could happen in practice, and every reason to believe otherwise. (p. 257)

This excerpt is strongly reminiscent of the passage seen earlier in which dancing qualia were said to be '*only just* logically possible' (NP4, examined at length in 3.1.1). We are asked to accept, of some scenario alleged to be logically (and so metaphysically) possible, that we have 'every reason' to believe ~~that~~ it could not happen 'in practice'. If that is not to say that the scenario falls short of *real, genuine* possibility, what *is* it to say? Remember that we are also to suppose that if the allegedly logically possible scenario were to be actualised, our evidence, including all that we could report or

believe concerning our experience, would be *precisely* as it is now!

In sum, appeals to supervenience theses do not seem to have rescued Chalmers from the collapse of some spuriously distinguished 'naturally' possible domain into the domain of the possible *simpliciter*. It may be that supervenience comes apart insidiously easily from world-based analysis – which analysis, if we are ontologically serious about worlds, will keep us honest enough. But all too easily does supervenience come to drift alongside the fleet of other linguistic, logistic devices, at the hands of a linguistically inclined philosopher. It becomes, by allowing a more distant foray from the ontological *terra firma* of modal discourse, mere word-talk, not world-talk. And we find ourselves at sea again.

But now it is time to go in search of zombies.

3.3 QUALIA AND THE ZOMBIES

3.3.1 *The secret life of zombies*

Daniel Dennett says of the zombie argument against the supervenience of consciousness on the physical (or on a functional system, however realised) that Bieri has nicknamed it *The Tibetan Prayer Wheel* (Dennett, 1991, p. 281; no, Bieri and Dennett are not guilty of an elementary misreading of Julien Offray de la Mettrie's *Homme Machine* as *Om Machine*). The gist of the argument is that, no matter what physical or functional description is given as a candidate complete specification for a fully sentient human person like one of us, it must fail, because, it is alleged, one can imagine or conceive of the thing specified lacking consciousness entirely – and *surely* we have consciousness. Ergo, consciousness is neither physical nor functional, and some sort of dualism must be true. The argument is one implementation of the basic vaguely Cartesian schema *I can imagine materialism (along with functionalism) not being true, so it isn't*.

Chalmers suggests an entire *zombie world* (1996, p. 94): a perfect physical duplicate of our world that is entirely devoid of conscious experience (that is, all phenomenal consciousness, as distinct from anything that could be characterised cognitively, or *psychologically*, as Chalmers would have it). As we have come to expect, this is supposed to be a *logically* possible world, in a usual sense; that is, it is a world that one could conceive of – all its detail, given sufficient processing capacity for such an operation, without any sort of incoherence or contradiction, but it is not to be taken as a practical, empirical, or 'natural' possibility. The *zombie world* represents a sensible move on his part, since it is, at least at first, most convenient to think about these issues in terms of global supervenience, but Chalmers does not always specify his zombie world quite as explicitly in such global terms, as seen earlier (3.1.1).

What is the causal story supposed to be, at the zombie-world equivalent of the actual world? What 'causal laws' should we presume to be in place at it? Exactly those that we should think operate at the actual world – if there are any.

Here, on the other hand, is what Chalmers says regarding how pervasive must be the operation of whatever principles will turn out to underwrite consciousness, in the context of showing how consciousness could not be subject to evolutionary pressures:

Presumably these principles – whether they are conceptual truths, metaphysical truths, or natural laws – are constant over space-time: if a physical replica of me had popped into existence a million years ago, it would have been just as conscious as I am. (1996, p. 121)

It is hard to see, though, how Chalmers is not compelled to allow a certain 'logical' possibility (and so, I would insist, a *real* possibility) for worlds. At these worlds, only *some* inhabitants at only some places and times (physically identical to their fully conscious counterparts in the actual world) lack consciousness entirely and permanently. And some other inhabitants lack consciousness of certain kinds only, or only some of the time, or a combination of these, and still other inhabitants have consciousness continuously – or at least, as continuously as we are presumed actually to have it. Now, at this stage it is just as well for Chalmers that he is *not* equipped with a robust framework such as modal realism, because if he did have one (and was free of Lewis's prejudices that enable him to believe the actual world to be the single epiphenomenon-free world in its physical equivalence class, for example), he would have to be entirely agnostic about how zombie-inhabited or otherwise *this* world is. My argument for this is essentially a version of the paranthropic argument I gave in Part 2.

One who has made a similar point is Joseph Levine, reviewing Chalmers' book. He suggests that progress might be made:

[...] if one could argue that zombie worlds are metaphysically possible, but not mixed worlds, where there are pairs of functional duplicates with one conscious member and one zombie member. However, given that the functional-consciousness connection is only nomological, there is no way to rule out the metaphysical possibility of mixed worlds. The question then becomes, how do we know our world isn't one? [...] The bizarreness of fading and dancing qualia according to the reductive functionalist is tantamount to conceptual incoherence. If she's right, then clearly they can't be nomically possible, since they're not even metaphysically or even conceptually possible. But Chalmers has to say – what he does say forthrightly – that he allows the conceptual coherence, and therefore the metaphysical possibility of fading and dancing qualia. [...] But what evidence do we have that they could really happen? Well, what evidence do we have that they couldn't? (Levine, 1998, p. 880)

Chalmers would respond by insisting that no matter what is metaphysically possible, it is a separate question which metaphysical possibilities are also *natural* possibilities – which of them could occur at *our* world. We just *know* that we have qualia, and that they do not invert, fade, or dance on and off; so we just know that we are not zombies, any of

the time. And, given the observed uniformity of the world in 'physical' ways, we should positively expect that this uniform, unbroken distribution of qualia holds throughout the world, also. What are we to say about this? It is time to make use of our results from Part 2, and of some of the work done earlier in Part 3.

3.3.2 *Fermi and the dancing zombies*

Chalmers is not strictly entitled to answer Levine's objection as outlined just now, perhaps; but Levine has not got available to him the *homogeneity of possibility* (see Figure 2, on p. 208), as I have. He shares too many of Chalmers' base assumptions for that. I think Levine is right, but that his point is impossible to drive home in a way that would convince so resourceful and committed an advocate of dualism as Chalmers, who has powerful 'logistic' arguments to fall back on.

I have urged that there is only one sort of possibility that is metaphysically interesting at the level of worlds. All other sorts are pretty well arbitrary restrictions on this broad possibility; when it comes to ontology, epistemology, and especially the formation of expectations concerning the world we are confined to, we must survey the evidence available to us at that single world, draw from this evidence what conclusions we can about the whole range of worlds, and then eventually use these conclusions to arrive at further conclusions concerning the world we started with in the first place. This was the naturalistic, induction-respecting, and science-respecting message of Part 2.

Enrico Fermi famously asked, concerning supposed extraterrestrial beings, *Where are they?* If they are out here, why should we not ever see evidence of them? (Leslie, 1996, pp. 137 ff.) If they *were* there, surely we *would* have evidence. The question may be taken to assume a Copernican attitude to our place in the cosmos (as discussed in Part 2). Our place in spacetime is not an especially privileged place – with respect to absence of life forms, in this case. (I do not follow any of the extensive literature on the so-called Fermi paradox here: I am hijacking it as a raw question, for my own purposes!) The question can equally be posed for zombies: if there are any zombies in the *worlds*, why do we not observe them?

(The short answer is: Because they are by their very nature not observable! Their zombiehood is not even detectable to themselves. Myself, I think that makes the whole

zombie program fall on its face. But Chalmers has arguments of a non-world-based sort to bolster the coherence of zombie-talk, even confronted with what I take to be the very sound suggestion that *we* are the zombies, if Chalmers is right that zombies are at all possible. And in that case, the very existence and so the possibility of phenomenal consciousness is directly questioned. But I am not interested in dealing with those short answers, and those very long responses to them: we are working with ontological seriousness about worlds.)

Chalmers will answer that they are not 'naturally possible', as all the most immediate evidence shows us that we – and inductively, all things physically or even functionally like us -- are constrained by natural necessity to have qualia. They are naturally supervenient on our physical and functional structure. But this can mean nothing at all in the context of my moderate necessitarianism! Any putative 'natural' species of modality is at best arbitrary; so it can carry no epistemic or ontological freight for us.

But, the objector might retort, zombies may be *sparsely distributed* in the worlds: just as extraterrestrials may be few and far between at *our* world.

To answer this we need a sort of paranthropic argument (along with our multiplicative strategy) applied in the context of accepting Lewisian recombination (LR; which Chalmers must accept, if he is to derive such notions as metaphysically possible zombiehood) and the notion of zombiehood itself. Several arguments of this sort are presented in Part 2, so I shall not go through all the nuanced details here. Here is one version among many for mixed zombie worlds, somewhat informally presented:

- P1: If zombiehood is possible but not necessary, and reasonably assuming that its presence is not made less likely by the presence of reflective observers (like us), then zombiehood should be roughly equally distributed throughout total reality. [a 'Copernican', paranthropic assumption of a sort discussed at length in Part 2, and supported by the homogeneity of possibility for worlds]
- P2: Zombiehood is possible but not necessary. [Chalmers' assumption]
- P3: LR is true. [Chalmers' assumption]
- P4: Some worlds are mixed zombie worlds, with zombiehood at some parts of them but not at others. [from P2, P3]

- P5: We can have no evidence either way concerning there being zombies at our world. [from the nature of zombiehood]
- P6: The mixed zombie worlds vastly numerically dominate both the pure zombie worlds and the pure non-zombie worlds. [from P3, P4, and multiplicative strategies to avoid problems with the principle of indifference]
- C: We should believe that our world is a mixed zombie world. [from P1, P5, and P6]

There are several moves we could make with this argument. Here are some: Chalmers would want to deny the conclusion, so he would probably deny P1 – which relies on the very homogeneity of possibility that he does not accept. Someone else not accepting the conclusion might reject P3 – Lewisian recombination. That is certainly a premise I reject. Or one could reject P2, and say that zombiehood is not possible; alternatively, one could say that zombiehood is necessary, and so also deny P2. In either of these cases one denies P4 (that there are mixed worlds), and so also the conclusion (that ours is a mixed world).

So we would have to go back and inspect what we already believe, and how strongly, before finally judging *how*, if at all, this argument should shift *which* of those beliefs. The conclusion is essentially that ours is a world of *dancing zombies*. If it is true, then these dancers are in the same world as Fermi, who should therefore accept that his 'paradox' is dissolved in respect of zombies, at least. The only glitch is that Fermi and we cannot detect them. The world is replete with them – and we are almost certainly among them, samba-ing our way in and out of zombiehood.

Zombie talk is really just another way of doing the arguing that can be done with old-fashioned qualia, but made chic and up-to-date. And now we turn instead to a brief review of some consequences for classic qualia puzzles.

A buzz of qualia

By my paranthropic style of reasoning, combined with Lewisian recombination, some surprising results are obtainable with qualia thought experiments. We are able to apply paranthropic arguments again and again in Part 2, to get more and more likelihood of rapid variability in parameters.

If it is granted that my arguments show we should expect changes in the zombie-status

of complex psychological beings like ourselves at the actual world, given Chalmers' LR-style worldmaking principles conjoined to my paranthropic, Copernican assumptions, and my *homogeneity of possibility* (there being 'objectively' only one sort of possibility), there is an interesting consequence. Assume a period of any arbitrarily short duration in a person's life, and consider the probabilities for that period having a change to or from zombiehood within it. By a conjunction of my assumptions and Chalmers' assumptions, we should expect there to be indefinitely many! For any such period, and for any number of such changes, there is a class of worlds at which that number of changes does occur in that period; the very rarest worlds would be those at which there occur none or *nameably few*. (For this recall discussion of π and γ in Part 2; and I am assuming here that cardinality problems and difficulties with the principle of indifference can be sorted out much as they were in Part 2. The technique used was quite robust.) So rationally, we should expect that at every moment of our supposedly conscious lives we are oscillating in and out of zombiehood, at an infinite rate. A 'white noise', if you like, of zombie-changes.

I should make it clear that I find all of this scarcely intelligible myself; I simply draw conclusions for Chalmers.

Now, we might think that this means little, in the end. We would have to say that Chalmers is mistaken: given only the additional assumptions that have been argued for at length in this work, there can be zombies in worlds like ours if there are zombies at *any* worlds. If zombies are possible there certainly *are* zombies at this world: at least, our credence for this should approach arbitrarily close to 1. We are all part-time zombies! Chalmers, while not being able easily to solve this problem, might answer that this flickering, buzzing, zombiehood would not be detectable to us, especially since he believes that in any 'logically possible' world with zombies, those zombies in principle cannot know that they *are* zombies. And after all, even if we were to grant 'introspectable' zombiehood, the case might be likened to what happens when we watch the flickering image on the screen on a television. We detect only the light, not the intervening darknesses between the rapidly pulsating illuminations.

But as I said, we should look at some classic qualia problems in the same way. If we consider the absent qualia scenario, which Chalmers turns into a dancing-qualia

scenario (essentially the same thing as flickering zombiehood), things are as just described. But what about the inverted spectrum? In such cases we should expect alternation between the standard orientations of our qualia and their opposites! Which, then, would *be* the standard orientation, anyway? Our qualia would be a uniform... what?... Grey? White? Brown? Who can say! It seems to me that for qualia-believers like Chalmers there is in each of us what Daniel Dennett (1991) likes to call a *Cartesian theatre*, at which the qualia (straight or inverted: who can say?) are appreciated and interpreted. And of course this issues in a vicious infinite regress.

Many have pointed out the incoherence of this sort of talk; Dennett is prominent among them (Dennett, *ibid.*, and 1988), and I have no change to suggest in his excellent analyses of the non-problem of qualia. I only urge the addition of the foregoing style of objection, which I claim as unusual in its ontologically informed and serious application of modal considerations. The route to this sort of application is long and arduous, but we have taken the journey and are now entitled to the grail.

Where to next?

We have touched on the problem of phenomenal judgement for believers in qualia, like Chalmers. He is an ingenious apologist for his views, and has a rich stock of speculative ways of dealing with seeming impasses. Many are linguistic and 'logistic', in my pejorative sense – disengaged from ontology. Mere *words* which cannot deliver results about *worlds*, if they are not annexed to experience, and inductive deliberation *founded* in experience. But he does speculatively address radical alternatives to standard ways of doing empirical investigation – at least in the abstract. We consider one now in some detail – or rather, we consider a certain appeal to authority, on Chalmers' part.

3.3.3 *In search of Russell's neutral monism*

One hope that Chalmers expresses is that neutral monism might come to the aid of his program. This is an ill-defined view that Russell dallied with, at least, early in the twentieth century, and concerning which he showed the most extraordinary and protracted ambivalence. Many sources are relevant to untangling how Russell negotiated his way in and out of acceptance of the doctrine, though among Russell's own texts Chalmers refers only to *Analysis of Matter* (1927a; he cites it some of the

time as 'Russell, 1926': it is based on the Tarnier Lectures of 1925).

Early on we find Russell citing Mach and James as independently proposing the view that:

the things commonly regarded as mental and the things commonly regarded as physical do not differ in respect of any intrinsic property possessed by the one set and not the other, but differ only in respect of arrangement and context. [...] Two objects may be connected in the mental world by association of ideas, and in the physical world by the law of gravitation. The whole context of an object is so different in the mental order from what it is in the physical order that the object itself is thought to be duplicated, and in the mental order it is called an 'idea', namely the idea of the same object in the physical order. But this duplication is a mistake: 'ideas' of chairs and tables are identical with chairs and tables, but are considered in their mental context, not in the context of physics. (Russell, in 1956, p. 139: *On the Nature of Acquaintance*, first published in 1914.)

Russell follows this with thirty-five pages of close analysis, in the course of which neutral monism is granted recognition as superior to various earlier theories, but is ultimately rejected. By the time we get to *Analysis of Mind* (1921), neutral monism appears to be accepted, though with apparently serious reservations:

My own belief [...] is that James is right in rejecting consciousness as an entity, and that the American realists are partly right, though not wholly, in considering that both mind and matter are composed of a neutral-stuff which, in isolation, is neither mental nor material. I should admit this view as regards sensations: what is heard or seen belongs equally to psychology and to physics. But I should say that images belong only to the mental world, while those occurrences (if any) which do not form a part of any "experience" belong only to the physical world. [...] entities subject only to physical laws, or only to psychological laws, are not neutral, and may be called respectively purely material and purely mental. Even those, however, which are purely mental will not have that intrinsic reference to objects which Brentano assigns to them and which constitutes the essence of "consciousness" as ordinarily understood. (pp. 25-26)

Precisely what the doctrine amounted seems to have progressively undergone subtle changes, so that with *Analysis of Matter* (1927a) the content is not clear, nor is Russell's precise relation to the doctrine. Lockwood, after citing Russell's 'plaintive passage' (1959, p. 15) in which he claims that he had 'solved' the problem of mind and matter but that no one had understood him, observes that:

the vast majority of contemporary philosophers have probably never read Russell's writings on this issue. [...] But of those who have read the relevant texts, it is perfectly true that most have failed to see what Russell was driving at; I have explored elsewhere (Lockwood, 1981) the extraordinary misapprehensions that have prevailed among Russell's leading philosophical commentators. [...]

Professional philosophers [...] read his exposition of these views, if at all, through the distorting lens of their own philosophical preconceptions and have mostly made nonsense of them [...]. (Lockwood, 1989, p. 157)

One who disagrees with Lockwood's assessment is Kilmister, who asserts that Russell's work in general after 1911 is 'lucid, readable, and easily available' (Kilmister, 1984, p. 227); but even a cursory survey of the evidence shows that it weighs heavily in favour of Lockwood's assessment. Kilmister claims that *Analysis of Matter* 'marked the end to the neutral monism and a movement towards what Ayer calls "physical realism"' (p. 230). This is in accord with Passmore's declaration (1968) concerning *Analysis of Mind*:

This is as far as Russell ever went in the direction of neutral monism: in *Analysis of Matter* – in which he comes to terms with Einstein's 'new physics' – he turns hard a-port to something more like, though very different from, that 'inferential' theory of physical objects which he had maintained in *The Problems of Philosophy*. (p. 236)

Passmore adds, with more than a hint of impatience: 'The twists and turns in Russell's argument after *The Analysis of Matter* we cannot follow in detail' (p. 237). Wedberg, in a work offering much valuable exposition and critique of Russell's views, usefully placing him in a 'Hume–Mach' tradition, at first refers with unruffled equanimity to the neutral monism of both *Analysis of Mind* and *Analysis of Matter* (Wedberg, 1984, p. 15). But there are questions about how idealised and abstracted Wedberg's reading of neutral monism is, bearing in mind his comment that: 'Russell's neutral monism in its different forms is a compromise: he retains the ideal as long as possible without getting obviously stuck in the difficulties' (*ibid.* p. 155; let us note also his observation that: 'Russell's writings are justly praised for their brilliance, but most of them are scarcely systematic, and their renowned "clarity" is often only on the literary surface'; p. 127). Later Wedberg informs us, somewhat ambiguously, that:

Up to *The Analysis of Mind* (1921), Russell believed that there was a fundamental difference between mind and matter. The name "neutral monism", which Russell applied to his theory after 1921, indicates the abandoning of that belief. (p. 159.1)

At the time of the publication of *Analysis of Matter* Russell himself writes in *Outline of Philosophy* (1927b):

The datum is a datum equally for physics and for psychology: it is a meeting point of the two. It is neither mental nor physical [...] This is the theory which is called "neutral monism", and is the one that I believe to be true. (p. 217)

And lest it be too easily supposed that the term is here appropriated to refer to some *new* view, let us note his remark on James's original suggestion:

He thus laid the foundations for what is called "neutral monism", a view advocated by most American realists. This is the view advocated in the present volume. (*ibid.* p. 218)

Russell still cites James's original thesis with obvious approbation in *History of Western Philosophy* (1945, p. 812, and p. 833, where it is clearly James's thesis that is still to be called 'neutral monism'). As Galen Strawson suggests, though, Russell 'adopts ["neutral monism"] as a term for his own view, and seems to mean something very peculiar (and phenomenalist) by it' (1994, p. 97, note 6).

In the light of these exegetical complications, it is odd that a doctrine Chalmers considers so close to his own concerns should attract so little careful citation and analysis from him. Perhaps we could speculate that he came upon Russell relatively late in the formation of his views. The twenty-six articles in response to the keynote article yield just one two-line reference to Russell, briefly mentioning his neutral monism (Hameroff and Penrose, 1997, p. 177), but in Chalmers' piece in reply to his critics there are at least thirteen mentions of what Chalmers calls the 'Russellian view' (Chalmers, 1997; in his 1996 book there were a few, but pretty well *mere* mentions).

Though the relevant (penultimate) chapter of *Analysis of Matter* has a decidedly problematic status in the Russell corpus, it is certainly an extraordinarily fertile tract of philosophy, inspiring not only Chalmers but also, though rarely with any acknowledgement, those who advocate a quantum-theory account of consciousness (e.g. Penrose, 1989, 1994; Lockwood, 1989; Hodgson, 1991). We may even discern a precursor of Jackson's famous knowledge argument (1982) in this passage:

It is obvious that a man who can see knows things which a blind man cannot know; but a blind man can know the whole of physics. Thus the knowledge which other men have and he has not is not part of physics. (p. 389)

But we *cannot* find a well articulated version of neutral monism. Contemporary theorists will have to do without moral support from Russell in that regard.

3.4 CONCLUSION: CONFABULATING CONSCIOUSNESS

My conclusion is easily stated, and will be briefly stated. No matter which turns are taken by its advocates, I find the very notion of consciousness incoherent (*phenomenal* consciousness, of course: that which gives us the *hard problem*, as Chalmers calls it). I know that something is going on at what I am as tempted as the next philosopher to call the phenomenal level – an ineradicable *sheen* on experience. But I cannot feel Chalmers' problem with it, and I am not convinced by his arguments that I ought to have a philosophical problem with experience that should lead me away from my materialist monism. I remain convinced that the sciences are very likely to be on the right track! And I include the neurosciences.

If phenomenal, non-causal consciousness has not *evolved* (given that evolution, however it is theoretically worked out in detail, is clearly a causal affair), it remains true that *belief* in consciousness has evolved, so it is very likely to be adaptive. (This belief is not universal, of course. There does not seem to be a problem of phenomenal consciousness, such as Chalmers would see us vexed with, in Aristotle. Nor in Descartes, nor Hume, nor Locke, for that matter.) I shall not speculate here on the likely nature of that adaptive quality, beyond the suggestion that it is closely related to belief in personhood – another favourite philosophical illusion, for another long thesis (well, perhaps in some possible world...). We have an enormous appetite for, and propensity to believe in, our own *wholeness*, so that whatever calamities befall our nervous systems we seem automatically to pick up the pieces and redefine ourselves as a *person*. This is the extraordinary message from the split-brain studies, in which the brain itself is bisected yet the sense of single personhood is entirely unruffled (despite the fact that information flow between the hemispheres is drastically reduced). To account for this seemingly unlearned virtuosity in split-brain patients, we may be drawn to the conclusion that they have had a great deal of pre-split practice at confabulating wholeness – that we all do that all the time. (For such striking accounts of confabulation of wholeness in these patients see the classic studies by M. Gazzaniga, 1970; and Gazzaniga and LeDoux, 1978.)

Perhaps, to convert a famous Humean phrase, consciousness is the *cement of personhood*. That is to say, *belief* in consciousness is one way we have come – or some

of us, in some cultures have come – to make some sort of sense of what we are. But this is no guarantee that the notion is ultimately philosophically respectable. I am enough of a Humean to be sympathetic to revisionist treatments of causation; and enough of a Dennettian to *quine consciousness*, as Dennett has convincingly quined qualia (which I claim amounts to the same thing, anyway).

I close with a quote from the neurologist Macdonald Critchley, whose *The Divine Banquet of the Brain* (1979; the phrase is from the writings of William Harvey) is an engaging and philosophically rich potpourri of observations on the human condition from a deeply empirical and scientifically astute point of view. (The condition in question is commonly called *Anton's syndrome*, though Critchley does not call it that.)

It is the more severe defect, however, which concerns us today, chiefly by reason of the diversity and unexpectedness of the psychological reaction. The sudden development of bilateral occipital dysfunction is likely to produce transient physical and psychical effects in which mental confusion may be prominent. It may be some days before the relatives, or the nursing staff, tumble to the fact that the patient has actually become sightless. This is not only because the patient ordinarily does not volunteer the information that he has become blind, but he furthermore misleads his entourage by behaving and talking as though he were sighted. Attention is aroused however when the patient is found to collide with pieces of furniture, to fall over objects, and to experience difficulty in finding his way around. He may try to walk through a wall or through a closed door on his way from one room to another. Suspicion is still further alerted when he begins to describe people and objects around him which, as a matter of fact, are not there at all.

Thus we have the twin symptoms of anosognosia (or lack of awareness of defect) and confabulation, the latter affecting both speech and behaviour. (*Modes of reaction to central blindness*, in Critchley, 1979, p. 156)

Perhaps we all have *super Anton's syndrome*, with respect to this phenomenal consciousness that so much is made of. Critchley goes on to observe that his patients often do change their beliefs and achieve insight into their condition, by rational persuasion. It may be that I and other resolute monists have diagnosed an even more insidious and baffling confusion in the human condition – and my hope is that arguments such as those I have brought to bear here will have their own part to play in our epistemic recovery.

BIBLIOGRAPHY

- Aaron, Richard I (1965) *John Locke*, (2nd ed.), Oxford: London
- Alexander, Peter (1985) *Ideas, Qualities and Corpuscles*, Cambridge University Press: London
- Aristotle *Metaphysics*, trans Ross, in McKeon, R. (ed.), 1941. *The Basic Works of Aristotle*, Random House: New York
- Armstrong, DM (1980) 'Naturalism, materialism and first philosophy', in *The Nature of Mind*, University of Queensland Press: Queensland
- Armstrong, DM (1989) *A Combinatorial Theory of Possibility*, Cambridge University Press: Cambridge
- Barrow, John, and Tipler, Frank (1986) *The Anthropic Cosmological Principle*, Oxford University Press: Oxford
- Bigelow, John, and Pargetter, Robert (1987) 'Beyond the blank stare', *Theoria*, Vol. 53, pp. 97-114
- Bigelow, John (1976) 'Possible worlds foundations for probability', *Journal of Philosophical Logic*, vol. 5, pp. 299-320
- Bigelow, John (1990) 'The world essence', *Dialogue*, vol. XXIX, pp. 205-217
- Bigelow, John (1996) 'God and the new math', *Philosophical Studies*, vol. 84, pp. 127-154
- Birnbacher, Dieter (1988) 'Epiphenomenalism as a solution to the ontological mind-body problem', in *Ratio* (New Series) vol. 1, pp. 17-32
- Bolingbroke (*Viscount*; Henry St John) (1754) *Philosophical Works*, London
- Braddon-Mitchell, David, and Jackson, Frank (1996) *Philosophy of Mind and Cognition*, Blackwell: Oxford
- Broad, CD (1925) *The Mind and its Place in Nature*, Routledge and Kegan Paul: London
- Campbell, Keith (1984) *Body and Mind*, (second edition), University of Notre Dame Press: Indiana

- Carter, Brandon (1974) 'Large number coincidences and the anthropic principle in cosmology', in *Confrontation of Cosmological Theories with Observation*, ed. Longair, MS, pp. 291-298, Reidel: Dordrecht. Reprinted in Leslie, 1990, pp. 125-133
- Castell, Paul (1998) 'A consistent restriction of the principle of indifference', *British Journal for the Philosophy of Science*, vol. 49, pp. 387-395
- Chalmers, David (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press: New York
- Chalmers, David (1997) 'Moving forward on the problem of consciousness', in *Explaining Consciousness: the Hard Problem*, ed. Shear, J, MIT Press: Cambridge, Mass., pp. 379-422
- Chalmers, David (1999) 'Materialism and the metaphysics of modality', *Philosophy and Phenomenological Research*, vol. LIX, pp. 473-496
- Chomsky, Noam (1972) *Language and Mind*, Harcourt Brace Jovanovich: New York
- Churchland, Patricia S (1986) *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT Press: Cambridge, Mass.
- Coleman, Stephen (2001) 'Fine-tuning and probability: does the universe require an explanation?', *Sophia*, vol. 40, pp. 8-15
- Crane, Tim, and Mellor, DH (1990) 'There is no question of physicalism', *Mind*, vol. 99, pp. 185-206
- Critchley, M (1979) *The Divine Banquet of the Brain*, Raven: New York
- Culpeper, Nicholas (1652) *The English physitian: or an astrologo-physical discourse of the vulgar herbs of this nation*, London
- Daly, Chris (1998) 'What are physical properties?', *Pacific Philosophical Quarterly*, Vol. 79, pp. 196-217
- Davidson, Donald (1994) 'Donald Davidson', in *Companion to the Philosophy of Mind*, ed. Guttenplan, S, Blackwell: Oxford
- Dennett, Daniel (1988) 'Quining qualia', in *Consciousness in Contemporary Science*, ed. Marcel, AJ and Bisiach, E, Clarendon Press: Oxford, pp. 42-77
- Dennett, Daniel (1991) *Consciousness Explained*, Little, Brown: Boston

- Descartes, René (1640) *Meditations on First Philosophy*, (trans. Cottingham, J) in *Descartes, Selected Writings*, Cottingham, J, Stoothoff, R, and Murdoch, D, Cambridge University Press: Cambridge, 1988
- Earman, John (1975) 'What is physicalism', *Journal of Philosophy*, Vol. 72, pp. 565-7
- Edelman, D (1990) 'A confidence interval for the center of an unknown unimodal distribution based on a sample size 1', *The American Statistician*, vol. 44, pp. 285-7
- Forrest, P (1982) 'Occam's Razor and Possible Worlds', *Monist*, vol. 65, pp. 456-464
- Forrest, P, and Armstrong, DM (1984) 'An argument against David Lewis' theory of possible worlds', *Australasian Journal of Philosophy*, vol. 62, pp. 164-8
- Galen *On the Natural Faculties*
- Gazzaniga, Michael (1970) *The Bisected Brain*, Appleton-Century-Crofts: New York
- Gazzaniga, Michael, and LeDoux, Joseph (1978) *The Integrated Mind*, Plenum Press: New York
- Goertzel, B (1992) 'Quantum theory and consciousness', *Journal of Mind and Behavior*, Vol 13, pp. 29-36
- Gott, J (1993) 'Implications of the Copernican principle for our future prospects', *Nature*, May 27, pp. 315-319
- Guttenplan, S (ed.) (1994) *A Companion to the Philosophy of Mind*, Blackwell: Oxford
- Hacking, Ian (1965) *Logic of Statistical Inference*, Cambridge University Press: Cambridge
- Halliday, D, Resnick, R, and Walker, J (1993) *Fundamentals of Physics*, Wiley: New York
- Hameroff, Stuart, and Penrose, Roger (1997) 'Conscious events as orchestrated space-time selections', in *Explaining Consciousness: the Hard Problem*, ed. Shear, J, MIT Press: Cambridge, Mass., pp. 177-195
- Hawking, Stephen W (1988) *A Brief History of Time*, Bantam Press: London
- Heidegger, Martin (1959) *An Introduction to Metaphysics*, (trans. Manheim, R), Yale University Press: New York
- Heil, John (1998) *Philosophy of Mind*, Routledge: London

- Heil, John, and Martin, CB (1999) 'The ontological turn', *Midwest Studies in Philosophy*, Vol. XXIII, pp. 34-60
- Hodgson, David (1991) *The Mind Matters*, Clarendon Press: Oxford
- Hume, David (1739) *A Treatise of Human Nature*, Selby-Bigge edition, 1888, Clarendon Press: Oxford
- Hume, David (1740) *An Abstract of a Book Lately Published*, London
- Hume, David (1748) *An Enquiry Concerning Human Understanding*, Selby-Bigge 3rd edition, 1975, Clarendon Press: Oxford
- Hume, David (1779) *Dialogues Concerning Natural Religion*, London
- Jackson, Frank (1982) 'Epiphenomenal qualia', *Philosophical Quarterly*, vol. 32, pp. 127-136
- Jackson, Frank (1994) 'Armchair metaphysics', in O'Leary Hawthorne, J and Michael, M, eds., *Philosophy in Mind*, Kluwer: Dordrecht
- Jackson, Frank (1998) *From Metaphysics to Ethics*, Clarendon Press: Oxford
- Kilmister, CW (1984) *Russell*, St. Martin's Press: New York
- Kim, Jaegwon (1993) *Supervenience and Mind*, Cambridge University Press: Cambridge
- Kim, Jaegwon (1994) 'Supervenience', in *A Companion to the Philosophy of Mind*, Blackwell: Oxford, pp. 575-583
- Koons, Robert (1997) 'A new look at the cosmological argument', *American Philosophical Quarterly*, vol. 34, pp. 193-211
- Kripke, Saul (1980) *Naming and Necessity*, Harvard University Press: Cambridge, Mass.
- Leslie, John (1989) *Universes*, Routledge: London
- Leslie, John (1990) *Physical Cosmology and Philosophy*, Macmillan: New York
- Leslie, John (1996) *The End of the World*, Routledge: London
- Leslie, John (2001) *Infinite Minds*, Oxford University Press: Oxford
- Levine, Joseph (1998) [Review of Chalmers, The Conscious Mind], *Mind*, Vol. 107, pp. 877-881
- Lewis, David (1970) *Anselm and Actuality*, *Noûs*, Vol. 4, pp. 175-188; reprinted in Lewis, *Philosophical Papers*, Vol. 1

- Lewis, David (1983a) 'New work for a theory of universals', *Australasian Journal of Philosophy*, vol. 61, pp. 323-377
- Lewis, David (1983b) *Philosophical Papers*, Vol. 1, Oxford University Press: New York
- Lewis, David (1986a) *On the Plurality of Worlds*, Blackwell: Oxford
- Lewis, David (1986b) *Philosophical Papers*, Vol. 2, Oxford University Press: New York
- Lewis, David (1991) *Parts of Classes*, Oxford University Press: New York
- Lewis, David (1994) 'Reduction of mind', in *A Companion to the Philosophy of Mind*, ed. Guttenplan, S, Blackwell: Oxford
- Locke, John (1690) *An Essay Concerning Human Understanding*, Nidditch edition, 1975, Clarendon Press: Oxford
- Locke, John (1700) *An Essay Concerning Human Understanding* [later edition]
- Locke, John *Letters to the Bishop of Worcester [i.e. Stillingfleet]*, reprinted in part in *An Essay Concerning Human Understanding*, 24th edition, 1823. William Baynes and Son: London
- Lockwood, Michael (1989) *Mind, Brain & the Quantum*, Blackwell: Oxford
- Lowe, EJ (1996) 'Why is there anything at all?', *Proceedings of the Aristotelian Society*, Supplementary Vol. LXX, pp. 111-20
- Lucretius *De Rerum Natura*, (trans. Leonard, WE), 1957, Ward Ritchie Press: Los Angeles
- Mackie, JL (1974) *The Cement of the Universe*, Oxford University Press: Oxford
- Mason, Richard (1988) 'Logical possibility', *Metaphilosophy*, vol. 19, pp. 11-24
- McGrew, T, McGrew, L, and Vestrup, E (2001) 'Probabilities and the fine-tuning argument: a sceptical view', *Mind*, vol. 110, pp. 1027-1037
- Monier-Williams, M (1899) *Sanskrit-English Dictionary*, Clarendon Press: Oxford
- Nagel, Ernest (1961) *The Structure of Science*, Routledge and Kegan Paul: London
- Nagel, Thomas (1974) 'What is it like to be a bat?', *Philosophical Review*, vol. 4, pp. 435-450

- Nolan, Daniel (1996) 'Recombination unbound', *Philosophical Studies*, vol. 84, pp. 239-262
- Nozick, Robert (1981) *Philosophical Explanations*, Belknap Press (Harvard): Cambridge, Mass.
- Oxford English Dictionary* (1971) Oxford University Press: Oxford
- Partridge, Eric (1966) *Origins: A Short Etymological Dictionary of Modern English*, Routledge and Kegan Paul, London
- Passmore, J (1968) *A Hundred Years of Philosophy*, Harmondsworth: Penguin
- Penrose, Roger (1989) *The Emperor's New Mind*, Vintage: London
- Penrose, Roger (1990) *The Emperor's New Mind*, Oxford University Press: Oxford
- Penrose, Roger (1994) *Shadows of the Mind*, Oxford University Press: Oxford
- Place, UT (1956) 'Is consciousness a brain process?', *British Journal of Psychology*, vol. 47, pp. 44-50
- Plato *Timaeus*, trans. Jowett, in *Collected Dialogues*, 1961, ed. Hamilton and Cairns, Princeton University Press: Princeton
- Poe, Edgar Allan (1832) *Bon Bon*, in *The Works of the Late Edgar Allan Poe*, 1850, vol II, pp. 479-495
- Pruss, Alex (2001a) *Possible worlds: what are they, and what are they good for?*, unpublished PhD thesis, University of Pittsburgh
- Pruss, Alex (2001b) 'The actual and the possible', (preprint:) forthcoming in *Blackwell Guide to Metaphysics*, Richard M. Gale (ed.), Blackwell: Oxford
- Quine, WV (1960) *Word and Object*, MIT Press: Cambridge, Mass.
- Quine, WV (1976) *The Ways of Paradox and Other Essays*, Harvard University Press: Cambridge, Mass.
- Quine, WV (1981) *Theories and Things*, Harvard University Press: Cambridge, Mass.
- Rosen, Gideon (1990) 'Modal fictionalism', *Mind*, Vol. 99, pp. 327-354
- Russell, Bertrand (1921) *Analysis of Mind*, Allen and Unwin: London
- Russell, Bertrand (1927a) *Analysis of Matter*, reprinted 1954, Allen and Unwin: New York

- Russell, Bertrand (1927b) *An Outline of Philosophy*, Unwin Books: London
- Russell, Bertrand (1945) *A History of Western Philosophy*, Simon and Schuster: New York
- Russell, Bertrand (1956) *Logic and Knowledge*, Allen and Unwin: London
- Scheffler, Israel (1950) 'The new dualism: psychological and physical terms', *Journal of Philosophy*, Vol. 47, pp. 732-752
- Schlesinger, George (1984) 'Possible worlds and the mystery of existence', *Ratio*, vol. XXXVI, pp. 1-17
- Shalkowski, Scott (1994) 'The ontological ground of the alethic modality', *The Philosophical Review*, vol. 103, pp. 669-688
- Shear, Jonathan (ed.) (1997) *Explaining Consciousness: the Hard Problem*, MIT Press: Cambridge, Mass.
- Smart, JJC (1989) *Our Place in the Universe*, Blackwell: Oxford
- Smith, Adam (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*, London
- Sterelny, Kim (1990) *The Representational Theory of Mind*, Blackwell: Oxford
- Strawson, Galen (1989) *The Secret Connexion*, Clarendon Press: Oxford
- Strawson, Galen (1994) *Mental Reality*, MIT Press: Cambridge, Mass.
- Teller, Paul (1984) 'A poor man's guide to supervenience and determination', *Southern Journal of Philosophy*, vol. XXII, pp. 137-162
- The Australian* (1994) 'Researchers finally slice the pi', 12 April, p. 22
- Themistius *On Aristotle on the Soul*, trans. Todd, 1996, Duckworth: London
- Thomas, Holly (1993) 'Modal realism and inductive scepticism', *Noûs*, vol. 27, pp. 331-354
- Tye, Michael (1996) *Ten Problems of Consciousness*, MIT Press: Cambridge, Mass.
- van Fraassen, Bas (1989) *Laws and Symmetry*, Clarendon Press: Oxford
- van Fraassen, Bas (1996) 'Science, materialism, and false consciousness', in *Warrant in Contemporary Epistemology: Essays in Honor of Alvin Plantinga's Theory of Knowledge*, (ed.) Kvanvig, Jonathan Rowman Littlefield, pp. 149-181

- van Inwagen, Peter (1996) 'Why is there anything at all?', *Proceedings of the Aristotelian Society*, Supplementary Vol. LXX, pp. 95-110
- von Weizsäcker, Carl (1981) *The Unity of Nature*, Farrar, Straus and Giroux: New York
- Ward, James (1899) *Naturalism and Agnosticism*. [2 volumes]. Adam and Charles Black: London
- Watts, Alan (1962) *The Way of Zen*, Penguin Books: Harmondsworth
- Wedberg, Anders (1984) *A History of Philosophy*, Vol. 3, Clarendon Press: Oxford
- Wilkes, Kathleen (1988) '——, yìshì, duh, um, and consciousness', in *Consciousness in Contemporary Science*, ed. Marcel, AJ and Bisiach, E. Clarendon Press: Oxford, pp. 16-41
- Wilson, Mark (1985) 'What is this thing called "pain"? The philosophy of science behind the contemporary debate', *Pacific Philosophical Quarterly*, vol. 66, pp. 227-267
- Yablo, Stephen (1999) 'Concepts and consciousness', *Philosophy and Phenomenological Research*, vol. LIX, pp. 455-463