OPEN ACCESS

University of BRISTOL

This electronic thesis or dissertation has been downloaded from the University of Bristol Research Portal, http://research-information.bristol.ac.uk

Author:
Paterson, Niall

Title:
Doxastic Dispositions

An Essay on the Metaphysics of Belief

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited on the University of Bristol Research Portal. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

•Your contact details
•Bibliographic details for the item, including a URL
•An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Doxastic Dispositions: An Essay on the Metaphysics of Belief

Niall J. Paterson

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Arts.

August, 2018

Word count: 69,344

Abstract

Is belief a dispositional state? This thesis looks at what sense, if any, belief is dispositional, and what that claim amounts to. My focus is on the extent to which dispositions enter into the individuation of belief-content pairs. It has commonly been supposed that dispositions cannot, by themselves, individuate belief-content pairs for at least two reasons. First, due to the influence of functionalist accounts of belief, on which beliefs (despite perhaps bearing dispositional properties) are taken to be the occupants of causal roles. Second, due to the 'Twin-Earth' style counterexamples to internalist theories of content. In response to the first, I argue against a functionalist account of belief in favour of a dispositional account. In response to the second, I argue that whilst content externalism shows that beliefs are not individuated in terms of actually possessed dispositions, that belief-content pairs are nonetheless dispositionally individuated, albeit by dispositional ideals, which I take to be a kind of functional norm.

Preface

This thesis is on the metaphysics of dispositions, and the extent to which belief-content pairs admit of dispositional individuation. Originally, my aim was to defend the knowledge norm on belief. Alas, it quickly dawned upon me that I had tackled a veritable Cerberus. A lesson on monster slaying: When starting out, tackle single-headed beasts!

Here is a brief summary of the thesis. There are two parts. At the start of each I have included a short introduction. The first part is on the metaphysics of dispositions, the second on the relationship between dispositions and the attitude of belief. There are seven chapters. Each chapter (quite satisfyingly, I might add) has two sections.

The first chapter, *The Conditional Analysis*, explores the semantics of dispositional ascriptions. The central aim is to argue against a host of 'conditional' accounts and to support a causal account of dispositional ascriptions, on which to ascribe a disposition is to ascribe the possibility of a specific sort of causal relation, holding between certain kinds of stimulus event and certain kinds of manifestation event.

The second chapter, *The Identity Theory*, looks at the ontological status of dispositional properties. According to what was once a popular account, dispositions are identical to their causal bases. Orthodoxy has it that, at most, one may endorse a *token-token* rather than a *type-type* identity theory, due to the problem of multiple realisability. I argue this is mistaken: There is multiple realisation at the token level. I then argue that dispositions are modal properties, more precisely those modal properties that constitute their *analysans*.

The third chapter, *Functions*, explores the relationship between functions and dispositions. On a fairly standard view, the canonical endorsement of which is given by Robert Cummins (1975), functional ascriptions entail dispositional ascriptions: to function as an *F* is, in part, to bear a dispositional property. In contrast, Mumford has argued that not only is this claim false, but its converse is true: dispositions entail functions, though functions do not entail dispositions. I argue that *both* Mumford's *and* Cummins' support for their respective positions rests on equivocation. I then provide a non-equivocatory argument of my own for Cummins' position.

In the fourth chapter, *Functional Norms*, I outline the notion of a normative judgement, and distinguish teleological from what I call 'rule' or deontological normativity.

I argue that they are distinct kinds of normativity, as the relevant judgements bear distinct causal powers. Whilst rule normative judgements play a *motivational* role in cognition, teleo-normative judgements do not. I finish by explicating the notion of *functional norm*.

In the fifth chapter, *The Attitude of Belief*, I outline belief, and the notion of a belief-content pair. I go on to distinguish several distinct ways in which beliefs may be said to be 'dispositional'. On one view, which I call *Simple Behaviourism*, beliefs are individuated in dispositional terms. On another, dispositions are employed to distinguish kinds of mental state: dispositional states vs occurrent states. I argue that the distinction between 'occurrent' and 'dispositional' belief has been misapplied, and has led to widespread endorsement of an overpermissive account of what agents believe. I then outline and reject three purported differences between dispositions and beliefs given by Armstrong (1973).

In the sixth chapter, *Belief as Dispositional*, I argue that belief is a dispositional state. First, I outline 'functional analysis', and then functionalism in the philosophy of mind, on which beliefs are individuated in terms of the occupation of causal roles. I then show how Functionalism is a better theory than Simple Behaviourism, but argue that a novel dispositional theory, which I call Complex Behaviourism, is preferable to Functionalism.

In the seventh and final chapter, *Dispositional Ideals*, I develop what I call a 'regulatory' theory of content. First, I outline externalism in the philosophy of mind, and the argument for the inconsistency of externalism and complex behaviourism. Following that, I outline and reject a number of ways to accommodate externalism. Finally, I provide my own regulatory theory, on which belief-content pairs are individuated by functional norms, which are taken to be a kind of dispositional ideal.

Over the last four years, I have had the immense fortune to be surrounded by numerous highly talented and intelligent individuals. I owe them more than I can express: to them, I give my thanks. Amongst those who helped me along the way, and in no particular order, are Geoff Keeling, Leo Graham, Aaron Guthrie, Chris Burr, Oliver Lean, Shaun Stanley, Nemo D'Qrill, Josh Habgood-Coote, Catrin Campbell-Moore, Neil Coleman, Mark Pinder, Charles Janson, Havi Carel, Lucienne Spencer, Julien Dutant, Jonathan Vogel, James Wilson, Alan Wilson, Leia Hopf, Richard Pettigrew, Andy Suttie, Vanessa Seifert, Ji-Young Lee, Issac Kean, Ben Springett, Arsham Nejad Kourki, Anya Farrenikova, Nick Axten, James Ladyman, Nick Cosstick, Anthony Everett, Finn Spicer, Kit Patrick, Karim Thebault, Ana Cretu, Maria Lasonon-Aarnio, and Samir Okasha. Incompetently, I am bound to have missed many off the list. To them, I extend my apologies. The team at Friska, where this thesis was written, also deserve a mention, for keeping moods high and the caffeine flowing.

I was also extremely lucky to have benefited from excellent supervision, both from

Alexander Bird, and in the final year by Jason Konek. Their critical comments have been both insightful and invaluable. To them, I offer special thanks, not only for their time and hard work, but for their support. My final thanks goes to my father. Without his help, this thesis would never have been written, and I would have not had the self indulgent pleasure of writing a Ph.D thesis. Certainly, I would never have had the joy of reading and thinking about these issues, and I would be worse off for it. I dedicate this thesis to him.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Contents

I Functions & Dispositions **9**

1 The Conditional Analysis **13**
1.1 Canonical Reduction 15
1.2 Dispositional Semantics 18
1.2.1 Three conditional analyses 18
1.2.2 The causal contingent account 27

2 The Identity Theory **37**
2.1 Two Identity Theses 38
2.1.1 Dispositions as causal bases 38
2.1.2 Reformulation: token identity 44
2.2 Dispositions as Modal Properties 46
2.2.1 Plural realisation 46
2.2.2 Dispositions as modal properties 51

3 Functions **57**
3.1 Four Senses of 'Function' 58
3.1.1 Mathematical and social 58
3.1.2 Causal role functions 60
3.1.3 Teleological functions 65
3.1.4 Relation between causal and teleological functions 70
3.2 Functions and Dispositions 73
3.2.1 Two equivocatory arguments 74
3.2.2 Functions entail dispositions 81

4 Functional Norms **85**
4.1 Normativity 86
4.1.1 Normative judgements 86
4.1.2 Upholding to standards 89
4.2 Teleological Normativity 92
4.2.1 Teleological standards 92

4.2.2 Norms and conventions 97
4.2.3 Functional norms 101

II Doxastic Dispositions 105

5 The Attitude of Belief 109
5.1 The Propositional Requirement 110
5.2 Belief as Dispositional 120
5.2.1 Simple behaviourism 120
5.2.2 Armstrong's three differences 127

6 Belief as Dispositional 131
6.1 Functionalism 132
6.1.1 Functional analysis 132
6.1.2 Functionalism 134
6.2 Complex Behaviourism 142
6.2.1 Token states 142
6.2.2 Back to plural realisation 147
6.2.3 Accommodating holism 148

7 Dispositional Ideals 153
7.1 The Externalist Challenge 154
7.1.1 Content externalism 154
7.1.2 The externalist challenge 157
7.1.3 Context relativity 159
7.2 Dispositional Ideals 165
7.2.1 Classification and regulation 166
7.2.2 Dispositional ideals 168

References 177

Part I

Functions & Dispositions

I want to perceive and understand the hidden powers and laws of things, in order to have them in my power.

Salvador Dalí

Introduction to Part I

Dispositions form a notoriously puzzling class of property.[1] Paradigmatic examples include fragility, flammability, solubility, and poisonousness. What is primarily puzzling is the way in which dispositions 'lie latent'; to use Goodman's phrase they are like *'threats and promises'*. To say that a glass is fragile is not to say it is broken, to say that a match is flammable is not to say it is on fire. Rather, the glass is fragile if it *threatens* to break, the match is flammable if it *promises* to ignite when drawn across the striker. Those events that reveal a dispositional property are called its *stimulus conditions*. Those events a disposition threatens or promises to bring about are called its *manifestations*. As Armstrong remarks, dispositional manifestations have an '*amazing property, metaphysically speaking*' — they may be non-existent.[2] Dispositions may never manifest; a glass may never smash, a match may never light. Such manifestations may, then, like the content of some threats and promises, be mere *possibilia*.

Philosophers often say that belief is dispositional. What does this claim amount to? The first two chapters of this part are on the nature of dispositional properties. In the first, I develop and support an account of dispositional ascriptions. The account developed construes dispositional ascriptions as ascriptions of a certain kind of causal, modal property. Some metaphysicians may then be tempted to identify the disposition with its *causal basis*, that is—the causally efficacious properties it bears, such that in virtue of its bearing those properties, the modal property obtains. In the second, I argue this is mistaken: Dispositions are not identical to their causal bases, *not even token-token identical*. Re-deploying the problem of multiple realisability at the token level, I argue that we have reason to not identify dispositions with their token causal bases.

In the second two chapters, I turn to functional ascriptions. In the third, I distinguish a number of distinct senses of the term 'function', and explicate two kinds of ascription used to denote what I call the 'causal role' sense. I will argue, *contra* Mumford (2003), that functional individuation is still a kind of dispositional individuation. I show that Mumford's argument, and arguments in favour of its negation, rest on equivocation. I then provide a non-equivocatory argument of my own. In the fourth, I look at the relationship between the 'teleological' sense of 'function' and normative judgements. The central aim is to explicate the notion of a 'functional norm'.

[1] It is worth noting that 'disposition' has somewhat of an 'umbrella sense'. It should most likely be taken to include, non-exhaustively, tendencies, potentialities, capacities, powers, and propensities.

[2] (Armstrong, 1968, p. xvii).

Chapter 1

The Conditional Analysis

Introduction

Long ago, in a philosophy dominated by logical empiricism, dispositions ran amok. For the empiricist, statements were either analytic, empirically verifiable, or empty. Dispositional ascriptions are patently not analytic: The presence of fragility does not follow from the definition of 'glass'. Moreover, dispositional ascriptions are hardly *meaningless*. If the trichotomy holds true, then dispositional ascriptions must be empirically verifiable. But therein lies the problem: Dispositions seem, in a certain way, to *resist observational verification*.[1] Of course, one can verify a dispositional ascription through observation of the disposition's manifestations. But some dispositions lie latent throughout their persistence: They *never* manifest. It seems to follow that some dispositional ascriptions are never empirically verifiable.

Mellor (1974) put things rather dramatically: Dispositions were as '*shameful in many eyes as pregnant spinsters used to be — ideally to be explained away, or entitled by a shotgun wedding to take the name of some decently real categorical property*'. Mellor's own attempt at nominal restoration was largely unsuccessful. But just three years prior, Kripke's *Naming and Necessity* had come into print. And with a new found respect of modality developing, dispositions promised to appear respectable again. And so it emerged: what is known as the 'simple conditional analysis'.

The analysis first appeared in Martin's '*Dispositions and Conditionals*', though by that time it was already a widely held position. As Martin, and subsequently a wide range of philosophers have shown, the simple conditional analysis fails. Nonetheless, metaphysicians did not lose hope. Many have thought—and still do—that a conditional analysis is still possible. It may require a tinker, and the end result may be less simple, but in principle the analysis is correct. In what follows, I am going to argue that all conditional analyses, no matter what form they take, are bound to fail. As I shall argue,

[1] Cf. (Papineau, 1987, p. 79).

dispositions are not analysable in terms of conditionals *of any kind*. Dispositions, on the view to be endorsed, are a certain kind of causal property. The lure of conditional analyses follows only from the close—though extensionally non-equivalent—relationship that holds between causal relations and conditionals. Loosely following Vetter (2014), I will then propose an alternative: What I call the *causal contingent* account.

Now before we get into the thick of it, I wish to say a few brief words on my understanding of the term 'analysis'. It is commonplace for contemporary philosophers—particularly epistemologists—to eschew talk of 'analyses'.[2] Much of this is due to the influential work of Timothy Williamson (2000), who argued that (a) there is no reason to suppose that a conceptual analysis of knowledge in the Fregean/Russellian sense (i.e., the *mereological* sense) is possible and that (b) a pessimistic induction may be run for the providing of non-circular necessary and sufficient conditions on knowledge. Similar doubts may be raised in the present case. Now, some may say that mere necessary and sufficient conditions do not, strictly speaking, constitute an 'analysis'. But as I use the term 'analysis', a set of necessary and sufficient conditions will count. Whilst truth-conditional semantics is irredeemably incomplete, truth-conditions do provide insights into concepts. That is all I claim.

Here is the plan. There are two sections. In the first, I distinguish two kinds of disposition: Conventional dispositions and canonical dispositions. I then outline the view that the former are analysable, or semantically reducible if 'analyse' is too strong, in terms of the latter. I suggest that such a view faces certain challenges, but assume that such analyses are possible, if epistemically inaccessible. In the second, I outline the standard way of analysing conventional dispositions, namely in terms of conditionals, rejecting a variety of attempts. Finally, I outline a recent 'contingent' account developed by Barbara Vetter (2014), and support a modified version, what I call the 'causal contingent' account.

Here are the aims:

[1.1] To outline the distinction between *canonical* and *conventional* dispositional expressions.

[1.2] To outline and reject three analyses of dispositions: the 'simple' analysis, the 'causal' analysis, and the 'fainthearted' analysis.

[1.3] To sketch and defend the 'causal contingent' account.

The first section deals with [1.1] and [1.2]. The second section deals with [1.3].

[2] For some vivid examples, see (Greco, 2010) and (Pritchard, 2012).

1.1 Canonical Reduction

Cigarettes are flammable, puddles of water are not. A rod of steel sufficiently cooled may become fragile, a match may lose its flammability if drawn across the striker. The upshot of these remarks is that dispositions apply to some objects but not to others and that the bearing of dispositions may wax and wane. Under what conditions is it true that a given object at a given time has some dispositional property? In this section, I will reject the standard 'conditional' answer to this question.

Dispositional ascriptions are typically taken to fall into two classes. There are *conventional* dispositions and *canonical* dispositions. The dispositional properties we have looked at so far (fragility, flammability, and so on) are conventional. In contrast, canonical dispositions are denoted by statements of the form:

Canonical Disposition x is disposed to M if S.

Instances include:

1a. Petrol is disposed to set on fire if ignited.

1b. Tim is disposed to get angry if provoked.

1c. Jelly is disposed to wobble if prodded.

The distinguishing mark of canonical dispositional ascriptions is that they make explicit reference to (i) stimulus conditions (igniting/provoking/prodding) and (ii) manifestations (setting on fire/becoming angry/wobbling). From now on, let 'Dx' denote conventional dispositions, and '$Dx_{(s,m)}$' denote canonical dispositions.

It has been widely thought that conventionals are analysable into canonicals. On this view, for a glass to be fragile is simply for it to have certain canonical dispositions: It would smash if dropped, crack if struck, and so on. Call this the *canonical reduction thesis*. Why should we accept the thesis? Surprisingly, the view is rarely offered much in the way of support. There do, however, appear to be at least two reasons to suppose it holds. They are as follows.

1. The semantic reason One can grasp a conventional dispositional concept by being given a set of canonical ascriptions. Suppose, for example, that Jones tells Smith 'Be careful, that glass is fragile!' If Smith were to respond, 'What do you mean 'fragile'?' then Jones could respond 'Well, if you drop it, then it will smash, and if you strike it, then it will crack...'. In such a case, Smith may come to grasp the concept of fragility by being given a set of conventional dispositions. Why is this possible? If the canonical reduction thesis is true, we have an answer.

2. The epistemic reason One can come to know that certain kinds of object possess a conventional dispositional property by coming to know that they possess certain

canonical dispositions. For example, Smith may come to know that matches are flammable by observing Jones draw several across the striker, and the subsequent ignitions that follow. In other words, Smith comes to know that matches are flammable by coming to know that they are disposed to ignite when drawn across the striker.

These reasons are admittedly inconclusive. Nevertheless, they give *prima facie* support for the canonical reduction thesis. Unfortunately, however, whilst intuitively plausible, getting precise on how exactly particular conventionals canonically reduce is less straightforward than one might like. The task is not so difficult for what Ryle called 'single track' dispositions, which are ones that reduce to a unique stimulus-manifestation pair. Put schematically, and letting $\rightarrow^r$ denote the relevant notion of reduction, we could formalise such reductions as:

Single Track Reduction $\mathrm{D}x \rightarrow^r \mathrm{D}x_{(s,m)}$

Simple though such reductions would be, it is far from clear that there exist any single track conventional dispositions. For any given conventional disposition, there appears to be a variety of stimuli which could cause the disposition to manifest, and a wide variety of reactions may be properly regarded as its manifestation. Fragility serves as a vivid example. A glass may manifest its fragility upon being dropped, thrown, twisted, or struck. Its manifestations may differ also; it may crack, smash, shatter, or splinter.

It might be thought that the various manifestations and stimuli, whilst appearing disparate, may nonetheless be classified. For example, the stimulus condition for fragility may be taken to be the exertion of a certain degree of force, and the manifestations may be classified in terms of their all being a kind of loss of structural integrity. But whilst this may work for some dispositions, it won't work for all. An example much discussed, originally given by Ryle, is 'knowing French'. Consider:

> A common view is that mental dispositions are often multi-track. To use Ryle's example, knowing French seems to be dispositional, but its manifestations may be various—talking French, writing French, obeying an order in French or even changing mental state when reading or hearing something in French. (Bird, 2009b)

Where a conventional disposition is multi-track, its canonical reduction will involve more than one canonical disposition. The problem then becomes specifying how the various canonical dispositions must be truth-functionally related in order to give rise to the reduced conventional property. Put schematically, and letting '$\mathrm{D}x_{(s_n,m_n)}$' denote specific canonicals, two possibilities would be as follows:

$\wedge$-*Reduction* $\mathrm{D}x \rightarrow^r \mathrm{D}x_{(s_1,m_1)} \wedge \mathrm{D}x_{(s_2,m_2)} \wedge \mathrm{D}x_{(s_3,m_3)} \ldots \wedge \mathrm{D}x_{(s_n,m_n)}$

$\vee$-*Reduction* $\mathrm{D}x \rightarrow^{r} \mathrm{D}x_{(s_1,m_1)} \vee \mathrm{D}x_{(s_2,m_2)} \vee \mathrm{D}x_{(s_3,m_3)} \ldots \vee \mathrm{D}x_{(s_n,m_n)}$

Both are standardly taken to be too crude: the first too strong, the latter weak. Consider again Ryle's example of knowing French. A conjunctional analysis would require that one is able to respond to *every* possible French sentence. But this is patently false; even native speakers of the language may not know the whole French vocabulary. Similar points apply to more standard dispositional properties. Tempered glass may not crack into large pieces, despite this being one of the possible manifestations of fragile objects. But that is insufficient to disqualify it from falling under the concept's extension; tempered glass is fragile, less fragile, but fragile nonetheless.

Do similar points apply to the disjunctional analysis? It may be thought that knowing how to respond to a single sentence of French is insufficient to count as someone who knows French. On closer reflection, however, this is unclear. After all, dispositions are gradational. Petrol is *more flammable* than cotton, and lead *less poisonous* than arsenic. It may be that when someone knows only one sentence of French, that one *does* know French, albeit very weakly. We would typically not attribute knowledge of French to such speakers, just as we would typically not attribute flammability to a plastic bag. But plastic bags *are* flammable; our choice of attribution in such cases is plausibly guided by pragmatics and is thus not evidence for the presence or absence of dispositional properties. So too, it may be, for knowing French.

Unfortunately, an alternative counterexample raised by Stephen Mumford may be raised in objection. He writes:

> In the case of love, for instance, buying some flowers, if passing a florist, may be one of the disjuncts in the analysis [...] one may buy flowers for a colleague recuperating in hospital. (Mumford, 1998, p. 80)

Our response to the previous case here is inert. To argue that one loves both one's colleague and one's spouse, the former just more weakly is to misgrasp the concept of love.

Further problems may remain for the disjunctional analysis. If so, it may be that conventionals are only analysable into canonicals related by complex blends of different truth functional relations. For instance, perhaps there are certain 'core' necessary canonicals, and other 'peripheral' unnecessary but sufficient canonicals. In what remains, I will assume that *some* kind of analysis of conventionals into canonicals is possible, complex though it may be. But in any case, even that assumption granted, an analysis of conventionals in terms of canonicals is, at best, incomplete, as it fails to analyse out of dispositional terms.[3] To complete the analysis, we require an analysis

[3]It might be argued that, like the concepts of modality, dispositional concepts are only definable in terms of other dispositional concepts. I'll assume for present purposes that this is false.

of canonical dispositions in non-dispositional terms. That is the task to which I now turn.

1.2 Dispositional Semantics

So far, I have dealt with the first aim, namely

> [1.1] To outline the distinction between *canonical* and *conventional* dispositional expressions.

In this section, I turn to the second and third. I consider and reject three analyses of dispositions, all united by a common 'conditional' theme. I then sketch and defend a modified version of Vetter's non-conditional alternative.

1.2.1 Three conditional analyses

The standard view, at least historically speaking, is that canonical dispositions are analysable into conditional statements that relate events; stimulus conditions for the antecedent, manifestations for the consequent. This garners support from the grammatical, and thereby apparent *logical* form of statements that contain canonical dispositional expressions. We say that the glass is disposed to break *if* struck, or that the match is disposed to ignite *if* drawn across the striker. Thus, *a priori* reflection makes it appear that canonicals are equivalent to statements of the form:

<u>Conditional Analysis</u>

$\mathrm{D}x_{(s,m)}$ iff.

1. $\mathrm{S}x \rightarrow^{x} \mathrm{M}x$

where '$\rightarrow^{x}$' denotes some kind of conditional.[4] The trick to making the analysis work, at least it has been supposed, is getting the right sort of conditional. Early attempts given by Carnap were hindered by the lack of a sufficiently powerful logic.[5] But in the wake of a Kripkean modal enlightenment,[6] modal conditionals became a respectable analytical tool. Dispositions were an obvious target for modal *analysans*, in part due to their connection, noted at the start, to *possibilia*. As Goodman wrote:

[4] If we were to be more precise, we would time index both the *analysandum* and the *analysans*. Thus, we would say that something is disposed, at a given time, to M given S, just in case if it were to undergo M at t, then it would manifest at t^*, where t^* is later than t.

[5] See (Carnap, 1936).

[6] (Kripke, 1980).

> To find non-dispositional, or manifest, predicates of things we must turn to those describing events [...] To apply such a predicate is to say that something specific actually happens with respect to the thing in question; while to apply a dispositional predicate is to speak of what can happen [...] they seem to be applied to things in virtue of possible, rather than actual occurrences... (Goodman, 1954, pp. 41-42)

It seems to follow, then, that '$\rightarrow^x$' must denote some kind of modal conditional. Which kind? Early attempts were influenced by the fact that explanations of dispositions are often given in the subjunctive mood. Consider:

> To say that the lump of sugar is soluble is to say that it would dissolve, if submerged anywhere, at any time and in any parcel of water. (Storer, 1951)
>
> The peculiarity of all such definitions is the occurrence of the type "If so and so *were to happen*, then such and such *would be* the case." (Ryle, 1949) (Italics in original.)

A fairly standard way to interpret these subjunctive expressions is as counterfactual conditionals, where the antecedent need not be false. With the developments in counterfactual logic, primarily due to Lewis (1973) and Stalnaker (1968), what was known as the 'simple conditional analysis' of canonical dispositions emerged, which may be formalised thus:

<u>Simple Conditional Analysis</u>

$\mathrm{D}x_{(s,m)}$ iff.

1. $\mathrm{S}x \mathrel{\Box\!\!\rightarrow} \mathrm{M}x$

Put in the language of possible worlds, and following the Lewisian semantics, this reads: In the closest worlds in which the stimulus condition obtains, so too does the manifestation. For example, a glass is disposed to shatter if struck iff. in the closest worlds in which the glass is struck, the glass shatters.

Despite the analysis' formal elegance and initial plausibility, philosophers soon provided what Bonevac *et al* call a 'bestiary' of counterexamples to the simple conditional analysis.[7] Two famous examples were given by Martin (1994). The first targets the necessity of the analysis, the second targets the sufficiency. They involve what he called an 'electro-fink'. We start with the assumption that the conventional dispositional term 'live', as it features in statements such as 'x is live', is analysable as follows:

> (α) An electrical current is disposed to flow from x to a conductor if touched by a conductor.

[7](Bonevac *et al.* , 2011).

on the simple conditional analysis this should be taken as equivalent to the following counterfactual conditional:

> (β) x is touched by a conductor $\Box\!\!\rightarrow$ an electrical current flows from x to the conductor.

Martin then gives a case in which the electro-fink makes it the case that (1) (β) is true, but (α) is false (this gives the lack of sufficiency), and (2) in which (α) is true, but in which (β) is false (this gives the lack of necessity). Here are the examples:

> *Insufficient Fink* A dead wire is connected to the electro-fink. The fink makes it the case that if the wire is touched by a conductor, then an electrical current is passed through the wire, which in turn makes electricity flow from the wire to the conductor.
>
> *Unnecessary Fink* A live wire is connected to the electro-fink. The fink makes it the case that if the wire is touched by a conductor, then no electricity is allowed to flow through the wire, which in turn makes it the case that no electricity flows from the wire to the conductor.

In *Insufficient Fink*, (α) is false (the wire is not live), but (β) is true, for if the wire were touched by a conductor, it would suddenly become live, and would thereby allow electricity to flow to the conductor. So the analysandum may be satisfied when the analysans is not. The analysis is insufficient. In *Unnecessary Fink*, (α) is true (the wire is live), but (β) is false, for if the wire were to be touched by a conductor, it would suddenly become dead, and would thereby *not* transmit electricity to the conductor. So the analysans may be satisfied when the analysans is not. The analysis is unnecessary.

It is worth noting that Martin's counterexamples work by fixing the semantic value of the counterfactual conditional, whilst changing the intrinsic properties of the wire, and by doing so its dispositions. This is not a necessary feature of counterexamples to the simple conditional analysis. Consider the following two from Johnston (1992):

> *Mimic* A gold chalice is not fragile. Nevertheless, an angel takes a disliking to it, taking it to be sacrilegious, and thus decides to cause the chalice to shatter if struck or dropped.
>
> *Mask* A fragile glass is carefully wrapped in bubble wrap, such that the bubble wrap causes, upon striking or dropping, the glass to be protected from breaking.

Consider the following canonical disposition:

> (γ) x is disposed to shatter when struck.

on the simple conditional analysis, this should be equivalent to:

(δ) x is struck $\Box\!\!\rightarrow$ x shatters

However, in *Mimic*, the gold chalice satisfies (δ). If the chalice were struck, then it would shatter. But *ex hypothesi*, the chalice is not so disposed. The *analysandum* is false, but the *analysans* true. So the account is insufficient. Moreover, in *Mask*, the glass satisfies (γ). The glass is disposed to shatter if struck. But the bubble wrap falsifies the *analysans* — given that the glass is wrapped in bubble wrap, it would not shatter if struck. The *analysans* is false, but the *analysandum* true. So the account is unnecessary.

Importantly, in both cases it is something *extrinsic* that does the work. Mimics are extrinsic insufficient finks, masks are extrinsic unnecessary finks. The intrinsic properties of the chalice and glass remain the same, the manifestations are either caused to occur or prevented from occurring by extrinsic factors. Similar points apply to their dispositions. A glass wrapped in bubble wrap retains its fragility, despite losing its ability to manifest its fragility. Similarly, a gold chalice does not become fragile upon the Angel's decision to make it shatter if struck or dropped. So the efficacy of the counterexamples do not require changes in either the dispositions, nor the intrinsic properties of the object in question.[8]

An appealing patch to the analysis, to help it avoid the counterexamples given above, is to introduce a casual requirement into the conditional. Lewis (1997) takes this approach, and whilst Armstrong does not attempt to analyse the concept, he takes causation to be amongst its essential components. A vivid example of this may be found in *A Causal Theory of Mind*, where he writes:

> An example of a causal concept is the concept of poison. The concept of poison is the concept of something that when introduced into an organism causes that organism to sicken and/or die. This is but a rough analysis of the concept the structure of which is in fact somewhat more complex and subtle than this. [...] But the essential point about the concept of poison is that it is the concept of *that*, *whatever it is*, *which produces certain effects*. (Armstrong, 1973, p. 20)

He does not say so explicitly, but his choice of concept here is dispositional, and it is precisely because the concept is dispositional that it is causal.[9] Whilst I will agree that the concept is causal, if the remarks below hold true, then the last of Armstrong's statements cannot be quite right. Nonetheless, I take it to be very much on the right track.

Here is Lewis' account:

[8]For further counterexamples and discussion, see (Fara, 2005), (Smith, 1977), (Manley & Wasserman, 2008), and (Manley & Wasserman, 2011).

[9]Cf. (Armstrong, 1969).

> Something x is disposed at time t to give response r to stimulus s iff, for some intrinsic property I that x has at t, for some time t' after t, if x were to undergo stimulus s at time t and retain property I until t', s and x's having of I would jointly be an x-complete cause of x's giving response r. (Lewis, 1997, p. 157)

As he duly notes, 'an unlovely mouthful!' Simplicity is *not* on the side of the causal analysis. Letting '$\rightarrow^c$' denote an x-complete causal relation, and t some time, we may formalise this as:

<u>Causal Analysis</u>

> $\mathrm{D}x_{(s,m)}$ at t iff. there exists some intrinsic property I, and time t^* later than t such that
>
> 1. ($\mathrm{S}x$ at $t \wedge \mathrm{I}x$ from t to t^*) $\Box\!\!\rightarrow$ (($Sx \wedge \mathrm{I}x$) $\rightarrow^c \mathrm{M}x$)

Let's consider a case. Suppose we take a glass that is disposed to shatter if struck. An ascription of the disposition means: the glass has some intrinsic property, such that if it were to be struck and were to retain that property, the striking and that intrinsic property would be an x-complete cause of the glass's shattering. What is this unfamiliar phrase: 'x-complete cause'? Lewis writes:

> We have the notion of a complete cause of an effect. (Mill called it the 'whole cause'. I use a different term to mark that we need not be committed to Mill's own analysis.) We can introduce a restriction of that notion: a cause complete in so far as havings of properties intrinsic to x are concerned, though perhaps omitting some events extrinsic to x. For short, an 'x-complete cause'. (Lewis, 1997, p. 215)

This passage is not very clear. Mill's notion is one of the sum of all of those events and conditions which, taken together, are jointly sufficient for the caused event's occurrence. Patent enough. But what is a complete cause 'in so far as havings of properties intrinsic to x are concerned' supposed to be? Lewis' interpreters appear to fare no better. Consider the following passages:

> I take this to mean [...] (i) s is a cause of x's giving response r; (ii) x's having B is a cause of response r; and (iii) if any of x's other intrinsic properties (besides B) had been different, x still would have given response r. (Hitchcock, 2015, p. 308)

> The meaning of 'x-complete cause' is as follows. Starting with the idea of the complete cause of a manifestation event, 'x-complete cause' is obtained by restriction [and] makes the having of the disposition independent of circumstances that are extrinsic to x. (Molnar, 2003, p. 92)

These two interpretations are not equivalent. The first cannot be right, for we could introduce interfering properties that would mask the disposition. Suppose that a match is flammable, perhaps due to its bearing a certain chemical structure in the tip. Nonetheless, had one of the other properties differed (perhaps the colouring of the tip acts in such a way as to starve the local environment of oxygen) then the match would may well not manifest its flammability, despite the match counting as flammable.

The second is more plausible. I take it that it means something like: That subset of the complete cause that constitutes either the stimulus, or the relevant intrinsic properties. So if, for instance, a glass is struck, which results in it gaining some novel intrinsic property I^*, such that I^* enters into the cause of the relevant manifestation, then we would *not* have x-complete causation. In essence, being an x-complete cause eliminates dispositions gained between stimulus and manifestation.

With that in mind, we find that the account avoids a number of the counterexamples outlined above. The finkish cases are avoided, for they involve a change in the intrinsic properties of the object. If the wire is disposed to pass an electrical current to a conductor upon being so touched, then it need only conduct in cases in which it maintains its intrinsic properties, properties which are changed by the presence of the conductor. Similarly, if the wire is *not* disposed to do so, then it will not have intrinsic properties that satisfy the requirement. Given that the finks *change* the intrinsic properties, in the first case by giving the wire an intrinsic property, and in the latter by taking it away, the cases are not counterexamples to the causal analysis.

Mimic is more tricky. After all, the striking of the (not fragile) chalice and some intrinsic property would constitute an x-complete cause of the shattering. Lewis may, however, have a trick up his sleeve: he may argue that the similarity relevant to the evaluation should give less weight to features such as the angel's presence. After all, it may be thought, there is a perfectly respectable sense in which the glass could easily *not* shatter when struck (we simply need to look at a world that is similar striking-wise, but not angel-wise). And given that it could easily have failed to shatter, by equivalence, it is false that it would shatter if it were struck.

So far so good. But the pressure point for the causal analysis lies in its treatment of *Mask*. For in such a case, the glass is disposed to shatter when struck, despite it being the case that if it were struck, then it would not shatter, *even though it would retain its intrinsic properties*. So the striking, and one of the glass' intrinsic properties, would not be an x-complete cause of the shattering, as there would be no shattering *simpliciter*. Lewis is aware of the issue. In response, he suggests that the issue is resolved by proper stimulus individuation. He writes:

> We might offhand define a poison as a substance that is disposed to cause death if ingested. But that is rough: the specifications both of the response

> and of the stimulus stand in need of various corrections. To take just one of the latter corrections: we should really say 'if ingested without its antidote'. Yet the need for this correction to the analysis of 'poison' teaches no lesson about the analysis of dispositionality in general. (Lewis, 1997, p. 153)

This response is highly dissatisfying. It is certainly right to suppose that the stimulus conditions require more detailed specification: The glass will not manifest its fragility if struck by a hammer made of sponge. But exclusions of the kind Lewis has in mind make the analysis appear circular. After all, to grasp what kind of conditions must be excluded, one must already grasp the relevant notion of a disposition: to grasp the concept of antidote one must first grasp the concept of poison. An antidote is something that counteracts the effects of a poison; the concept cannot therefore be used as part of the analysis of the disposition to kill when ingested.

Worse still counterexamples remain, for there are cases in which the antidote does *not* appear to be part of the stimulus conditions. Bird (1997) gives the following case:

> *Antidote* Smith ingests a poison at t. Sometime later at t^*, before his untimely death, Smith is injected with an antidote, which prevents the poison's deadly effects.

In *Antidote*, the stimulus occurs, and *then* the masking. Why is this problematic? He writes:

> In this case, the disposition and its causal basis remain throughout. The object in question receives the appropriate stimulus, but does not give the expected response. (Bird, 1998, p. 228)

If this is correct, then Lewis' amendment to include the lack of antidote as part of the stimulus condition is ineffective, for the disposition can be masked *post-stimuli*. It follows, then, that Lewis' account, even so modified, cannot deal with all cases of masked dispositions.

Should the modal conditional analysis, then, be committed to the flames? Some say 'no'. A number of contemporary metaphysicians have opted for what Goodman called a 'fainthearted' conditional. A fainthearted conditional is one that requires that the consequent be satisfied only under certain conditions. We could, then, eliminate masks by requiring that the conditional holds only in cases in which these further conditions obtain. Where they do, the case would be what Fisher calls an 'auspicious circumstance'.[10] Notice that these will *not* be built into the stimulus conditions, *contra* Lewis, but will rather be supplementary features held fixed. Denote those further conditions by 'ζ'. We could formalise this thus:

[10](Fisher, 2013).

Fainthearted Analysis

> D$x_{(s,m)}$ at t iff. there exists some intrinsic property I, and time t^* later than t such that
>
> 1. $((\mathrm{S}x \text{ at } t \wedge \mathrm{I}x \text{ from } t \text{ to } t^*) \wedge \zeta) \mathrel{\Box\!\!\rightarrow} ((Sx \wedge \mathrm{I}x) \rightarrow^c \mathrm{M}x)$

An even more unlovely mouthful![11] The analysis now avoids *Antidote*, for we may say that the poison is disposed to kill when ingested, and when no antidote is present (rather than: disposed to kill when ingested without its antidote).

But there are serious issues with adopting the fainthearted analysis. The first is that the counterfactual condition becomes redundant. Now that both the consequent and antecedent have been hyper-specified in order to discount the problematic cases, it may be thought that the analysis should make use of some kind of strict implication. But more problematically, and irrespective of which modal conditional we choose, it is far from obvious how we are to specify the correct value of 'ζ', which results in a painfully nebulous analysis.[12] Mumford argues they are the 'ideal' conditions for the disposition, but then the worries raised against Lewis re-emerge. As Fara duly notes, what counts as ideal *'can be fully specified only by appeal to the notion of masking [...] [t]he ideal conditions for a disposition are surely those conditions most conducive to its being manifested'*.[13] Since masking is understood only with reference to the analysed disposition, disclusion of ideal conditions is viciously circular. The 'ideal' conditions for a poison will be those in which an antidote is not administered, but we can know this only because antidotes mask the dispositions of poisons.[14]

Another tempting approach would be to specify the conditions in terms of what 'normally', 'ordinarily' or 'typically' occurs. But masks can be perfectly normal or typical. In a world in which it was normal or typical for glasses to be bubble wrapped, or for antidotes to be administered (or vaccinations, for an actual case) fragile and poisonous objects would still bear their dispositions.[15] Others have argued that the

[11] It should be noted that there are a variety of ways to give fainthearted analyses, some of which do not involve the causal element. The view may also be formalised through the use of a novel conditional. See (Morreau, 1997) for an excellent example, and (Fara, 2005) for discussion.

[12] Cf. (Lewis, 1997, pp. 157-8), (Martin, 1994, pp. 4-5).

[13] (Fara, 2005, p. 52).

[14] For an argument against the charge of circularity, which is to my mind unsuccessful, see (Choi, 2008).

[15] A valiant attempt at specifying the conditions has been undertaken by Contessa (2013). The analysis involves an absence of 'interfering conditions', of which there are two kinds, both explicated in non-dispositional terms. I won't have too much to say about Contessa's analysis, other than it suffers from the following shortcomings. Firstly, it is *highly* complex. Secondly it requires that different treatment be given for what Contessa calls 'intrinsic' as opposed to 'extrinsic' dispositions. To my mind, the notion of an extrinsic disposition is highly contentious. But even granted that the dichotomy holds, the analysis is irredeemably disjunctive. In that respect, it lacks unity. It also commits Contessa to implausible consequences, namely that a match's flammability is an extrinsic disposition. For more see (Contessa, 2013, p. 415).

correct value is determined by the context of utterance.[16] But appeals to context sensitivity are generally hindered by lack of a plausible account of how exactly the context is determined. Whilst context sensitivity is an opaque phenomenon, we should expect *some* account of those features that set the context; there appears to be *nothing at all* that we can use to determine which features are relevant, for the counterexamples that must be ruled out *are so numerous and rich in variance* that they appear to resist unification, at least without reference to the analysed dispositional concept. If such an account cannot be provided, then exclusion of problematic cases are irredeemably *ad hoc*. If not normality, typicality, ideality, which features are salient? Without a plausible answer, appeals to context sensitivity should be revoked.

More worryingly, *even if* such circumstances obtain, the objection retains its force, for it is not only environmental features that would have to be set as part of the context, as which circumstances must be excluded is also determined by the object that bears the disposition. This point is well made by Hardin, when writing on colour concepts.[17] He notes that what counts as standard conditions for a given colour concept is often not *colour* relative, but *object* relative. What I mean by this is as follows: The normal conditions are not simply determined by the object's colour, but partly by the kind of object it is. To take his own examples, rainbows and bioluminescent fish simply cannot be viewed under the 'standard conditions' relevant for Munsell chips. For an analogous case, and to appease those who are anti-dispositionalists about colour concepts, consider an ordinary kitchen match. What counts as a 'normal circumstance' for the kitchen match would most likely include sufficiently dry conditions, but would not be included when considering a waterproof match. That is to say; the conditions of masking are not *just* determined by the context, but *also* by the object that bears the disposition. It seems as though one can truly say of two matches, one waterproof and one not, that they have the disposition to ignite when drawn across the striker, and such ascriptions are not—not obviously at least—ambiguous. We can attribute *the same disposition* in *the same context* to *two different objects*, even though the auspicious circumstances that must obtain for each object differs. So it seems as though 'ζ' cannot be individuated in terms of context alone.

A brief recap is due. So far, I have distinguished between conventional and canonical dispositions, and have outlined three analyses of the latter, the simple conditional analysis, the causal analysis, and the fainthearted analysis. All three were rejected. That deals with

> [1.2] To outline and reject three analyses of dispositions: the 'simple' analysis, the 'causal' analysis, and the 'fainthearted' analysis.

[16] See, for example, (Fisher, 2013), (Hawthorne & Manley, 2005), (Manley & Wasserman, 2011), and (Bird, 1998).

[17] See (Hardin, 1988, pp. 68-69).

1.2.2 The causal contingent account

In what remains, I would like to sketch an alternative. I am in sympathy with Bird when he writes that the conditional element is a '*red herring*'.[18] That is not to deny that there is an important connection between conditionals and dispositions, but only to deny that the connection is one of coextension. The account is based upon one recently given by Barbara Vetter (2014). However, whilst I take the spirit of her account to be correct, it is less plausibly motivated than one might like, and benefits from some modifications. In what remains of this section, I outline Vetter's account, suggest some modifications, and then support the modified version.

Vetter's account is as follows:

<u>Contingent Analysis</u>

$\mathrm{D}x_{(m)}$ iff.

1. $\Diamond\, \mathrm{M}x$

There are a few points to note. The first is that the disposition is now formalised as a two place relation between manifestation and object. This is because, on her view, dispositions are individuated solely in terms of manifestations. Why so? Firstly, she is sceptical on linguistic grounds of the notion of canonical dispositions, and thus restricts her attention to conventionals. Secondly, she suggests that the stimulus conditions for a given disposition are too disparate to be satisfactorily unified. She writes:

> Fragile parchments break upon being merely touched, and a fragile old wooden chair may split when transferred into a different temperature. The various conditions that cause fragile objects to break have little in common, apart from their being non-extreme causes of object's breaking. (Vetter, 2014, p. 132)

It is not obvious we should accept this. Two objections can be raised. Firstly, an objection she considers, is that it fails to accommodate the fact that different dispositions may result in the same manifestation. Her own examples are claustrophobia (the disposition to feel anxiety in response to being in an enclosed space) and acrophobia (the disposition to feel anxiety in response to heights). Both have anxiety as their manifestation; they differ only in their stimuli. But if dispositions are individuated in terms of their manifestations alone, then these ought to count as the same disposition. But they are not. So dispositions must, in some cases at least, be individuated by more than their manifestations alone.

Her response is dissatisfying. She writes:

[18](Bird, 1998, p. 233)

> [T]he manifestation of a disposition [...] may or may not be complex [...] it may be the property of *being caused by flowers to sneeze*, of *being caused by an enclosed room to feel anxiety* or of *being caused by height to feel anxiety*. (Vetter, 2014, p. 149) (Italics in original)

This move does not work, for she avoids individuating dispositions in terms of stimulus conditions by simply building in the stimulus conditions into the manifestations. The manifestation/stimulus distinction is supposed to distinguish *what happens to an object* that results *given some event's occurring which has an effect on the object*. By allowing these features as part of the manifestations, Vetter just changes what she means by 'manifestation' to 'stimulus-manifestation'. This fails to show that dispositions are individuated in terms of their manifestations alone.

Vetter may reply that whilst the account does sometimes include something akin to 'stimuli' as part of the manifestations, that it is nonetheless silent *vis-á-vis* the specifics of the relation between stimulus and manifestation. But that can't be right. Consider again the disposition to be caused by flowers to sneeze. Suppose that Smith suffers from a rare condition, such that he is disposed to sneeze upon undergoing a perceptual experience of flowers. In such a case, Smith is disposed to be 'caused by flowers to sneeze', but the flowers are *not* the cause of the kind of the dispositional manifestation she has in mind: Namely *being allergic to flowers*. The natural solution is, of course, to build into the clause that it must be caused by *the pollen* of flowers to sneeze. The problem with pursuing this line, however, is that we now have what looks suspiciously like the relevant specifics of the causal relation between manifestation and stimuli. So in the end, 'manifestation' ends up being structurally analogous to 'stimulus-manifestation'.

The second problem is that exactly the same applies to dispositional *manifestations*. Fragility can be manifested, as noted at the start, by cracking, breaking, shattering, splintering, and so on. So *even if* her claim holds, we have just as much right to claim that dispositions are individuated only in terms of their stimulus conditions. In either case, her defence fails. We should, then, maintain the tradition of individuating dispositions in terms of stimulus conditions and manifestations, and conventional dispositions in terms of complex sets of canonicals where required.

The second point to note is that the analysis does not involve a conditional, but requires only that the manifestation *could* occur. Some might take this to be over permissive: Almost anything *could* break when struck, but it is not the case that almost anything is fragile. To accommodate these worries, Vetter classifies dispositions into two types. The first involves bare possibility and include flammability, irascibility, breakability, and so on. Something is flammable just in case it could set on fire when ignited, something is breakable just in case it could break when struck. The second type, which include the problematic counterexamples, does not involve what merely

could occur but rather what *could easily* occur. They are, in that sense, gradational dispositions. Something is fragile just in case it *could easily* break, something is poisonous if it *could easily* kill.

To my mind, there is a genuine dichotomy here, but not one of gradationality. Flammability is gradational; logs are less flammable than kindling. So too for breakability; champagne flutes are both more fragile *and* more breakable than steel flasks. Whilst it is true that we often have certain strengths of disposition in mind when we make dispositional ascriptions, this teaches us nothing about the meaning of dispositional concepts, for similar points apply to predicates of all kinds. When I ask for a red jumper, a jumper that is *just barely* red may not be what I had in mind. When I ask for an alcoholic drink, I would generally not be asking for a drink with an alcohol content of less than 1%. What distinguishes the kinds, in contrast, is that the former class are individuated by a single manifestation (something is breakable just in case it could *break*), whereas the latter are not (fragility may manifest through shattering/cracking/splintering and so on). Of course, there are many ways of breaking, but that is not the point. A term like breakable or flammable specifies a particular manifestation; terms like fragility do not.

That is not to deny that some dispositional ascriptions bear a built-in stability requirement. Sometimes, the term 'flammable' is contrasted with 'combustible', in that the latter applies to entities that could catch on fire, whilst the former applies to entities that could easily catch fire. Usage differs; even dictionaries offer distinct definitions. Lexicography is an imperfect art. But even if some dispositional terms do bear a stability aspect, that is not to say that the stability is part of the meaning of dispositional concepts. Rather, we may say that such terms are *partly* dispositional, *partly* gradational. In that sense, they are analogous to other non-dispositional terms. For instance, the term 'huge' may be thought of as expressing a complex concept: the combination of 'big' and 'very'. Similarly, to be drenched may be thought of as the complex of 'wet' and 'very'. Stability is not a unique feature of dispositional expressions: Rather, many terms of English express blends of concepts, some of which are gradational, some of which are not. In short: Gradationally is no distinctive feature of dispositional expressions.

I think that Vetter is correct that we should analyse dispositions as contingencies, rather than as conditionals. But her defence of the contingent analysis is primarily linguistic; she argues that it best respects the way we talk about dispositions. To my mind, we should be suspicious of conceptual claims made on the basis of linguistic evidence. Moreover, I part ways from her in two important respects. Firstly, I take it that the analysis should incorporate stimuli and manifestations—she has given us no reason to suppose that it should not. Secondly, and more importantly, we should re-introduce Armstrong's insight: *Dispositions are causal concepts*. I suggest, then, that

we construe the relation of stimulus to manifestation in causal terms and let the truth-functional work be done by the modal operator. Thus, on the present account, for an object to possess the disposition to break when struck is simply for a certain *causal modal contingency* to hold; the possibility to be caused to break upon being struck. In formal terms:

Contingent Causal Analysis

$\mathrm{D}x_{(s,m)}$ iff.

1. $\Diamond\, \mathrm{S}x \rightarrow^{c} \mathrm{M}x$

The contingency at issue should most likely be construed nomologically; dispositions are possible causal relations given the actual laws of nature.[19] We should also hold fixed the object's intrinsic properties at the start of the evaluation. Like before, a more precise formulation would stipulate that the object undergoes no change in intrinsic properties other than those caused by the relevant stimulus. Note that these features are not unique to dispositional ascriptions, but are features of modal ascriptions more generally construed. When I say that Jones can run a five minute mile, I generally mean that he can given the laws of nature, and given the properties he currently bears. It may be that Jones can run a two minute mile if gravity were weaker; that would not make an utterance of 'Jones can run a two minute mile' true. Similarly, it may be that Jones could run a two minute mile after extensive surgery to increase his speed, but such a world would not usually be counted as accessible on the relevant modal evaluation.

In what remains, I offer an alternative non-linguistic defence of the reformulated account. I then consider and reject some objections.

First, the account best respects the purpose of dispositional ascriptions. The basic thought is that we characterise objects dispositionally in order to provide information about how the object may be caused to change or to bring about changes in other objects, by certain stimuli. We do so in order to modify our behaviour in such a way as to better satisfy our desires. In that sense, dispositional ascriptions play a central role in practical rationality. And this is all, I contend, best explained by their being possible causal relations.

Here is an example. Suppose I tell Jones that the glass is fragile. Why would I do this? What would the *purpose* of my doing so be? Well, I contend, I am telling Jones that the glass could be caused to smash by being dropped. It follows, given Jones' desire to keep the glass intact, that he should be careful not to drop the glass. Thus, by pointing to the causal modal properties of the glass, Jones may modify his behaviour in order to better satisfy his desires. And once his behaviour has been so-modified, my purpose

[19] Cf. (Mumford, 1998, p. ch. 10) and (Armstrong, 1968, p. xvi).

in ascribing the dispositions has been satisfied. Similar points apply to dispositional ascriptions in the sciences. The discovery of possible causal relations is part and parcel of scientific inquiry. Its purpose is, often enough, to guide future behaviour, whether that be the design of new artefacts or of new methods of empirical verification.

Second: The account is theoretically virtuous, in the following three ways:

> *1. Simplicity* The analysis is simpler, or at least as simple as the simple conditional analysis. There are no 'unlovely mouthfuls'. Of course, the subject of simplicity is vexed. But those sympathetic to simplicity as a theoretical virtue should be sympathetic to the account.[20]
>
> *2. No counterexamples* The account is resistant to *all of the standard counterexamples to the simple conditional analysis*. Let there be finks, masks, mimics, antidotes—populate one's bestiary! The examples are inert in regards to the contingent causal analysis.
>
> *3. No ad hocery* The counterexamples are avoided without *ad hoc* stipulations that certain cases be removed, and without vague appeals to 'normal contexts'.

I contend the reasons above make the contingent analysis preferable to the other analyses. It is able to avoid the relevant counterexamples simply and without *ad hoc* explanations of why certain cases aren't to be included, and without appeal to unspecified contexts.

I will now consider some objections. The first I anticipate is as follows: The analysis is *wildly* over permissive. There are two ways to argue for this. They are as follows. The first would be to consider *Mimic*. The angel has decided to make the chalice shatter if struck. The chalice is struck, the striking causes the angel to shatter the chalice, so there is a possible world in which the chalice is struck, and the striking causes the shattering. So the *analysans* is true, but *ex hypothesi* the *analysandum* false. Too much is permitted.

In reply, we should note that the notion of 'cause' in the relation '$\rightarrow^c$' cannot be *any* kind of causation whatsoever. Rather, it should invoke what we may call *direct* causation. In *Mimic*, there is a weak, transitive sense in which the shattering is caused by the striking, but really *the direct* cause of the shattering is not the striking, but the angel. What do I mean by the 'direct' cause? An event e is the *direct* cause of some other event e^* just in case e causes e^* in a more liberal sense, and e^* is spatiotemporally contiguous with e. If I strike a glass and it thereby shatters, the striking is contiguous with the shattering, but in *Mimic* it is not, for the shattering is contiguous not with the

[20] I will discuss simplicity, alongside other theoretical virtues, in more detail in Chapter 2.

striking, but with that long complex event the Angel ensures occurs, which results in the shattering of the chalice.

But the amendment faces resistance. Consider, for example, the following passage from Lewis:

> A certain virus is disposed to cause those who become infected with it to end up dead before their time, but not to undergo the direct and standard process whereby lethal viruses mostly kill their victims. For this virus does not itself interfere with any of the processes that constitute life. Rather, it interferes with the victim's defences against other pathogens – whereupon those other pathogens [...] do the dirty work. Do we call this a lethal virus? Of course we do. After all, my story of the virus is not just another philosophical fantasy! It is the true story of HIV, slightly simplified. We should not think, therefore, that dispositional concepts generally have built-in response-specifications requiring a direct and standard process. (Lewis, 1973, p. 154)

The basic thought, then, is that HIV can be disposed to kill upon infection despite the virus causing death by setting off of a long complex causal chain, and one that can take many forms at that. So for something to bear a dispositional property does *not* require that there is some direct causal process involved that holds between the stimulus condition and the manifestation. I doubt, however, that the case is as problematic as Lewis makes out. True enough, we sometimes say that HIV is deadly. But if we were to be precise we would not say so. Consider the following claim found in information about HIV infection:

> HIV stands for human immunodeficiency virus and on its own it does not kill you. The virus can survive and grow only by infecting, and destroying, the immune system. This continual assault on the immune system makes it weaker and weaker until it is no longer able to fight off infections. Without treatment, it takes about 10 years from infection to the development of AIDS - acquired immune deficiency syndrome. It is then that "opportunistic infections", ones a healthy immune system could fight off, become deadly.[21]

Notice that the following claims are made in this passage: (1) HIV does not kill you, and (2) it is the infections which normally do not kill that 'do the dirty work' by becoming deadly. I suggest this is correct—upon learning that HIV works as it does, the natural temptation is to withdraw the dispositional ascription. In fact, the case appears to be a one of unmasking. Our immune system usually masks the dispositions

[21] http://www.bbc.co.uk/news/health-15853743

of bacteria and viruses to kill upon infection; the HIV removes the mask. HIV, then, is *not* a deadly virus; but rather is a virus that allows bacteria and other viruses to manifest their deadly dispositions.

Nonetheless, further cases bring further worries. One such case would involve *bona fide* poisons; cyanide starves the mitochondria of the oxygen they require to produce energy. Deadly though cyanide may be, its mortal manifestations are highly indirect. One response would be to invoke levels of description. The death of an individual poisoned with cyanide, it may be thought, is a long, complex event of which the purported events that make the deaths seem indirect are mere parts. But further cases may be constructed at the same level of description. An alarm clock has the disposition to ring once the minute hand passes a certain threshold, but the mechanism which runs from the minute hand to the bell striker may be highly indirect, and *certainly* cannot be included as part of the manifestation, i.e., the ringing.

Fortunately, proponents of direct causation need not fear. For the purported indirectness occurs in all the purported counterexamples *as part of the manifestation*. The manifestation may be a highly indirect process—it does not follow that the stimulus does not directly cause the manifestation. When I say that a glass is disposed to shatter if struck, I mean to say that the *striking* could have some effect on the glass's properties which result in its shattering. So, when some x is disposed to M given S, it must be possible that S is a direct cause of M. With that amendment in place, the account is saved from the charge. *Mimic* is now avoided, for it is a case of indirect causation between stimulus and manifestation.[22]

One may still worry. Is the notion of a direct cause coherent? My reply: of course it is! The direct cause just is the event that caused the effect. What's problematic about *that*? Compare Mellor:

> We can identify events, in this concrete occurrence sense, simply by their spatio-temporal regions, well enough to allow one event but not another roughly contemporaneous event to be the cause of a certain effect. (Mackie, 1974, p. 257)

Like similarity, causal directness is a notion we are adept at employing. There is nothing illegitimate or 'spooky' about it.

A second way to argue that the account is over permissive would be to consider objects that would be caused by the stimulus to manifest only in abnormal environments.

[22]It is worth briefly noting that Lewis appears to have missed that the case he gives—HIV—is *not* counted as deadly on his own analysis. For HIV does *not* (in fact: perhaps *cannot*) act as an x-complete cause of death. Rather, something extrinsic (a novel infection) must also occur. That is to say: there are no properties of an individual, such that those properties and HIV are a sufficient cause of the individual's death. So, even by his own lights, HIV should not be counted as deadly, just as we should expect.

A ball of steel wool is not flammable, but when placed in a sufficiently oxygen-rich environment, the ball would set on fire when ignited. So something can *possibly* be caused, given the relevant conditions, to manifest, even if it does not have the relevant disposition. My response here is to bite the bullet. Intuitively, to me at least, the discovery that steel wool can set on fire when ignited in certain conditions shows that it *does* have the relevant disposition. Steel wool is flammable, not very flammable, but flammable nonetheless. True enough, we would typically not cite steel wool as a flammable item when listing the items in one's house that are flammable. But again, this is merely pragmatics, when we ask what is flammable, we are interested not in what is flammable *tout court*, but what is flammable if, say, a house fire were to break out. So steel wool is no counterexample, but rather *supports* the contingent analysis. The discovery that something *could* be caused to manifest given the relevant stimuli comes with the discovery that the object bears a disposition it was previously thought to lack.

That deals with [3.1]. The causal contingent account has been sketched and defended. Now, before we finish, we should ask: What of the relationship between dispositions and conditionals? We saw at the start that the conditional analysis arose, in part, due to the fact that dispositions are often explained with reference to conditionals. If no conditional analysis is possible, why would that be?

We can, I think, explain the temptation away, for at least two reasons. Firstly: Causal relations and conditionals bear close residence. When possible causal relations hold, in the absence of cases that interfere with the causal process, the effects of the cause will typically follow. Put another way: causal relations are typically counterfactual-supporting. Not always, though. The causal process may be interfered with (think masks), or deviantly duplicated (think mimics). So we can, admittedly somewhat heuristically, make use of conditionals when ascribing causal relations. Secondly: It is commonplace for contingencies to be explained with reference to conditionals. Suppose, for example, that Smith claims that it is possible to dip one's hand in molten lead, without any pain or tissue damage. A contentious claim indeed! Jones, suspicious as he is, asks how. Smith says that *if* one dips one's hand in water prior to dipping one's hand in the molten lead, *then* the water will vaporise upon contact with the heated metal, which will result in a protective layer of steam that enables one's hand to avoid contact with the lead. In this case, a possibility claim is explained with the use of a conditional. This is no surprise, one can show that something is possible by showing that it is entailed by possible antecedent events. I contend, then, that this is why there is such a strong link between conditionals and dispositions. It is not because dispositions are *analysed* in terms of conditionals, but rather because a dispositional ascription, like ascriptions of contingencies of other varieties, can be explained and supported through the use of conditionals.

Conclusion

This chapter had the following aims:

[1.1] To outline the distinction between *canonical* and *conventional* dispositional expressions.

[1.2] To outline and reject three analyses of dispositions: the 'simple' analysis, the 'causal' analysis, and the 'fainthearted' analysis.

[1.3] To sketch and defend the 'causal contingent' account.

Those aims are now satisfied. Now, recall that at the start of this chapter, we saw that dispositions were problematic for the empiricists. If what was argued in this chapter is correct, then their problematic nature derives from the fact that dispositions must be analysed in modal – though non-conditional – terms. In the next chapter, I am going to take the argument one step further: I will argue not just that dispositions may be *analysed* in terms of causal contingencies: They may be *identified with them*.

Chapter 2

The Identity Theory

Introduction

Once upon a time there was a theory. Its name was: *The Identity Theory*. In its embryonic stages, the identity theory was surrounded by hope and praise. It promised much: A solution to the mind-body problem, the problem of psychophysical overdetermination — philosophical spoils! Alas, as it grew, hope was lost. It faced a problem. The identity theory fell out of philosophical fashion.

This short story tells a piece of philosophical folklore: The rise and fall of the identification of body and mind. Now, whilst many suppose the identity theory to be, by now, a dud hypothesis, not all do. *Some* have thought that its central problem is avoidable. What is the problem of which I speak? Answer: The problem of *multiple realisability*. Now, some have denied the problem to be fatal for the identity theory. All that it reveals — so the story goes — is that we must endorse a *token-token* as opposed to a *type-type* identity theory. Once that amendment is made, the spoils are ripe for the picking!

'Hold on', you say, 'this part is supposed to be on *dispositions*, but the identity theory is a position in the philosophy of mind.'

Well true, but the identification of mind and brain *ran alongside* the identification of dispositions and their causal bases. For it was thought that states of mind *are* dispositions, and it is *precisely because* dispositions are identical to their causal bases that mental states are identical to their causal bases — states of the central nervous system'.

Thus, in this chapter, I will have little to say on the identity theory as a position in the philosophy of mind, though implications may naturally be drawn. My primary focus in this chapter will be on the move outlined above, from a type-type identity theory to a token-token identity theory. I will argue the move is dialectically inert.

I have three central aims. The first is to outline the identity theory. The second

is to reject it. I will argue that the problem of multiple realisability holds not only at the type-level, as is well known, but at the token-level also. The third is to support a *modal* account of dispositions: I will argue that dispositions are identical to those causal contingent properties they were analysed in terms of in the previous chapter.

Here are the aims, explicitly put:

[2.1] To outline the identity theory.

[2.2] To argue that the problem of multiple realisability holds at both the type and the token level.

[2.3] To defend the view that dispositions are modal properties.

Here is the plan. There are two sections, both composed of two subsections. In the first, I outline the identity theory, which states that dispositions are identical to their causal bases. I then outline a notorious problem such identifications must face: the problem of multiple realisability. I then outline a reformulation of the view intended to overcome the worry, which takes the identification to hold not at the *type* level, but at the *token* level. In the second section, I argue that the reformulation fails. I show that there exists a class of the multiply realised dispositions which are as threatening to the token-token identity theory as standard cases of multiple realisation are to the type-type identity theory. I call these *plurally realised* dispositions. I then argue that dispositions are modal properties, defending the view from several objections, two of which are given by Mellor (1974). Finally, I offer the view some support.

2.1 Two Identity Theses

2.1.1 Dispositions as causal bases

According to

> *The Identity Theory* Dispositional properties are identical to their causal bases.

What are causal bases? As noted at the start, the causal basis of a given disposition is simply that intrinsic property, or that conjunction of intrinsic properties if there are many, that are causally efficacious in the disposition's manifestations. For instance, the fragility of a glass is identical to, say, those structural properties of the glass which are responsible for its shattering once struck. Those non-dispositional properties are typically called *categorical* properties.[1] I will follow suit.

[1] There is, in fact, a real trouble with the categorical/dispositional distinction, *especially* given attempts

Major proponents include Quine (1960), Armstrong (1968, 1973), Mumford (2003), Dennett (1989), and Heil (2004). Before I reject the thesis, it is worth outlining its merits. Why should we accept it? What benefits does the identity theory yield?

Merit 1 The identity theory provides a neat identification of the mind and the body. The main thrust of Armstrong's *A Materialist Theory of Mind* essentially rests on the identification of dispositions with their causal bases. The basic argument runs as follows:

Australasian Identity Argument

1. Mental states are dispositions.
2. Dispositions are identical to their causal bases.
3. The causal bases of mental states are bodily states.

therefore

C. Mental states are identical to bodily states.

Nevertheless, as I will try to show, valid though it may be the argument is unsound, for the second premise is false.

Merit 2 The second reason in its favour concerns the status of scientific concepts. Recall that at the start of the last chapter, we saw that dispositions are entities that 'lie latent'. The latency of dispositions makes us, as Goodman once wrote, *'moved to inquire whether we can bring them down to earth; whether, that is, we can explain disposition terms without reference to occult powers'*.[2] But this is no simple task: The latency of dispositions makes them appear to fall outside the scope of observable reality, at least whilst they persist unmanifested. This is especially clear in the case of scientific concepts, many of which are taken to be dispositional. Dispositions appear occult, mysterious; such mystery has no place in naturalistic inquiry.

The natural response to their latency is to take dispositions to be *possibilia*. But *possibilia*, many have maintained, are no less occult than dispositions themselves. To subsume dispositions under the merely possible is to fail to bring them down to earth. To identify the dispositional with the modal merely trades the occult for the occult. If the sciences rest only on *mere possibilia*, at least so it was maintained by empiricists of the day, then the sciences appear *less scientific*. If what follows is correct, then Goodman's inquiry is stillborn. Dispositions are modal properties; they cannot be brought down to earth, unless *possibilia* may be brought with them. It should be noted that to the naturalistically minded of contemporary philosophy, scientific concepts are essentially

at identifying the two. For, strictly speaking, 'categorical' in this context means 'non-dispositional', and evidently dispositions cannot be identical to non-dispositional properties. Despite the inadequacies of this terminology, it will be worthwhile to adopt it. For interesting discussion on this issue, see (Crane, 1996).

[2](Goodman, 1954, p. 40).

modal will be less worrying than it was to the likes of Quine and Goodman. Modality has been legitimised; with it, suspicions have been quelled.[3]

In summary, the identity theory (a) gives us a neat identification of mind and body and (b) de-modalises dispositional concepts. The thesis has pleasing consequences. Are there independent reasons in its favour? According to a number of philosophers, the answer is 'yes'. I will consider two.

The first concerns its explanatory virtues. In particular, the view is *highly* parsimonious, and parsimonious in the 'kind' sense at that.[4] Our best total ontology need not include both dispositional properties and non-dispositional properties. We must grant only the latter ontological status. The second is semantic. Suppose that Jones utters 'that glass is fragile'. According to Mumford, Armstrong, and Quine, what Jones means by this is something akin to: 'the glass has some categorical property C, such that C would cause the glass to exhibit certain manifestations given certain stimuli'. Importantly, the latter part is but a description of the dispositional property—the property is identical with C. This also appears to be given tacit endorsement by Lewis for, as we saw in the last chapter, his analysis of dispositional concepts included the existence of some causally efficacious intrinsic property as part of its *analysans*.

The semantic reason is not self evidently true. Why should we accept it? In later work, Armstrong writes:

> [I]t is linguistically proper to *identify* the disposition with this state of the disposed object. It is linguistically proper, for instance, to say that brittleness *is* a certain sort of bonding of the molecules of the brittle object. The ground for saying this is simply that scientists and others often speak in this way, and there seems to be no objection to such speech. (Armstrong, 1973, p. 14)

The passage gives two reasons. They are as follows: (1) scientists (and others) often speak as though the identity theory is true, (2) there is no objection to such speech.

Now, (1) is true, but we don't *always* speak this way. After all, we sometimes do not say that brittleness of a glass is *identical* to the bonding, but rather that the bonding is the *reason why* the glass is brittle. But even if this were false, that scientists and others often speak a certain way is not strong evidence for its truth. Scientists (and others) are notoriously liberal in their attribution of identity relations. Plausibly, such liberality is objectionable on philosophical grounds. We might, then, reasonably object to such speech, *contra* (2). For if someone states 'The fragility is the bonding', we might reasonably reply: 'That is not obvious, but certainly the bonding is the *reason for* its fragility'.

[3]Cf. (Williamson, 2016).
[4]Cf. (Lewis, 1986b).

More seriously, it is far from clear that such speech is guided by purely semantic motivations. For, it may be thought, it was very much unknown whether science would discover the bases of dispositional properties. This was no *conceptual* truth, at least. It may be thought that a better explanation of our tendency to *suppose* that we are characterising objects so, is our semantic grasp of dispositional concepts *combined with* our knowledge of scientific discoveries. Now that we know science seems to find the causal correlates of dispositional manifestations, we may infer that such correlates exist, whenever a disposition is ascribed. We need not commit to that being part of the *meaning* of dispositional ascriptions, we may take it to be an inferential habit, informed by knowledge gained *a posteriori*. So Jones may not *mean*, merely by making a dispositional ascription, that there exists some categorical property which the relevant disposition is identical with, though it may be amongst the beliefs that drive his utterances.

Now, the considerations above are, at best, weak support for the identity theory. Linguistic considerations and mere parsimony alone are insufficient for outright acceptance. Are there other reasons that we should accept the view? According to Stephen Mumford (2003), there are. For, so he contends, there is a strong argument in its favour. In essence, the argument is a generalisation of Lewis' (1966) argument, later developed by Peacocke (1979), for the physchophysical identity theory.

It runs as follows:

The Argument from Causal Role

1. For any dispositional property D, D is the occupier of some causal role R.
2. For any such casual role R, R is occupied by some categorical property C.
3. If x occupies the causal role R, and y occupies R, then $x = y$.

therefore

C. For any dispositional property D, D is identical to some categorical property C.[5]

The argument is valid, but the premises are contentious. Why should we accept the first? Mumford's support essentially rests on the semantic reason offered above. We found that wanting. He does give the following brief further support:

> How could the opponent [...] be persuaded? One line of argument is to point out the conceptual absurdity of dispositions occupying inappropriate

[5]Talk of properties occupying causal roles is somewhat misleading. Strictly speaking, we should we speaking of, as Shoemaker puts it, the causal contribution to the object the property provides. See (Shoemaker, 1980).

> causal roles. If it was claimed that something dissolved in water because of its *fragility*, then, unless some explanation could be produced, this claim would appear not just false but also absurd [...] because of the conceptual connection between the causal role of causing dissolving when in water and the causal role of causing breakage when dropped. The idea of a disposition occupying a different causal role to the one it actually occupies involves a conceptual, rather than a factual, confusion. (Mumford, 1998, p. 149)

But this gives us no reason to suppose that dispositions occupy causal roles. After all, we can explain the absurdity, and agree that it arises from conceptual confusion, without thereby accepting the premise. For suppose that the modal analysis defended in the previous chapter is correct. We can say that proper grasp of the concept of fragility involves grasping that something's being fragile is a matter of it being possibly caused to break if struck. It would then be a conceptual confusion to claim that something dissolved because it is fragile; but this does not require that it occupies any causal role whatsoever.[6]

What, then, about the second premise? Mumford's support is that:

> There has been a long history of categorical properties or mechanisms being offered as explanations of disposition manifestations. Their explanatory value lies in the empirical or theoretical claims that they occupy, as a matter of fact, those same causal roles which are 'pre-scientifically' understood only in terms of dispositions. (Mumford, 1998, p. 149)

This certainly seems right. We now understand that an object is combustible, for instance, just in case it contains a substance capable of undergoing rapid oxidation. It is, of course, an empirical matter whether this will hold for *all* dispositional properties, but this does appear well supported by the sciences. I am prepared, then, to grant that the second premise is true. So too for the third. The premise states that where two properties occupy one and the same causal role, those two properties are identical. This is but the standard account of property identity. If water is identical in its causal role to H_2O, then the two properties are identical. The *concepts* may be distinct, even if the properties are not.[7]

A notorious hurdle that the argument must face is the well-known problem of *multiple realisability*, famously raised by Prior *et al*. We can illuminate the problem with the use of a *reductio*. First, we suppose that the identity theory is true. Then, we take two objects that possess the same disposition: Perhaps an elastic band, and an elastic metal rod. By the identity theory, elasticity is identical to the causal basis of the rubber

[6] For more discussion, see (Squires 1968, 1970), (Stevenson, 1969), (Armstrong, 1969), and (Kim, 1993).
[7] Cf. (Shoemaker, 1980) and (Alston, 1971).

band. So too for the causal basis of the metal rod. Now, the causal basis of elasticity in the rubber band is its possession of polymer chains, but in the rod the basis is atomic lattices. So elasticity is identical to (a) polymer chains and (b) atomic lattices. By the transitivity of identity it follows that polymer chains are identical to atomic lattices. But polymer chains are non-identical to atomic lattices. The absurdity is now revealed: We have generated a contradiction.

The argument is powerful, though some have wondered whether it succeeds. Mumford, for example, questions whether or not the contradiction genuinely arises. He gives two arguments. The first is as follows. It may be that dispositions are *determinables*, whilst categorical bases are *determinates*. If that is true, dispositional properties would be analogous to the property of being red. A deep red tunic and a light red Ferrari may both fall under the extension of 'red', despite the categorical properties that are responsible for giving rise to the redness differing. Just as there is no contradiction in two objects being red, despite arising from two distinct categorical properties, there is no contradiction with one and the same disposition arising from two distinct categorical properties.[8] This response is rather puzzling, however. After all, it seems perfectly respectable to suppose that the relation between determinates and determinables, at least insofar as there are multiple determinates, is not one of identity, for exactly the transitivity worries given above. We do not want to identify red with any particular categorical property, arguably, for precisely the presence of non-identical determinates.

The second argument is that the purported multiple realisability arises due to the inclusion of 'causally irrelevant properties in the identity statement'. The basic thought is that we should not take the causal bases to be, in the instance given above, the atomic lattice and the polymer chains, but rather some property that they both share which is responsible for the disposition to deform reversibly under stress. As he notes:

> Once this is done, we find in each case that just a single categorical correlate, no matter how general, can be found for each disposition. (Mumford, 1998, p. 431)

Whether this strategy works, of course, must also be settled empirically. But it should be noted that the kind of properties which *would* unify the seemingly disparate properties, on the surface at least, appear to resist the generality Mumford requires. For often, it appears that we must appeal to properties that are suspiciously dispositional to unify the causal bases, which would yield unwanted circularity. What is it that all fragile objects have, in virtue of which they count as fragile, other than the disposition to shatter if struck? Some would argue that they are unified in terms of their all sharing a *functional* role. But functional-role realisation does not obviously denote what may

[8]See (Mumford, 1998, p. 158).

be properly called a 'categorical property'. We cannot rule out such identifications *a priori*, but in the absence of empirical verification, we should not accept they hold. Without the required empirical support, to rest one's case on this alone seems like an instance of *ad hoc* wishful thinking.

Mumford is aware that the arguments are seriously wanting. He claims that '*[T]his line is a difficult one to follow and it need not be pursued if variable realisation and a form of property monism can be reconciled.*'

2.1.2 Reformulation: token identity

In order to avoid the problem, Mumford follows Armstrong in modifying the kind of identity required. The basic thought is that the argument relies on what is called a 'type-type' identification between dispositions and categorical bases. Instead, they opt for a 'token-token' identification.[9] Consider:

> But I did miss something important, though. If the mental is nothing but that which plays a certain *causal role* [...] then there is the possibility, which may even be an *empirical* possibility that the causal role of tokens of the same mental type should be filled by tokens of significantly different physical types. Instead of type-type identity, one might have no more than a mental type correlated with an indefinite disjunction of physical types. [...] Every mental token is a purely physical token. (Armstrong, 1968, p. xv)

> The monist wants to say that there is just one attribute of x, or state that x is in, that makes it true of x that Dx and that Cx. This requirement can be satisfied even if the extensions of D and C do not coincide. Thus there need not be an identity of universals for monism. [...] each instance of a disposition is identical to some instance of a categorical base [this] amounts to a token-token identity theory. (Mumford, 1998, p. 159)

Talk of 'types' and 'tokens' is notoriously obfuscating.[10] But the intended distinction is as follows. As formulated, the argument requires that for any instance of a dispositional property, there is some unique type of categorical state, the tokens of which are responsible for the instance's manifestations. What cases of multiple realisability show is that, for many dispositions at least, there *is* no such type of categorical state. Different objects may bear the same disposition, despite the manifestations being caused by categorical tokens of distinct types.

[9] Cf. (Davidson, 1970).
[10] Cf. (Steward, 1997).

In contrast, a token-token identity theory makes no such demand. All that is required is that *each instance* of a dispositional property is identical to *some instance* of a categorical property. The rubber band's elasticity is identical to its polymer chains, the metal rod's elasticity is identical to its atomic lattice. But there is no requirement that the atomic lattice is identical to the polymer chains. We can, then, re-run the argument as follows:

The Argument from Token Causal Role

1. For any token *d* of a dispositional property D, *d* is the occupier of some token causal role R.
2. For any such token casual role R, R is occupied by a token *c* of some categorical property *C*.
3. If x occupies the causal role R, and y occupies R, then x = y.

therefore

C. For any token *d* of a dispositional property D, *d* is identical to the token *c* of some categorical property *C*.

Because the identification is at the token, and not at the type level, the *reductio* may no longer be run; our contradiction is avoided. At least, that is how the argument goes.

Not all are convinced that token-token is the way to go. Instead, the relation between categorical causal bases and dispositions may be thought to be one of *realisation*. Dispositional properties are said to be 'higher order' properties of lower level categorical properties.

Call this the

Higher Order Theory Dispositional properties are higher order properties of distinct categorical causal bases.

Why would one opt for the higher order, as opposed to the token identity theory? One motivation concerns the modal status of the laws of nature. Those who hold that the laws of nature hold only contingently may be dissatisfied with an identity relation holding between dispositions and causal bases, for identity holds necessarily, and if the laws of nature are contingent, then it is plausible that the relationship between causal bases and dispositions is too.[11] A glass may be fragile in one world, not fragile in another, despite being intrinsically identical across both. Some may object: Does this not merely show that the identity must be contingent? Many would say 'no', particularly those tempted by Kripke's (1980) highly plausible view that the notion of contingent identity is but a philosophical chimera, based only on conflations of the epistemic with the metaphysical. If there are no contingent identity relations, then

[11] Cf. (Prior *et al.* , 1982).

dispositions cannot be contingently identical to categorical causal bases. The typical response would be to take the identity to be at the level of non-rigid designators, thus making the identity hold at the level of reference.[12] Those properties that are the referents of token dispositional and token non-dispositional ascriptions are one and the same; they differ only in their presentational mode. But other worries remain. For one, there is serious dispute as to whether the notion of a token property, or a 'property-instance', is coherent. In what follows, I won't attempt to run the standard objections to the token-token identity theory, for if what follows is correct, there is a problem that dwarfs them. Namely, token-token identity theories do not avoid the problem of multiple realisability. Moves from type-type to token-token identities, then, no matter what else we may say, are dialectically inert.

Further still, I will have little to say about the higher order identity theory. I will argue that dispositions are modal properties, the view does not entail that they are higher order properties, but is consistent with it. It may be thought, for example, that the relevant modal properties are higher order properties. There are certainly structural similarities between modal properties and dispositions: for one, they are both multiply realisable. It may be that both Smith and Jones could kill a man with their bare hands, for instance. But the possibility associated with Smith may be based in his knowledge of martial arts, whereas in Jones it may be based in his enormous size and uncompromising strength. Nonetheless, I will jettison this issue. On whether such properties are best conceived of as higher order properties, I will remain silent.

2.2 Dispositions as Modal Properties

We are due a recap. So far, I have outlined the identity theory, and a reformulation of the theory in response to the problem of multiple realisability. In what remains of this section, I argue that the reformulation fails to avoid the problem. I then argue that dispositions are modal properties.

2.2.1 Plural realisation

I am now going to offer an argument designed to show that the move from a type-type to a token-token identity theory fails. The argument is relatively straightforward. It aims to show that there exists a class of the multiply realised dispositions, where the members of the class are individuated by the fact that their manifestations are not caused by a unique token of any categorical property. Because of this appeals to token identity fail. I call this class the *plurally realised* dispositions.

[12] Cf. (Mumford, 1998).

Plurally realised dispositions should be distinguished from what I shall call the *distinctly realised* dispositions, and the *variably realised* dispositions. We can define the latter two as follows:

> *Distinct Realisation* A disposition 'D' is distinctly realised just in case there exist two distinct entities, x_1 and x_2, such that Dx_1 and Dx_2, and the causal basis for Dx_1 $\neq$ the causal basis for Dx_2.
>
> *Variable Realisation* A disposition 'D' is variably realised just in case there exists an entity x_1, at two times t_1 and t_2, such that Dx_1 at t_1 and t_2, where the causal basis for Dx_1 at t_1 $\neq$ the causal basis for Dx_1 at t_2.

The former concerns the prototypical case, such as those given above. Two objects may be elastic, even if both are elastic due to distinct categorical properties. The latter concerns change in causal basis over time. The very same object has at one time a given causal basis for its disposition, but at some other time the causal basis has changed. The standardly discussed cases involve neuroplasticity, where the causal basis for an agent's mental state is said to change with neurological adaptation.[13]

In contrast, I am concerned with dispositions that satisfy:

> *Plural Realisation* A disposition 'D' is plurally realised just in case there exists an entity x, such that Dx, C_1 is a causal basis for Dx, C_2 is a causal basis for Dx, and $C_1 \neq C_2$.

Plurally realised dispositions are ones that are realised by *more than one causal basis in the same object at the same time*. One purported example is given by Mackie. He writes:

> Even in the same material, the same disposition may have more than one ground. A piece of cloth may absorb water in two ways, by the water being taken into the individual fibres and by its being held in spaces between the fibres: its absorbency then has two different bases, the molecular structure of the fibres and the larger-scale structure in which those fibres are spun and woven. (Mackie, 1972, p. 148)

Mackie's case may be ersatz. Those tempted by the view that the bases of a disposition must be intrinsic properties may resist the example. It may be thought that the spaces are *not* intrinsic properties, but relational and thereby extrinsic. After all, we could imagine that the threads were so tightly packed that the cloth would no longer trap spaces between the fibres despite retaining its intrinsic properties. In any

[13]For interesting discussion on neuroplasticity, see (Hurley & Noë, 2003).

case, there are more obvious examples. Cigarette smoke has the disposition to damage the lungs once inhaled, but that disposition arises due to a wide variety of chemicals present in the smoke's composition.

I believe that the potential force of such counterexamples has largely been overlooked by those attempting to advocate token-token versions of identity theory. The reason this class is so problematic is as follows: The multiple realisation occurs not at the type level, but at the token level. Because of this, we can run a *reductio* that is still problematic for the identity theory, even the token-token version. Consider:

> *Overkill* A vial of poison x contains two deadly chemicals, 'DEATH_1' and 'DEATH_2'. In most individuals, each chemical taken by itself is sufficient to give rise to the disposition to kill when ingested. But some individuals are perfectly resistant to 'DEATH_1', and others to 'DEATH_2'.

I will start with the token-token identity theory. With the instance to hand, we may now re-run our *reductio*. First, consider:

> *Resistance-1* Jones ingests x. Jones is perfectly resistant to 'DEATH_2'. Unfortunately, however, Jones is not at all resistant to 'DEATH_1', and thus, as a result of ingesting x, Jones dies.

In *Resistance-1*, Jones' death is caused by 'DEATH_1'. Not only that, but Jones' death is a manifestation of the vial's disposition to kill when ingested. Given the token-token identity theory, it follows that the vial's disposition to kill when ingested is identical to the causal basis 'DEATH_1'.

The problem should now be obvious. We can re-run the argument for another case, merely switching the substance our agent is resistant to, and by doing so generate a contradiction. Consider, for instance:

> *Resistance-2* Smith ingests x. Smith is perfectly resistant to 'DEATH_1'. Unfortunately, however, Smith is not at all resistant to 'DEATH_2', and thus, as a result of ingesting x, Smith dies.

In *Resistance-2*, Smith's death is caused by 'DEATH_2'. Not only that, but Smith's death is a manifestation of the vial's disposition to kill when ingested. Given the token-token identity theory, it follows that the disposition to kill when ingested is identical to the causal basis 'DEATH_2'. By the transitivity of identity, it follows that $\text{DEATH}_1 = \text{DEATH}_2$. But *ex hypothesi*, $\text{DEATH}_1 \neq \text{DEATH}_2$. We have our contradiction.

Notice that I *do not* deny that disjunctional states are the *truth-makers* of certain propositions concerning the presence of dispositional properties. It is, rather, only to deny that the dispositions may be reduced to or identified with them. The basic

problem is that what is responsible for a disposition's manifestations may be *two distinct* bases. If that is so, then we cannot identify a disposition with its causal basis, for the disposition is occupied by two distinct categorical properties. Plurally realised dispositions are, then, as threatening to the token-token identity theses as standard cases of multiple realisability are for the type-type identity theory.

How might the identity theorist respond? Other than accepting the identity of DEATH_1 and DEATH_2, or denying the transitivity of identity, I will consider two options.

Reply 1. Complex basis In some cases, a disposition has some conjunction of causal bases. The flint, the gas, and the sparker all together give rise to events which cause the manifestations of the lighter's disposition to ignite when sparked. In such a case, we would identify the disposition with the *conjunction* of those bases: The complex composite of the flint-gas-sparker. Could we not say that the causal basis is the conjunction of DEATH_1 and DEATH_2?

The answer is no. For the reason we would identify the disposition with the conjunction in such a case is *mereological*. It is because the flint-gas-sparker whole, of which the individual components are parts, causes the manifestations of the lighter. But in the present case, the manifestations are not (always) caused by some whole of DEATH_1 and DEATH_2. Each is individually unnecessary for the manifestation's occurrence. The obvious response is to appeal to a *disjunctional* causal basis. But this also won't do. For disjunctional states are causally inefficacious. As Shoemaker writes:

> But if we characterize a disjunctive property as "the property of being F or G or . . . ," specifying it by a list of its disjuncts, it can easily seem that the property is defined into existence, is in some sense a logical construction out of its disjuncts, and is not the sort of property that could enter into causal laws or have causal efficacy in its own right. (Shoemaker, 2007, p. 17)

I cannot be killed by DEATH_1 *or* DEATH_2, at most this would express ignorance. There simply are no complex disjunctional causal bases.[14] So dispositions cannot be identical to them. My opponent may resist the last step: Why not allow the causal efficacy of disjunctional states? But their question may be answered readily. For suppose that disjunctional states are causally efficacious; we would have admitted gross overdeterminacy into our causal ontology. For if Jones may be killed by DEATH_1 *or* DEATH_2, what reason is there to exclude DEATH_1 *or* the water the substance is dissolved in, or the other infinite possible disjunctions one may form?

Perhaps it will be argued that to count as a causal disjunct it need only be *possible* for each of the disjuncts to bring about the effect. But this won't do. For suppose

[14] Cf. (Lewis, 1986a).

that x causes e, yet it was possible that y and z cause e also. It would follow that x *and* the disjunction of y and z caused e! Further overdetermination looms. Another modification would be to require that one of disjuncts *in fact* caused the event, but recall that we are trying to establish that the *disjunction* is causally efficacious. Further still, the move is not independently motivated. We should not be moved by *ad hoc* retreats of this form. There is no good independent reason to take the notion of disjunctional causation seriously. The 'cement of the universe', as Hume put it,[15] does not involve disjunctional states of affairs.

Reply 2. Double dispositions Perhaps the most plausible reply is that such cases are ones in which objects have not one, but two dispositions. This reply is lent credence through consideration of the fact that in *Overkill*, the two chemical compounds may be thought to cause death via two distinct methods. It may be thought, then, that this is not a case where x has two categorical bases for a single disposition, but rather two categorical bases for two distinct dispositions. This would almost certainly be the response Heil would offer. When writing on multiple realisability, he notes:

> All this will seem plausible only so long as you are content to characterize fragility in a relatively non-specific way. If being fragile is described as shattering when struck by a massive solid object, for instance, this is something shared by a light bulb, an ice cube, and a kneecap: same higher-level dispositional property, different lower-level realizing properties. Light bulbs, ice cubes, and kneecaps shatter in very different ways, however. These "ways" reflect these objects' possession of distinct, though similar, dispositions. (Heil, 2004, p. 246)

On this view, the problem is resolved at the level of predication, we have essentially mistaken *'a non-specific predicate satisfied by a range of imperfectly similar states or properties for a specific predicate satisfied by a unique higher-level "multiply realized" state or property. Putative lower-level realizers of fragility are really just different ways of being fragile'*.[16] For analogy, take the predicate 'being red'. This is, on Heil's view, a general imprecise predicate, which may denote a range of similar, though distinct properties. A handkerchief that is both scarlet and crimson is red in has *two* properties of 'being red', for it is red in *two ways*. An alternative way to pitch this would be with use of the distinction between sparse and abundant properties. It may be thought, for instance, that the predicate 'red' denotes an abundant property, perhaps some disjunction of sparse dispositional properties.

Now, Heil's contention is that multiple realisability *in general* is not problematic for the identity theorist. It is worth bearing that in mind, for without multiple realisability

[15](Hume, 1965).

[16]See (Heil, 1999) for a complete defence of this view.

the type-type theorist has no (or at best significantly less) motivation to move to a token-token account. Thus, at minimum a conditional claim will hold: *if* multiple realisability is a problem, *then* token-token identity theories fail. Nonetheless, there are good reasons to accept its non-conditional analogue, for it is far from obvious that appeals to the imprecision of predication are sufficient to dull multiple realisability's blade. I will give two.

Response A The view fails to respect the way we talk about dispositions: We do not say that an object has *two of the same* disposition. For instance, we would not say that a glass has *two* fragilities. I do not deny that the *parts* of the mixture may be said to have distinct dispositions, but the *whole* does not. One would, rather, say that the mixture is poisonous *for more than one reason*. This is particularly vivid in the case of states of mind: If one's belief that p was disjunctively realised, one would not thereby be in *two* states of belief that p. Put another way: Dispositional properties do not, *prima facie*, admit of double counting.

Response B It may be that DEATH_1 and DEATH_2 both kill in *exactly the same way*. Perhaps, for instance, they both block a certain kind of receptor. Notice this is consistent with each chemical being causally distinct *vis-á-vis* their masking, as in the cases above. What makes the two chemicals *masked* may concern causal properties irrelevant to the manifestation of the disposition. If the chemicals both enter into the same casual relations when the disposition manifests across cases, it seems plausible to suppose that they base the very same disposition.

That completes my replies. I turn now to my defence of the modal account.

2.2.2 Dispositions as modal properties

Above, I argued against the causal identity theories. The argument that follows is more constructive in spirit: I will support what I shall call

The Modal Identity Theory Dispositions are modal properties.

If dispositions are to modal properties, which modal properties are they identical to? Recall in the last chapter I argued that dispositions may be analysed as causal contingencies. I will argue: Dispositions are those modal properties. If that is correct, it is precisely because dispositions are identical to such properties that they are coextensive with them. The defence will be comprised of both a positive and a negative component. I will start with the latter: I will defend the view from three objections. I will then briefly offer some support.

We start with the objections. Here they are:

1. Loss of parsimony One of the major reasons given in support of the causal identity theory is its *parsimony*. Once dispositions have been identified or reduced

to their causal bases, we no longer need to appeal to occult *possibilia*, nor anything else not already within our ontology. If dispositions are modal properties, we lose that parsimony. This loss of parsimony is a strike against the modal identity theory. Now, whilst I agree that the view is less parsimonious, in that it requires *possibilia* to account for dispositions, I maintain that it is *no strike* against the modal identity theory. Parsimony is only desirable when we have two competing explanations, where one is more parsimonious *ceteris paribus*, and only when we *can* explain the relevant phenomenon with fewer theoretical resources. If what I have argued in this chapter is correct, we cannot. Causal bases *do not* account for dispositional properties. So the parsimony may be lost, but loss of parsimony is not always a bad thing. Dispositions are a case in point.

It should be finally noted that I do not intend to disclude reductionist accounts of modality. It may be, for example, that modal properties reduce down to some other kind of property. In which case, so too for dispositions. For what it is worth, I doubt a such reductionist accounts will work; I am in sympathy with a fictionalist view of modality, most likely some form of combinatorialism.[17] But importantly, I claim only that dispositions are not identical to their causal bases, not that they are irreducible and non-identifiable *tout court*. So, in a sense, the view is *as parsimonious as our best total theory*. We need to account for modality. Whatever account is given of modality will yield an account of dispositions.

2. The argument from causal efficacy I turn now to Mellor's objections. First, he provides an argument from the causal efficacy of dispositions. It may be found in the following passage:

> [I]f mere possibilities distinguished fragile from other glasses, fragility would be no real property, and change in it can be neither cause nor effect. The latter view has indeed been held but it is evidently false. Consider a rod so twisted that, when put in liquid helium to make it brittle, it breaks. Its becoming brittle is caused by the cooling and in turn causes it to break [...] So dispositions are real properties in a sense that rules out any account of them as mere potentialities or possibilities. (Mellor, 1974, pp. 172-173)

It may be formalised thus:

<u>Argument from Causal Efficacy</u>

1. If x is a modal property, then x is causally inefficacious.
2. Dispositional properties are causally efficacious.

therefore

C. Dispositional properties are not modal properties.

[17] For a classic example, see (Armstrong, 1989).

The argument is valid, the first premise has initial plausibility, and the second is supported by the example of the rod. Dispositions can be *effects* (the effect of being put in the liquid helium) and *causes* (the cause of the breaking). So it follows that dispositional properties can be causally efficacious, and thereby cannot be mere modal properties. It follows that the modal identity theory is false.

We should not be perturbed by the argument; it may be rejected on both fronts. First note that, strictly speaking, the example given does not include dispositions as causal relata, for properties do not enter into causal relations: *events* causally relate.[18] Of course, manifestations are events, but manifestations are non-identical to dispositions, for manifestations need not occur, despite the presence of a dispositional property. Now, in the example given, it is the *rod's becoming* brittle that we are told causes the breaking, and the dipping causes the rod *to become* brittle. Dispositional properties do not enter into causal relations, because properties do not enter into causal relations *simpliciter*. So the example of the rod gives no support to the second premise.

In reply, Mellor may reformulate the argument as follows:

Argument from Causal Efficacy*

1*. If x is a modal property, then events of the kind 'becoming x' are causally inefficacious.
2*. Such events are causally efficacious.
therefore
C*. Dispositional properties are not modal properties.

but once reformulated, pressure may be put on the first premise. After all, one might say that the event 'the safe's becoming locked' causes the event 'the treasure's becoming impossible to steal'. The effect here is an event, i.e., the attaining of a property, and a modal property at that. Similar points apply to causes: The treasure's becoming impossible to steal may well cause it to persist unplundered. So the attainment of modal properties *can* enter into causal relations, *contra* the first premise.

It might be doubted that such events are really causally efficacious.[19] Perhaps to say that the treasure was not plundered because it became impossible to steal is at most an *explanation*, one that does not simply give information about the actual causal history of the event.[20] But such suspicions, I contend, would carry over to the dispositional events in question. It may well be argued that the rod's becoming brittle is *not* the cause of the rod's breaking. Rather it is the sudden change in temperature, or the rod's

[18] I assume here the standard view. No doubt, some will be discontent. For instance, Bennett (1988) takes *facts* to causally relate. But I take it that the view that properties are causally efficacious is, at best, contentious. See also (Mackie, 1974).

[19] For example, Bennett writes 'I submit that no modal facts—as distinct from their nonmodal bases—are causally efficacious'. See (Bennett, 1988, p. 27).

[20] Cf. (Lewis, 1986a).

being so dipped, may be properly called the cause of the rod's breaking. So to disallow modal properties to enter into the causal relations given above would be to disallow the dispositional properties to enter into the causal relations Mellor requires. Either way, I am in the clear: The argument from causal efficacy fails.

3. Modal Irrelevance The second argument Mellor gives is found in the following passage:

> Dispositional ascription entails statements of (admittedly conditional) fact, not statements of possibility. A fragile glass is one that does break (if dropped), not one that can break. Whether it can break depends inter alia on whether it can be dropped, and its being fragile entails nothing about that. (Bear in mind it must be physical, not logical, possibility at issue here.) The safety precautions at our nuclear power station are intended to prevent an explosion by making impossible the conditions in which the fuel would explode. It is ridiculous to say that their success robs the fuel of its explosive disposition and thus the precautions of their point. (Ibid., p. 173)

The basic thought is that whether or not an event is possible or impossible is sensitive to irrelevant details of a given case. In the case given, the precautions make it the case that it is not possible for the fuel to explode—yet it is still disposed to do so. It follows that there are some cases in which an object possesses dispositional properties where the relevant modal property is not possessed. It follows that dispositions are non-identical to modal properties.

This argument is no good, for Mellor may be charged with equivocation: The argument moves from one sense of 'can' to another. Recall that my analysis requires only that the fuel in Mellor's example *could* be caused to explode by being ignited, given the laws of nature. To say that it *could not* explode, given the laws of nature, due to the fact that there are safety precautions alters the kind of contingency at issue. It may be the case that it could not explode, *given that there are safety precautions*. But this does not imply that it could not explode *tout court*. Or, to take his other case: one can sensibly say in one breath that the glass could break if dropped, and in the next that it cannot be dropped without any pain of inconsistency. The consistency arises from the fact that the modal evaluations concern different classes of possible world. I do not say that a fragile glass can be dropped, given the various details about a particular case. All I claim is that it could—on a weak nomological reading of 'could'—break if dropped. Mellor's claim is silent with respect to that requirement.

I have defended the modal identity theory against three objections. A question remains: Why should we accept it? The argument in its favour I will now offer rests primarily on its theoretical virtues, ones which, I contend, make it favourable in contrast to its contenders. I give three. They are as follows:

1. Unity The account is more *unifying*. The notion of unity is highly vexed[21], but a slapdash account, which will suffice for present purposes, is that a theory T unifies a set of data Δ more than some rival theory T^*, just in case T is able to offer a more similar explanation of the members of Δ than T^*. For example, suppose our data set is as follows: Smith's car is not parked in front of his house, his keys are not on their hook, and the back window has been smashed. We could offer a theory that explains all of these independently. Perhaps the window was smashed by a cricket ball, an opportunist entered and stole the keys, and sometime later the car was hot -wired by a professional thief. But a better explanation is one that unifies the data: Smith locked himself out, broke in through the back window, and took his car wherever he needed to go. Of course, a better example would provide two theories equally confirming in other respects, but for purposes of brevity the present case should suffice. The point is that, in general at least, unification is a virtuous thing, theoretically speaking.

Evidently this is no scientific theory, and we are not here primarily concerned with the explanation of *events*. But explanation is not only given of events, but also of the holding of theoretical claims.[22] The sense of unity at issue here concerns the ability to offer answers to philosophical puzzles. More precisely: The modal account is more unificatory than its competitors, because it is able to give a single answer to both the *semantic* and the *ontological* questions that guide the present inquiry. For on a standard version of the causal identity theory, as already noted, to ascribe a dispositional property to some subject is to pick out an unknown categorical property, and to characterise it in modal terms. On this view, we get a distinction between the ontological question and the semantic question. What a statement such as 'that yarn of cotton is flammable' *means* is 'there is some categorical property of that yarn of cotton, which would cause it to catch fire if ignited'. But the dispositional property simply is the referent of the existential quantification, not the existential quantification itself. On the present view, in contrast, the two questions are given one and the same answer: What one *means* is the very same as what the disposition *is*. When I say 'that yarn of cotton is flammable', I simply mean 'the yarn could be caused by igniting to catch fire'. Moreover, the disposition is the very thing I mean: It is the possibility of being caused by ignition to catch fire. In that sense, the modal account, in contrast to its categorical competitor, has greater unificatory power.

2. Simplicity The second theoretical virtue is *simplicity*, a virtue equally vexed. Typically simplicity is given a two-fold classification: There is *syntactic* simplicity (which is, strangely enough, sometimes called 'semantic' simplicity), and *ontological* simplicity. The latter may be crudely put in Ockham's terms: No entities beyond

[21] For interesting discussion, see (Friedman, 1974), (Kitcher, 1981), (Kitcher & Salmon, 1989), (Schwartz, 1999), (McGrew, 2003), (Myrvold, 2003), and (Patrick, forthcoming).

[22] Cf. (Lipton, 2004)

necessity. But it is the former with which I am concerned. Syntactic simplicity concerns a theory's principles. For example: The Copernican theory of the celestial system was less simple than its Ptolemaic predecessor. What I mean by 'simpler' here is as follows: The account makes use of fewer conceptual resources. On the causal identity theory, to explicate a disposition requires the use of both modal properties *and categorical* properties. They are categorical properties with certain modal profiles i.e., a categorical property *that gives rise to certain modal properties*. On the present view, in contrast, a dispositional property *just is that modal property* which it gives rise to. The account is simpler in that regard.[23]

3. Elimination The final argument is one from elimination. We have seen that dispositions cannot be identical to causal bases. What is left? The objects must have modal properties which must explain why they manifest. I say: The only option left is identification. If they are not the causal bases then they must simply be the modal properties themselves.

Conclusion

This chapter had the following aims:

[2.1] To outline the identity theory.

[2.2] To argue that the problem of multiple realisability holds at both the type and the token level.

[2.3] To defend the view that dispositions are modal properties.

which have now been satisfied. That completes my investigation into the metaphysics of dispositions. All that is left now is to investigate the relationship that holds between dispositions and *functions*, and to outline the notion of a *functional norm*.

[23]For discussion on simplicity as a theoretical virtue in the sciences, see (Alan, 2003), (Cowling, 2013). (Goodman, 1961), and (Grünbaum, 2007). For interesting discussion on simplicity arguments (such as given here) in metaphysics, see (Willard, 2014) and (Brenner, 2017).

Chapter 3

Functions

Introduction

In the previous two chapters, I explored the truth-conditions and ontological status of dispositional properties. The purpose of this chapter is to outline the relationship that holds between dispositions and functions. On a widely endorsed view, the occupation of functional roles implies the possession of dispositional properties. Its acceptance is perhaps in part due to the influential work of Robert Cummins (1975), who argued just that, and outlined the method of 'functional analysis', a method many hold to be applicable to our mental conceptual scheme. Unfortunately, however, the view is rarely given independent support. To some, it may sound platitudinous. Nonetheless, as Lewis once wrote:

> It is the profession of philosophers to question platitudes that others accept without thinking twice. A dangerous profession, since philosophers are more easily discredited than platitudes, but a useful one. For when a good philosopher challenges a platitude, it usually turns out that the platitude was essentially right; but the philosopher has noticed trouble that one who did not think twice could not have met. In the end the challenge is answered and the platitude survives, more often than not. But the philosopher has done the adherents of the platitude a service: he has made them think twice. (Lewis, 1969)

If it is the profession of philosophers to question a platitude, and our claim is platitudinous, then Stephen Mumford (2003) has done its adherents a service. For in *Dispositions*, Mumford presents a challenge for standard theorising on the relationship between functions and dispositions. His challenge bears both a negative and a positive component. The negative component is that functions do not entail dispositions. The positive is just the rejected claim's converse: That dispositions entail functions. In this

chapter, I will defend the platitude. By doing so, I hope, it may be better understood, for the arguments both for *and* against it rest upon a widespread conceptual confusion. In particular, they rest upon equivocation over distinct (albeit highly related) senses of the term 'function'. There are two central aims. They are as follows:

> [3.1] To disambiguate and outline four senses of the term 'function'.
>
> [3.2] To argue against the claim that bearing dispositions entails the occupation of a functional role, and to argue that the occupation of a function role entails the possession of certain dispositions.

Here is the plan. There are two sections. In the first, I deal with aim [3.1]. There are four subsections: The first outlines the mathematical and social senses of 'function', the second outlines the causal sense, the third outlines the teleological sense, and the fourth outlines the similarity and differences that hold between the latter two. In the second, I deal with aim [3.2]. There are two subsections: The first outlines two arguments, one for the claim that functions entail dispositions, the other for the claim that dispositions entail functions. Both are rejected on the grounds that they rest on equivocation. The second provides an independent argument for the platitude: That functions entail dispositions.

3.1 Four Senses of 'Function'

Mental states are functionally individuated. This sentence is ambiguous, for it contains the term 'function' and as Wright (1973) notes *'[l]ike nearly every word in English, "function" is multilaterally ambiguous'*. Since Wright's paper, much work has been done to keep the term disambiguated. Nonetheless, its ambiguity *still* leads to arguments based on fallacious equivocation. The present aim is the clear and careful disambiguation of the term 'function'. I will finish by briefly reflecting on the similarities and differences that hold between two central senses: The causal role, and the teleological senses.

3.1.1 Mathematical and social

I will briefly start with two senses that are *not* particularly relevant to the discussion that follows. The first we should do away with quickly; it is the 'social' sense of 'function'. It appears in statements such as:

S1. There is a function at the church this weekend.

S2. The Christmas Ball is an extravagant function.

The referents of such expressions are those most curious entities: *events*.[1] When I say 'the function' was well attended, the entity of which I speak is some occasion, an occasion attended by a significant number of individuals. Whilst there may be interesting connections between this sense and other senses of the term, for present purposes, the sense will be jettisoned. The reason for doing so is simple: It is far from obvious that mental states are events, and even if they *are* to be treated as events, they are not social events of the relevant sort.

I turn now to the second: The mathematical sense of the term. This will require a little more detail, for whilst it is not the sense relevant to the individuation of mental states, it *is* relevant with respect to the charges of equivocation I will employ in the second section. The mathematical sense features in statements such as:

M1. Distance is representable as a function of time.

M2. A production function holds between outputs of production and factors of production.

The mathematical sense of function is technical, and is thus explicatively defined. Functions are a class of relation—those relations which hold between inputs and outputs, where for each value of the input I, there exists a unique relation between I and a unique output O. A simple example would be the relation that holds between individual agents and ages. For each individual (at a given time) there exists a unique natural number which constitutes that individual's age. Each individual has an age, and each individual has *only one* age. In contrast, the relation that holds between individuals and siblings is non-functional. For each individual there may be none or more than one other individual that counts as that individual's sibling.

Mental states are not individuated in terms of the mathematical sense of 'function'.[2] Why not? For one, mathematical entities are (at least on a fairly standard view) abstract entities that do not admit of spatio-temporal location. It may be argued that mental states, in contrast, at least admit of temporal location; Smith may believe that p at t, but cease to do so at t^*. Some won't buy that: On Williamson and Steward's view, which we shall discuss, mental states are general, and are thus plausibly abstract. But regardless, even if abstract they are non-identical to the relevant mathematical abstracta. After all, functions may be operated on; mental states may not. None of this is to say that mental states cannot be *represented* or *characterised* with the use of mathematical functions. But that tells us little: The scope of functional representation is broad.[3] That mental states may be represented with mathematical functions is to say

[1] Cf. Davidson (1969a, 1969b, 1970).

[2] That is not to say that the notion of a mathematical function has not been employed in functionalist theorising. Early functionalists such as Putnam (1967), clearly used the term 'function' in places in its mathematical sense. See also (Bird, 2018).

[3] Cf. (Crane, 2003, p. 86).

little about the distinctive nature of mental states. Whilst the mathematical sense will be required for my argument against Mumford, it will not be the adopted sense of the claim that mental states are functionally individuated.

3.1.2 Causal role functions

I turn now to the third sense of the term 'function': what I shall call the *causal role* sense, or just *causal* for short. Typically, when taking its causal sense, the term features in two kinds of statement. The first are statements of the form 'x functions as an F'. For example:

CR1. The organ functions as a heart.

CR2. The button functions as a screen dimmer.

Notice that with ascriptions of this kind we typically would not count the relevant causal functions as 'the function' or 'one of the functions' of the entity in question. Rather, to attribute a causal function is to say that the entity occupies a certain kind of causal role. If counting functions, what would be counted are functional role occupations, not the possession of functions. Under what conditions is it true that an individual x functions as an F?

In what follows, I will defend the following account:

<u>Function-as</u>

x functions as an F iff.

F1. $(\exists\chi)\ x \in \chi$

F2. $\Diamond\ (\mathrm{S}x \rightarrow^{c} e_f)$

The notation must be explained, and the conditions require support. That is the task to which I now turn. Take F1 for starters. The notation is as follows: 'χ' denotes what I shall call a *complex*, the set membership relation should be interpreted mereologically, as 'is a proper part of'.[4] So put into plain english, F1 says that there exists a complex, such that x is a proper part of that complex.

Why should we accept that? Here is some support. Suppose that we say of a given object that it is functioning as a water pump. Presumably, it must be functioning as a water pump *in some complex* that works in such a way as to pump water. Some would prefer to use the term 'system'. But the notion of a system is dissatisfying, for there

[4]Some might whinge at my use of set theoretic notation for mereology. Fairly standard practice dictates the use of non-alphabetic infix notation, but there appears to be no standard token non-alphabetic symbols that may count as 'standard notation'. Given the similarities that hold between mereology and set theory (Cf. (Lewis, 1969)) the abuse of notation will do no harm.

are entity complexes of which the parts may be properly regarded as functioning in certain ways, despite the complex in question not obviously falling under the extension of the concept expressed by 'system'. For consider tools: Something may function as a hammer, and may be doing so in virtue of its being in the hands of a hammerer. In such a case: What counts as the system? Are we to say some obnoxiously abundant entity—the 'human-*cum*-hammer'—is to count as a system? That seems strange to me; the term 'complex' gives a better fit. Because functioning in this sense partly requires complex integration, and complex integration is a relational property, functional occupation too is at best a partly relational property of objects. Because the properties are relational in this way, they are extrinsic. Intrinsic duplicates may differ with respect to their causal role occupation. Remove a heart from the body, and the heart will retain its intrinsics, despite ceasing to occupy its causal role.

It may be objected that for x to function as an F, that x need be part of no containing complex. For suppose that a water pump is dismembered from its whole; could it not still function as a water pump if, say, placed in a paddling pool and turned on? In such a case, the pump would not be *part* of any complex, it may be contended, despite functioning as water pump, *contra* F1. In reply, however, and to some extent this vindicates the more liberal notion of a 'complex', we should note that whilst the water pump will not be part of some mechanical system, that it may still be properly regarded as part of a complex in the relevant case; namely the 'paddling pool-*cum*-water-*cum*-pump'. The body of water in which the pump rests and the pool that contains the liquid may be thought of a complex in which the pump plays its functional role.

Pressure may be returned. With such a liberal conception of complexes, could not *anything* be properly regarded? After all, could we not take something to be functioning as part of the 'pump-*cum*-jungle-*cum*-trombone?'. The answer is 'no'. For the complexes in question must exhibit a certain degree of spatiotemporal continuity which our putative counterexample lacks. I do not say that such complexes are sparse entities, nor distinct from their components; but not *any* blend of entities will do. Only those with the right kind of spatiotemporal connection, whatever that may be.

And we can see why spatiotemporal continuity matters by considering F2. The notation that requires explaining is as follows: '$\rightarrow^c$' denotes a causal relation (cause for the antecedent, effect for the consequent), 'Sx' denotes some kind of stimuli x receives, and 'e_f' denotes what we may call an 'F-event'. What is this strange expression, 'F-event'? The basic view is as follows. For any function ascription, the function F may be linked to some kind of event e, an event which an entity, in order to count as possibly functioning as an F, must be able to cause once causally acted upon. Gosh, that was a mouthful! Here are some examples to clarify. First, consider 'functioning as a deadly weapon'. For something to function as a deadly weapon, all that is required is that it could be caused (by murderous intentions) to kill, either in some designed system,

or in the hands of an inventive criminal. Similarly if something functions as a set of brakes then it must possibly be caused (via an activating device) to cause deceleration.

The relevant causal relation must, as we saw with dispositions, be direct. It is for this reason spatiotemporal continuity of some form is required: No action at a distance—for x to serve a function, it must possibly enter into direct causal relations, and to enter into direct causal relations, x must be spatiotemporally contiguous with the relevant complex. A letter to a hitman may result in the death of Jones, but this does not show that the *letter* has functioned as a deadly weapon. The reason is that the causal chain is indirect. That is not to deny that a letter could *not* function as a deadly weapon, of course. Just that the deliverance of a message does not make it count as such.

It should be noted that for many functional ascriptions, there are a plurality of events that the entity must cause. To function as a liver, an entity must bear a range of functions. Strictly speaking, then, we should construe e as a set, of which the entity must possibly cause each member. Some may object that the requirement is too weak. For *far too much* would function as a deadly weapon on this account: rocks, spades, even bubbles of oxygen could kill, supposing they are injected into the bloodstream. In reply, however, the bullet should be bit: oxygen bubbles may be functioning as a deadly weapon, in the wrong hands at the right time. Being a deadly weapon may not be their *proper function*, but they may function as a deadly weapon nonetheless.

What kind of modality is at issue? Typically nomological, though it may depend on context. It does not appear *senseless* to ask whether something could function as a hammer if the laws of nature differed, though that depends upon the necessity of those laws. Similarly, it does not appear senseless to suppose that something *couldn't* function as it usually can, given such changes. But in any case, for all such ascriptions, there must be some laws held fixed. Similarly, just as the laws of nature are to be held fixed, F2 should *also* fix both the intrinsic properties of x, *and most* of those relational properties it bears to the complex in which it is so-integrated. Otherwise, it could possibly enter into the relevant causal relations simply by changing its integration. A heart duct taped onto a human body is, in some sense, integrated, and may possibly cause circulation given changes in the properties relevant to the integration. But we are not interested in cases in which those properties change, but only in which the heart remains affixed in the same rough-and-ready fashion. Of course, we cannot hold *all* of those extrinsic relational properties fixed – after all, entities often rely upon undergoing change in their extrinsic properties in order to manifest their functionally relevant dispositions. In order to function as a set of brakes, the brake pads and wheels cannot retain their spatial relations: the pads must touch the wheels. Put another way: We must exclude the relations relevant to the 'moving parts' of the complex. Determining which extrinsic properties we are to allow to vary should most likely be

done on a case-by-case basis; they are simply those that are sufficiently relevant to the manifestations of the function-relevant dispositions.

Another objection is as follows: x might not function as an x at some time t, despite satisfying F1 and F2, for it may be causally inert at t. For instance, the pancreas functions as an insulin producer, but only releases insulin upon the intake of a sufficiently calorific meal. But this relies upon the tempting though mistaken supposition that x is functioning as an F just in case it is *serving its function*, i.e., actually part of a system S, and causing the F-event relevant for the functional ascription. But this would be a mistake. For suppose that something is functioning as a doorstop. Does that imply that it is in fact, at every moment at which it is doing so, stopping the door? I say no! The door stop may be there, in front of the door, waiting for the wind to push it towards the frame, despite occupying a functional role. Or take to another example, a set of brakes may be functioning in a vehicle even if the brake lever is untouched. I suggest, then, that to say something is functioning as an F, is not to say that it is in fact causing such an event, but rather that it is part of some system, and that within that containing system it *could* cause an F-event. But, it may be asked, if that is so should we not replace the modal expression 'could' with the stricter 'would'? The answer again is 'no'. For to say that the brakes are functioning does not imply that they will *always* stop the vehicle, just that they could. True enough, to count as *fully functioning*, or to be functioning *well* they may need to stop the car on most occasions. But entities can function as brakes, even if they function poorly as brakes. That is accommodated by formulation with the stronger modality. So the chosen modal expression should remain.

With that in mind, F2 reads that x's being a member of the complex χ entails that it is nomologically possible that x cause some F-event, where the value is determined by the relevant given functional role. An example may serve to clarify. If x functions as a heart, then (1) x is part of some complex (perhaps a human body), and (2) x's being a member of that complex implies that it could possibly cause a 'heart-event', given certain stimuli (in this case, the circulation of the blood, given impulses from the sinus node).

That deals with the first class of ascription. Now for the second. The class I have in mind contains members that appeal to contingent modality, namely statements of the form: 'x could function as an F'. For example:

MC1. The water pump could function as a heart.

MC2. The kitchen knife could function as a deadly weapon.

Such ascriptions tell us not what an entity is in fact *doing*, but rather what it *could do*. But, it may be asked, have we not defined the first type above in terms of just that? After all, F2 requires that the entity, once subsumed under some complex, *could*

cause a certain kind of event. So the two kinds of statement cannot be distinguished in terms of modality, as I claim. Fortunately, the reply fails; it rests upon improper individuation of the scope of the modality at issue. For to say that x functions as an F is to give information about what x can do, to say that x *could* function of an x is to say of that entity that it *could* bear certain modal properties. The former kind are *singly* modal, the latter *doubly* so.

The account I have in mind is as follows:

Contingent function-as

x could function as an F iff.

F1*. $(\exists\chi) \Diamond x \in \chi$
F2*. $\Diamond (\mathrm{S}x \rightarrow^{c} e_f)$

The main difference lies in F1*. To say that something *could* function as an F is to make a claim concerning the possibility of its being subsumed under some system complex, in which it bears the relevant contingent causal relation to F-events. We are not interested in whether it *is functioning in some case* as an F, but rather whether *there is some case* in which is does. Put another way: To say that x could function as an F is not to say that it *in fact* occupies some causal role, but rather that it *could* occupy that causal role. The relevant kind of possibility, it should be said, fixes the intrinsic properties of the object.[5] For suppose that we take a heart so diseased that it could not function as a heart, at least in the state that it is in. There is a sense still in which it could function as a heart if it were to be the target of extensive surgery, perhaps. But this is not the sense that is at issue; when I say that a water pump could act as a heart, I am saying that it could be integrated into some system in which, fixing its intrinsic properties, it would circulate the blood. In short: The difference between actual and contingent function ascriptions does not concern the modality of the causal relation between x and certain F-events, but rather modality surrounding systemic integration. F2* must differ slightly, in that the accessible worlds will not be ones in which x bears most of the same integration properties it does in the *actual* world, but rather most of the relational properties it bears in the worlds in which the proposition ranged over by the possibility operator in F1* holds true. If something could function as a set of brakes, there is some world in which it is part of some system, such that holding fixed *those* integrational properties, it could cause an F-event.

Interestingly enough, and in contrast to the non-contingent analogues, possible causal occupations are plausibly intrinsic, namely for the reason given above: The intrinsic properties are fixed. After all, they make no requirement of mereological relations holding between the possible occupier, and the relevant complexes. Of course,

[5] Vetter, following Lewis, makes a similar remark. See (Vetter, 2014, p. 136), and (Lewis, 1976).

whether it could do so will be dependent, in part, by the properties of those complexes. Nonetheless, they are properties intrinsic duplicates will share, and are thus plausibly intrinsic.

In this section, I outlined what I called *causal role functions*, which are functions typically denoted by statements of the form 'x functions as an F', or 'x could function as an F'. These are the occupiers and the possible occupiers of causal roles respectively. It was argued that such ascriptions are essentially ascriptions of modal causal contingency. I turn now to the final concept expressed by the term 'function', the *teleological* sense of the term.

3.1.3 Teleological functions

If Smith had seen Jones place a pistol, a mask, and a bag marked 'loot' in his bag, then he would have understood the purpose of Jones' asking for a lift to the bank. The aim of the honey bee's visiting flowers is the collection of nectar; this behaviour also serves a greater purpose, as in collecting nectar the bee serves a crucial service to flowers—the dissemination of pollen. The tassels on some dresses are there to make pretty patterns during a dance, but there is no purpose to dancing other than having fun. These claims are teleological. That is to say, they concern the *aim*, *goal*, *point*, or *purpose* of phenomena.

Sometimes, the term 'function' expresses a teleological concept. Though as Wright notes:

> The notion of function is not all there is to teleology, although it is sometimes treated as though it were. Function is not even the central, or paradigm, teleological concept. But it *is* interesting *and* important... (Wright, 1973, p. 139)

I think Wright is correct that 'function' is not a central or paradigmatic teleological concept. In fact, strictly speaking, we might say that the term 'function', when used to express a teleological concept, is short-hand for the phrase 'proper function'. The teleological aspect of the phrase concerns its first constituent: The *proper* function of x is the function of x that x is *supposed* to serve. Alternately put: It is when the serving of a causal function is the purpose, point, aim, or goal of some entity.

Before we consider the teleological notion of function in detail, it will be helpful to give a general outline of the notion of teleology. As I shall be taking it, the term 'purpose' is the prototypical teleological term. What are purposes? So long as we are allowed an abundant reading, we may say that purposes are properties of entities. At least, we certainly *speak* as though they are; we certainly attribute purposes to entities, just as we attribute other properties. Just as we might say 'the colour of the ruby is red', we may say:

OP1. The purpose of the fuse is to prevent electrical fires.

OP2. The purpose of the yeast is to make the dough rise.

and to take examples involving events, just as we might say 'the turnout to the gala was poor' we may also say:

EP1. The purpose of Harry's telling the joke was to impress Sally.

EP2. The aim of Sally's rolling her eyes was to signal her boredom.

No doubt, language misleads. But assuming grammatical form to be ontologically indicative, purposes appear to be properties. Whenever an entity, whether it be an object or an event, has a purpose, that purpose is correlated with a kind of *content*. In the above, the purpose is denoted in the infinite mood: the purpose is *to* prevent electrical fires, or *to* impress Sally. Call the content of a teleological ascription its *τ-content*.

Not all purposes however, even so divided, are of the same ilk. There are three prominent kinds. The first we may call *intentional* purposes. These are purposes that feature in the descriptions of action such as those found in EP1 and EP2 above. They involve the intended purpose of a given act. It was Harry's intentions in telling the joke that determine the purpose; what Harry was *trying to achieve* by doing so. Notice that intentional purposes are, to use Williamson's (2000) terminology, non-luminous conditions. It does not hold that, for any behavioural event ϕ caused by an agent S, where the purpose of S's ϕ-ing is τ, that S is in a position to know that the purpose of ϕ is τ. Harry may join the gym, believing that his purpose in doing so is to keep fit, where his true motivation is developing a muscular physique to impress Sally.

The second class involves not human action, at least not directly, but rather human artifice. We may call these *artefact* purposes. These are purposes typically attributed to the objects of design, such as those found in OP1 and OP2 above. Note that in the latter case, what is being attributed the purpose is, despite appearances, not the yeast, but rather the appearance of 'yeast' on the recipe for bread. The third class are not the purposes of artefacts or action, but rather to what we may call *natural* entities. What is a natural entity? We may tentatively, in Quinean spirit, answer: Those entities stipulated as real by the biological sciences. Typically, they would be *traits*. They may be found in statements of the form:

NP1. The purpose of the heart is to pump the blood.

NP2. The purpose of the lizard's display is to attract a mate.

Notice that, as the examples above serve to show, natural purposes may be attributed to events or objects; the lizard's display is an event, the heart is an object. Such

ascriptions have presented somewhat of a challenge to philosophers of biology, for many take the use of teleological ascriptions in the biological sciences to be evidence of its inadequacy; that is, evidence that it relies on fundamentally non-naturalistic explanatory frameworks. Some vivid examples of this view are as follows:

> Biology is sick. Fundamentally unscientific modes of thought are increasingly accepted, and dominate the way the subject is explained to the next generation. The heart of the problem is that we persist in making (literally) sense of a world that we now know to be senseless by attributing subjective values to the objects in it, values that have no basis in reality. (Hanke, 2004, pp. 143-144)

> The notions of proper function, disease, and damage apply here and in a thousand other contexts; thinking in these terms is natural and apparently unavoidable for human beings. Of course, the compelling character of this sort of thought does not prove it coherent; nor does it show that these notions can in face be correctly applied, not just to artefacts, but also to natural organisms. [...] The bulk of mankind, however, has applied the notions of purpose and proper function to natural organisms, and has done so without any confusion or incoherence at all: for most human beings have thought that natural organisms and their parts *are*, in fact, designed. Most human beings *now* think so; certainly theists of all stripes do. (Plantinga, 1993, p. 196)

On this view, the attribution of teleological explanations to the natural world is to trade in fables; to engage is compelling but fundamentally mistaken anthropomorphic descriptions of nature. There is good reason, however, to resist this line of thought. After all, speakers of the language are perfectly aware that natural phenomena are not the product of design; it seems strange to attribute modern man with a belief that is commonly known to be false. We *know* nature is not designed, but this does not commit us to the denial of teleological explanations of the natural world. To suppose that people always presuppose design when making teleological ascriptions is simply wrongheaded. Moreover, the identification of natural purposes plays a central role in legitimate naturalistic explanation. At least, so says Karen Neander:

> [T]he apparent explanatory power of teleological explanations which appeal to biological functions is quite robust. That the koala's pouch has the function of protecting its young does seem to explain why koalas have pouches. That the bee's dance is for directing other bees to pollen does seem to explain why bees dance. I suppose it is just barely possible, perhaps, that

> this apparent explanatory power is illusory, based on hangovers from our Creationist past, or due to our mistaking the metaphorical for the literal, when we speak of 'Mother Nature's intentions', 'evolutionary design', and so on. However the thesis that we are persistently irrational in this respect is psychologically implausible in contrast to a theory of functions that shows such explanations to be legitimate. (Neander, 1991b, p. 457)

On whether teleological explanation is legitimate I will remain silent. But the literature that surrounds its endorsement provides important points for what follows, as the desire to legitimise such explanation resulted in a significant effort to analyse teleological ascriptions. A wide range of views emerged, though the dust seems to have settled, and it appears agreement has been reached that teleological explanations are essentially a form of *etiological* explanation; that is, explanations that make essential reference to the causal history of the bearers of purposes. On this view, an entity inherits its purpose from its *origins*.

Call this the

Origins Account An object's purpose is determined by its origins.[6]

The earliest version was developed by Wright (1973). Consider the following passage:

> The treatments we have so far considered have overlooked, ignored, or at any rate failed to make, one important observation: that functional ascriptions are—intrinsically, if you will—explanatory. Merely saying of something, X, that it has a certain function, is to offer an important kind of explanation of X [...] functional ascription-explanations are in some sense etiological, concern the causal background of the phenomenon under consideration. And this is indeed what I wish to argue: functional explanations, although plainly not causal in the usual, restricted sense, do concern how the thing with the function got there. Hence they are etiological, which is to say "causal" in an extended sense. (Wright, 1973, p. 156)

Why would we accept the account? Part of the motivation Wright offers is that the account makes sense of the explanatory role of functional ascriptions. When we demand the function of x, we are typically interested in *why* or *for what reason* it came about. For instance, the explanation-seeking questions:

[6]For a recent, though to my mind unsuccessful alternative, see Nanay's 'modal account' of proper functions. (Nanay, 2010).

Q1. Why did Harry wink at Sally?

Q2. Why is the emergency brake painted yellow and black?

Q3. Why do humans have hearts?

we may offer teleological *explanans*. To these questions, for example, we may answer:

A1. The purpose was to seduce her.

A2. The purpose is to make the emergency brake visible.

A3. The purpose of the heart is to circulate the blood.

And in all three cases, a statement with teleological content answers an explanation-seeking question. When we ask for information of the origins of x, then, a teleological ascription may serve as an explanation. If the origins account holds true, then that datum is accounted for.

The origins account certainly appears to work in the case of intentional and artefact purposes. After all, when we ask what the purpose of Harry's winking at Sally is, we are primarily interested in the content of the intention which played a causal role in the production of the behaviour i.e., the wink. Similar points apply to design functions; when we ask what the purpose of an artefact is, we are asking about the content of those intentions that gave rise to the entity's existence. The τ-content *just is* the content of the intentional origin. If I ask: What is the purpose of the tassels on the dress, then on this account I am asking what the tassels were intended to do when the dress was designed. Nonetheless, the origins account is less obviously applicable to natural purposes, for precisely the reason that there appear to be no intentional origins, *contra* the postulations of the pre-Darwinian religiously dominated biological sciences. If there is no intentional content, how are we to determine the τ-content of a given natural purpose?

A widely endorsed view, developed primarily by Karen Neander (1991a, 1991b) and Ruth Millikan (1989) puts to rest the requirement of intentional origins.[7] On this view, what matters is not that the origins are intentional, but rather that the entity arises from design: Intentional action is but one form of designed behaviour.[8] Natural selection is design also, albeit one unguided by the contents of intentional mental states. So to say that the purpose of the heart is to pump the blood, but not to make a noise, is to say that its pumping the blood is what the trait was *selected* for; its making a noise is merely accidental, a property that cannot enter into a legitimate explanation of the persistence of the trait in heart-bearing organisms. The τ-content of a natural purpose arises from natural selection, in contrast to their intentional and artefact counterparts. A fairly

[7]Cf. Godfrey-Smith (1993, 2012) and Schwartz (1999).

[8]If forced to fudge, we could speak of 'design-like' processes also.

standard objection to the origins account that is worth noting, especially with respect to natural purposes, is that traits cannot make contributions to their own selection; it is the beating of *previous* hearts, not a heart in fact attributed a purpose, that explains the presence of the trait. But this objection rests on a confusion: The relevant bearer of the purpose in the case of natural purposes are not tokens, but types. It is the *trait* 'heart' that has a purpose; at most, individual hearts inherit their purpose from their type.

One neat implication of this is that the origins account is highly unificatory. τ-content, no matter the purpose's kind, is determined by design origins, whether the design be selectionist or intentional. In other terms: To demand the purpose of x is to ask what x was *designed to do*. This is independently plausible. In trying to ascertain Jones' purpose in winking, one is trying to ascertain the intentional component of the design plan that resulted in the lashes' movement.

A final point worth making is that the origins account tells us something important about purposes; namely that they are, like the functions-as kind of function, extrinsic properties of entities. Why so? The argument is as follows: It is possible that there are two intrinsic duplicates which are nonetheless non-identical with respect to their purposes. Given that purposes are properties, they must be extrinsic. Consider for example the human heart. In our species, the heart functions as a blood pumper, but it does not function as a noise maker. But it may function as a noise maker, without any change in its intrinsic properties. This is, of course, because the purpose is set by its origins which are plausibly extrinsic. For take a particular human heart, and consider what its purpose would be in another world, had its origins differed in the following way. Suppose that it had developed in some other species, 'Schumans', who live on a quiet planet. On this planet the beating of the heart serves a purpose; when a heart beats strong, a member of the species is seen to be a more attractive mate, and so the noise made by a given heart features in an explanation of its persistence through the species. In such a case, the heart, identical in its intrinsic properties though it may be, differs in its purpose. So purposes are not intrinsic properties; they are extrinsic, relational properties—ones determined by the entity's origins.

So far, I have distinguished four senses, the 'social', the 'mathematical', the 'causal role', and the 'teleological'. In what remains of this section, I would like to briefly give attention to the relationship that holds between the two senses detailed over the last two subsections: The teleological and causal senses of 'function'.

3.1.4 Relation between causal and teleological functions

What is quite remarkable about the causal and teleological senses of 'function' is the extent to which they yield equivocation; it is fairly standard practice to conflate these

senses of the term. As Godfrey-Smith writes:

> Although it is common in the philosophical literature to understand these two views as competing analyses of a single concept, there is no obstacle to accepting both concepts as legitimate. On this view, 'function' in scientific contexts is ambiguous – or rather, more ambiguous than it is usually thought to be. (Godfrey-Smith, 2012, p. 15)

This is all bad business, partly due to the fact that the senses bear striking differences. Consider for example the following two:

Difference 1 The truth-makers are temporally distinct.[9] What makes a teleological functional ascription true is an entity's history. In contrast, what makes a causal role functional ascription true are the details about its current dispositions and its integration (possible or actual) into certain complexes. Thus, the truth-makers for the latter are *current* or *obtaining* states of affairs, whereas the truth-makers for teleological functional descriptions are *past* or *non-obtaining* states of affairs.

Difference 2 They may enter into distinct explanations. I will give two examples. First, whilst both causal role and teleological ascriptions can enter into explanations of events that occur when things are 'going as they should', causal role ascriptions can enter into explanations when things do not. For example, I can explain how Smith got from A to B by saying that he took a car, and by citing how cars function (read teleologically). But if the car is *malfunctioning* at least one non-proper function must enter into the explanation. If the engine breaks and the car crashes, I may appeal to how things were *functioning* in contrast to how they are *supposed to function*. So some events causal functions can enter into proper functions cannot, because some events occur due to the serving of non-proper causal roles. Second, teleological explanations can enter into etiological explanations that causal role functions cannot. If I wish to explain the persistence of a trait, then I can cite the function of previous instances of the trait. In such a case, what explains why the token is present is not its actual causal profile – again, this token may malfunction. Rather, what does the explanatory work is the proper function – that causal function it was selected to serve.

If so distinct, what explains the readiness with which the senses are conflated? The answer: Certain *similarities* between the usage of the two senses. Again, I offer two:

Similarity 1 The first similarity is that whilst both enter into different kinds of causal explanation, that often a complete explanation of a phenomenon must make explicit reference to *both* kinds of function. Consider Lean:

> [E]ven the strong theory of functions relies in some sense on the non-historical, causal-mechanistic notion of function discussed by Cummins

[9] For an excellent read on truth-making, see (Armstrong, 2004).

> (1975). This notion of function is needed to make sense of the causal contribution to the organism's fecundity that selection favours in the first place. Etiological functions therefore entail a Cummins function, albeit one that was performed by past tokens of some trait, rather than the current token. (Lean, 2014)[10]

Lean's point here is that when it comes to natural purposes, teleological explanations often appeal to some causal function that the trait in question bears. We saw that in *Difference 2* above, though a particularly vivid case may be found in Cummins' later work. In *Psychological Explanation*, he attempts to re-cast Freud as a functional analyst. The details are not too important, but his method of determining the causal roles that are to be attributed involves terms quite obviously used to express teleological concepts. The following passage may serve:

> Functional questions about Freudian theory and about particular analyses are naturally couched in functional language: (i) When we ask what repression is, we want to know how a repressor would function, i.e., what its input-output properties would be. (ii) Once we grant that repression occurs, the next question is, why? Here, what we want to know is: What is the function of repression? (Cummins, 1983, p. 146)

In the latter part of this passage, he is in essence asking what repression is *for*? This questions asks the purpose of repression: What it was 'designed' to do. That is, what its *purpose* is in mental systems. Recall, it may have a function, without in fact serving that function. Those who excessively repress, for instance, may be thought of as undergoing some kind of cognitive *mal*function, and to classify such repression as malfunction is to say that the repression does not do what it is supposed to – does not occupy its appropriate causal role.

Or, to consider a non-fictional functional analysis, when we ask what the purpose of the heart is then we must make reference to what tokens of the type 'heart' did in the organism complex which resulted in the trait being inherited by subsequent generations. Similar points apply to artefact purposes; to understand the purpose of the brakes, one must make reference to what the brakes do in systems with certain features i.e., those causal roles brakes are supposed to occupy. So too for intentional purposes; the intention is the content of the state(s) which is causally responsible for the event's occurrence. The τ-content of Harry's purposive wink is the content of that intention that caused Harry to wink. The upshot of these remarks is that the relationship between the two is close, and thus their conflation is explained, because functional purposes are causal roles. That is to say: the *function* (read teleologically)

[10]See also (Buller, 1998).

of an object is just the way it is supposed to *function* (read causally). Put another way still: proper functions are the causal functions some entity is *supposed* to bear.

Similarity 2 We often *answer* questions that demand how an entity functions (in the causal sense) by offering their teleological function. If Smith asks, pointing at the brakes on my bike, 'What does this do?', I will typically give him the purpose, i.e., 'It stops the bike once the lever is pressed'. This may be false: the brakes may be malfunctioning, or functioning improperly. Nonetheless, it is of heuristic value to assume entities are working as they are supposed to be, and to correct once that is discovered to be false, rather than to determine, for each such question, what exactly the brakes are capable of doing in the system of which it is a part.

There are, then, similarities that hold between usage of the terms. But importantly, the two concepts *are* distinct. For an object may be *supposed* to bear some causal role C even if it plays no such role. As Aristotle put it, something may be part of a function category 'only in name'. Harry may wink at Sally, and the wink may be designed to seduce, even if it disgusts. Similarly, the button may be designed to dim the screen, even if the light cannot do so due to some fault; the heart may be designed to pump the blood, even if it cannot due to disease. This tells us that whilst it follows from:

A. F is the causal function of x.

that

B. x can F.

no such statement follows from:

C. x's proper function is to F.

That is, if something occupies a causal role, it must be able to serve that role. But it does not follow from something's being *supposed* to occupy a causal role that it does.

We are due a recap. The aim of this section was to disambiguate the term 'function'. That aim is now complete. First, I outlined four senses of the term, and looked at the relationship between the causal role and teleological senses. I now turn to the second aim of this chapter, which concerns the relationship between functions and dispositions.

3.2 Functions and Dispositions

Above, the term 'function' was disambiguated with care. The fruit of the disambiguation will be found below. I will reject a number of arguments in the literature concerning the relationship between functions and dispositions, on the grounds that they equivocate between the various senses. Following that, I will provide a (non-equivocatory) argument of my own for the view that functions entail dispositions.

3.2.1 Two equivocatory arguments

What is the relationship that holds between causal functions and dispositions? Sometimes, authors talk as though dispositions *are* functions. For example, consider:

> Biological functions are dispositions or effects a trait has which explain the recent maintenance of the trait under natural selection. This is the "modern history" approach to functions. The approach is historical because to ascribe a function is to make a claim about the past, but the relevant past is the recent past; modern history rather than ancient. (Godfrey-Smith, 1994, p. 344)

Godfrey-Smith is being liberal with ontological ascriptions here: if we were to be precise, then we would not say that functions of any sort *are* dispositions, though they may entail them. This point is of importance, for if functions simply *are* dispositions, then a functionalist theory of mind *just is* a dispositional theory of mind, albeit a dispositional theory of a certain stripe. But if they are not, as we should accept, then functionalism in the philosophy of mind is no *mere* dispositional theory, though it contains an important dispositional element. Let us call the weaker of these theses the 'functions entail dispositions' thesis:

FED Functions entail dispositions.

The thesis is rarely supported, and where it is, as I shall argue, its support rests on equivocation. Similar points apply to the thesis' converse, namely the 'dispositions entail functions' thesis:

DEF Dispositions entail functions.

This thesis is also rarely supported, though in this case the absence of support is likely due to its *prima facie* implausibility. Nonetheless, support exists; but like before, that support rests on equivocation. In this section, I am going to (1) outline and reject an argument for **FED**, and an argument for **DEF**, and (2) provide an alternative defence of the former; that functions entail dispositions.

Cummins on FED

Do functions entail dispositions? According to Cummins (1975), they do. His argument may be found in the following passage:

> Something may be capable of pumping even though it does not function as a pump (ever) and even though pumping is not its function. On the other hand, if something functions as a pump in a system *s* or if the function of

> something in a system s is to pump, then it must be capable of pumping in s. Thus, function ascribing statements imply disposition statements; to attribute a function to something is, in part, to attribute a disposition to it. If the function of x in s is to ϕ, then x has the disposition to ϕ in s. (Cummins, 1975, p. 757)

He also gives a case for elucidation:

> For instance, if the function of the contractile vacuole on fresh-water protozoans is to eliminate excess water from the organism, then there must be circumstances under which the contractile vacuole would actually manifest a disposition to eliminate excess water from the protozoan that incorporates it. (Ibid)

There is a lot going on in these passages. Let us start by drawing out the various claims Cummins is making. They are as follows:

> W. Something may be disposed to pump even though it does not function as a pump (ever).
>
> X. Something may be disposed to pump even though pumping is not its function.
>
> Y. If the function of x is to pump in s, then x must have the disposition to pump in s.
>
> Z. If x functions as a pump in s, then x must have the disposition to pump in s.

I hope the reader is aware of the equivocation here: Cummins runs together 'F is the function of x' and 'x functions as an F', as though the expressions are synonymous, which they are not. This is particularly clear in the case of Y and Z. Notice also that Y is patently false. We have already established that. For, if Y were true, then a statement of the form of B would follow from C, but no such statement follows. The function of x (in the teleological sense) may be to F, even if x is not disposed to F. A heart may be diseased, and may thereby not be disposed to pump the blood, even if its purpose is to do just that. Similar points apply to X: Something may be disposed to pump even though pumping is not its function, because something may be disposed to F, even though F-ing is not its purpose. Again, this is a teleological claim. The heart is disposed to make a beating sound, but that is not its proper function.

What, then, about the remaining claims? On closer reflection, W is rather puzzling. As Mumford notes:

> Cummins's claim that something can be disposed to *F* though it does not function to *F* seems to be using a sense of 'function' where *x* has a function to *F* iff *x* is actually used at some point to *F*. (Mumford, 1998, p. 200)

We could put Mumford's criticism in even clearer terms: W seems to use the term 'function' to mean 'is actually functioning' or 'has actually functioned' i.e., is *currently or previously serving* its function. So this gives us:

> W*. Something may be disposed to *F*, even if it never actually *F*-s.

to which an immature response is warranted: *Duh!* Of course something may be disposed to *F*, even if it never does *F*. This is a prototypical feature of dispositional properties, it is none other than that feature already much discussed: their latency. Quite remarkably, when criticising the argument along these lines, Mumford manages to equivocate. He writes:

> The function of a can-opener is still to open cans even if it never actually opens a can: for instance, if all cans are destroyed. (202)

To say that 'the function of a can-opener is still to open cans' is to make a teleological, not a causal claim. It tells us what the object is supposed to do, not what it in fact can do. That the purpose of something may not be manifested is irrelevant here: We are concerned with causal functions, not their teleological counterparts. That leaves us with Z. If something *causally functions* as an *F*, then it is disposed to *F*. Whilst I take it that this claim is true, it garners *no* support whatsoever from Cummins' chosen example. The statement 'the function of the contractile vacuole' is teleological; it asks what the purpose of the vacuole is. So the example fails to support **FED**, for it targets the wrong sense of the term 'function'. So Cummins' argument for **FED** fails, as it rests on equivocation.

Mumford on DEF

Mumford has argued not only that **FED** is false, but that its converse, **DEF**, holds true. But like Cummins', Mumford's argument is fallacious, as it rests of equivocation. We will call the former the *negative* claim, and the latter the *positive* claim. Let us start with

> *The negative claim:* Functions do not entail dispositions.

What reason does Mumford give? He writes:

> The response of the functionalist about dispositions to Cummins' attack should thus be along the following lines. Cummins is using a sense of

> 'function' where a function-ascribing statement is understood as a causal-role ascribing statement [...] [t]his is too narrow a sense of 'function'; because not all function ascriptions are ascriptions of dispositions of any kind. (Ibid)

For support, he gives the following series of examples:

> On this view, things can have functions but not dispositions. Examples could be the function of a flag being to add grandeur or the function played by a premiss in an argument. There are legal functions played by a magistrate, the function of a protest, the function of a rule in a game, the function of a road sign. These are fairly commonplace cases where in saying that something has a function, we are not saying that it has a disposition to do something. (Ibid, p. 203)

Mumford then breaks down the examples into two classes, giving support for their exclusion from the category of dispositions. In what follows, I will outline the classes, and reject his support. We will start with:

> *Class 1: Conventions* "There are functions of things that are determined by *convention* such as flags being a symbol of grandeur or a road sign being an instruction. These functions depend essentially on our responses to certain objects or symbols and for this reason it would seem a mistake to ascribe a disposition to such an object." (Ibid)

The first point that cries out for noting: These examples are clearly teleological. To say that the function of a flag is to add grandeur is to say what the *purpose* of the flag is, not to say what it *in fact does*. So too for the road sign; it is *meant* to be an instruction, irrespective of whether its message is ignored. So the examples appear to be non-starters, on equivocatory grounds. But let us assume that Mumford has in mind how a *particular flag* functions. Mumford's claim is that because the flag's functioning in a certain way is response-dependent, that it may not be properly regarded as dispositional.

But we should ask: Why should response dependency have this consequence? Prototypical dispositions are response dependent in this way: Whether a poison manifests its deadly disposition is determined by an agent's responses to the substance. Similar points apply to certain dispositional theories of colour. The response dependency is *built into* the theory.[11] So Mumford owes us a reason as to why this is problematic.

Let us now consider:

[11] See, for example, (Johnston, 1992) and (Peacocke, 1983).

> *Class 2: Relations* "There are functions that are determined by *relations* to other items in a system such as rules of a game or the function of a premiss in an argument. These functions would make no sense and would not be possessed outside the system. A premiss which is part of no argument is no premiss at all but only a proposition. Dispositions, in contrast, are possessed in the absence of conditions which would provoke their manifestation." (Ibid)

This problem here appears to be one of *complex integration*. The basic thought is that dispositions are possessed irrespective of their integration into complexes, but various functions are not. So not all functions entail dispositions. The example he chooses to support this, however, is somewhat puzzling. For one, it is unclear that propositions *are* the bearers of dispositions. If propositions are abstract entities, then at most it is that which expresses a proposition i.e., a statement, whether that be a sentence written or uttered, or the content of a state of mind that plays a role in an argument. Certainly, on a Russellian view, on which propositions are constituted from facts, propositions cannot play such roles: Facts cannot be premises, though they may be *expressed* by premises. Moreover, and *again*, the most natural interpretation of this claim is teleological. A proposition may *serve the purpose* of being part of an argument but it *need not* serve that purpose, if, say, it is never employed in a derivation.

Nonetheless, it may be thought that an objection remains. After all, Mumford is correct to suppose that an object can bear a disposition whilst not being part of a relevant complex, despite the relevant functional ascription holding true. But this is silent on whether functions entail dispositions. It may help us see that dispositions do not imply functions, but it cannot support the converse. It is true that when we say that x is functioning as an F, we are committed to x's being actually integrated into a system. And it follows that functional ascriptions are not straightforward ascriptions of dispositions. Nonetheless, the functional ascription *does* imply a disposition, it is just that a necessary condition on its manifesting that disposition is its being so integrated. So, whilst it is true that x will not be *functioning as* an F when not integrated, it still bears the relevant disposition. Let us now turn to

> *The positive claim* Dispositions entail functions.

This claim is given little support, though he does offer an argument from unification. Consider:

> I have made the claim that some dispositions may be understood as abstract, such as being divisible by 2. A functionalist theory of dispositions has the advantage of being able to explain why it is plausible that such abstract

> powers are dispositional in addition to the more commonly cited concrete dispositions of fragility, solubility, and the like. If a theory of dispositions can include such cases, at no added cost, then there seems no objection to including them even if there may be certain grounds upon which abstract dispositions are atypical. (Mumford, 1998, p. 203)

Presumably, when he says that the disposition is abstract, Mumford means that it can be attributed to abstract objects. At least, in the discussion that follows, that appears to be his claim. But read that way, it is unfortunate that our given example is 'divisible by 2', for this is not obviously an example of a disposition. Whether its inclusion is a benefit to the functionalist account, then, depends upon his argument for their inclusion, to which I now turn.

His argument is as follows. Properties of *abstracta* such as 'being divisible by two' share the prototypical features of dispositions, and thereby should be classified as such. There are four features. Consider first:

> *1. Generality* "[B]eing divisible by 2 is something that can be true of more than one number in the same way that being soluble can be true of more than one substance." (Mumford, 1998, p. 165)
>
> *2. Similarity* "Second, being divided by 2 is something that can be done to certain numbers in a similar way that being dissolved in water is something that can be done to certain substances." (Ibid)

Let us pause to consider these two. Firstly, are they different? Not obviously, 'being divisible by 2' is both true of many numbers, and many numbers can be divided by two. These two claims are synonymous. But more seriously: *So what*? Why does this give support to the view that they are dispositional? The property 'being dispositional' is presumably not dispositional, but it applies to a range of different entities. So it is not clear how this gives support to the view.

Consider now the second pair:

> *3. Divisibility* "Third, to predicate divisibility by 2 is to ascribe a functional role that a number can play; it is to say that if divided by 2, then a whole number will remain." (Ibid)
>
> *4. Stimulus conditions* "Fourth, focusing just on the example of divisibility, though it applies also to other mathematical properties, numbers can have different functional roles in different conditions. These may be regarded as corresponding to the different stimulus conditions for a non-abstract disposition where the functional role is a causal role. Hence, a single number may be divisible by two or by three or by five, as in the case of the number thirty." (Ibid)

The third assumes the conclusion under question, and thus may be jettisoned. The fourth does not. The thought is that we can see 'dividing by x' as a kind of stimulus condition. Just like other properties that bear dispositional properties, different stimulus conditions will cause different manifestations. Numbers, if we construe their divisors as stimuli, are alike dispositions in that respect. Closer inspection, however, reveals that treating divisors as stimuli is very odd. In fact, by some plausible assumptions, we can generate a strong argument against the claim that 'divisibility by 2' is a *bona fide* disposition. Firstly, it would entail that a disposition can both be part of its own stimulus, and its own manifestation. After all, 2 is both divisible by 2, in which case the bearer of the disposition i.e., the number 2 is part of its own stimulus, and divisible by 1, in which case the 'manifestation' would be itself—the number 2. This yields a substantive difference from standard dispositional expressions. If we are to accept that a disposition cannot be either part of its own stimulus, nor its own manifestation, then we should reject the claim that divisibility is a dispositional property, at least insofar as they are predicated over numerical subjects. That is *not* to say that 'divisible by 2' is not a disposition (though I doubt it is), where the subject is concrete. But the sense of 'divisible by 2', when applied to concreta, is distinct from the sense as it applies to mathematical abstracta. A melon is divisible by 2 because it could be caused to split in half. Numbers cannot be 'split in two' in *that* sense.

Moreover, the account fails to accommodate the conceptual centrality of causation involved in dispositional ascriptions. Mumford notes this diligently, and in response writes that:

> The argument from causal role clearly has no application to these cases, as causal roles are not involved. However, the property monist is not committed to monism with respect to all kinds of property. A fundamental split between 'abstract' and 'concrete' properties could be accepted [...] [t]o ascribe divisibility by 2 is just to give a functional characterization of that number's properties: it is to say what can be done with it... (Ibid, p. 167)

Ergh! More equivocation. To reiterate: One cannot 'do things with numbers' in the relevant sense. Numbers are abstract entities—one cannot literally divide a number by two. To say 'n is divisible by 2' is not to say what can be done with the number, it is simply to say that the value of x in the equation '$n \div 2 = x$' takes a natural number. Moreover, it is straightforwardly false that functional ascriptions tell us what can be done with the bearers of those functions. If to give a functional characterisation of x is to say what can be done with x, then a hammer is disposed to be a prop in a play, a piece of art, or to be thrown by the irascible. Not all modal properties of that ilk are dispositional properties.

With that in mind, Mumford has offered us *absolutely no* way to classify entities in such a way as to allow for abstract dispositions, without including a whole host of obviously non-dispositional properties.[12] In fact, it seems as though the causal requirement is *precisely why* we are moved to exclude divisibility (at least as it applies to numbers, and not objects) as *bona fide* dispositions. A dispositional ascription makes a claim about what an object could be *caused to do*, not what *can be done* with it in some broader sense. If Mumford wants to argue that dispositions are a kind of function, then he needs to offer us a plausible account of what he means by the term 'function'. His examples equivocate between the causal and teleological senses of the term, and his argument appears to endorse a strange new sense which identifies functions with mere possibilities. Now, it *may* be that Mumford has in mind the mathematical sense here. That would certainly unify the abstract dispositions at issue. But, as was noted above, mathematical functions may be used to *model* dispositions, just as they may be used to model a vast range of phenomena. Taken on this sense, then, the claim is empty at best.

In this section, I outlined and rejected two equivocatory arguments. The first was given by Cummins in support of **FED**, the second by Mumford in support of **DEF**. In the next section, I provide a (non-equivocatory) argument for **FED**.

3.2.2 Functions entail dispositions

When the term 'function' takes its causal sense, functional ascriptions entail dispositional ascriptions. An entity only has functional properties if it has certain dispositional properties. But not all dispositional ascriptions entail functional ascriptions. This is the view I now support. As before, we may split the claim into a positive and a negative component. The positive component states that all functions entail dispositions. The negative component states that not all dispositions entail functions. We will start with the former.

All functions entail dispositions: Both 'actual function as' and 'contingent function as', where true, entail the possession of dispositional properties. From either a statement of the form:

> x could function as an F.

or a statement of the form:

> x is functioning as an F.

[12] Admittedly, Mumford believes that *all* properties have both a 'categorical' and a 'dispositional' aspect. He is not the only one. See (Mumford & Anjum, 2011) and (Marmodoro, 2010) for defences. Nonetheless, the view is wildly implausible. For a powerful (no pun intended) argument to that effect, see (Bird, 2016).

it follows that:

$$(\exists \mathrm{D}_{(s,m)})\ \mathrm{D}x_{(s,m)}$$

where the value(s) of $\mathrm{D}_{(s,m)}$ are determined by the value taken by F. For instance, if it is true that either x could function as a set of brakes, or x is functioning as a set of brakes, then it is true that there is some disposition, in this case the disposition to decrease acceleration upon activation, which x bears. Why so? The first avenue I will explore relies heavily on the account of dispositional expressions developed in Chapter 1. The basic thought is that the truth conditions for dispositional ascriptions are themselves necessary but insufficient for the truth of functional ascriptions. Recall that on the account given dispositional truth conditions are as follows:

<u>Contingent Causal Analysis</u>

$\mathrm{D}x_{(s,m)}$ iff.

1. $\Diamond\ (\mathrm{S}x \rightarrow^{c} \mathrm{M}x)$.

Now consider the condition that was argued to hold on both senses of 'function-as', namely:

F2. $\Diamond\ (\mathrm{S}x \rightarrow^{c} e_f)$

It may be thought that the dispositional requirement is now obvious: If we take e_f to be a manifestation, then they are equivalent. Given that a function ascription, whether contingent or not, will be true only if F2 is satisfied, it follows that function-as (both contingent and actual) ascriptions entail dispositional ascriptions. But, it may be asked: Should we take e_f to denote manifestations? The answer: Unfortunately not. The manifestation of the brakes may not be the deceleration of the vehicle, but rather, perhaps, the pulling of a brake-cable. Nevertheless, in order to decelerate the vehicle, there must be *some* manifestation which, transitively or no, results in the deceleration of the vehicle. Put another way: For x to cause an F-event requires that it manifest, though the F-event itself need not be identical to that manifestation.

It may be objected that the support relies too heavily on the account of dispositions offered: The argument's force relies upon the veracity of our truth conditions for dispositions, of which some will not be content. Are there other ways, not laden by the contingent causal analysis, that we may support the requirement? The answer, I think, is 'yes'. In what remains, I give two independent reasons.

Reason 1 Suppose that Jones asserts that x is functioning as a set of brakes, and then in the same breath asserts that x does not have the disposition to decelerate vehicles. There is something strange about this assertion. If the thesis is correct, that much is

explained: it is a conceptual confusion to say that x functions as an F, and then to deny that x has the disposition to cause the relevant F-events.

Reason 2 Functional components bear many of the properties of dispositions. For one, they lie latent; for the brakes to be functioning, they need only manifest upon being activated, they need not do so at all times. Moreover, like dispositions, they may never manifest. The brakes may never be activated, if, say, the vehicle is owned by a reckless individual.

The remarks above, if true, establish that functional ascriptions entail dispositional ascriptions. But that is not the end of the story. For such ascriptions also involve claims concerning the integration into complexes. If something *could* function as an F, then it could be integrated into some complex where it will be in a position to manifest its disposition, if it *actually* functions as an F then it is a member of such a complex. Not all dispositions, then, entail functions, because the content of functional ascriptions is not exhausted by its dispositional content: to ascribe a function is to do more than just ascribe a disposition, such ascriptions also make claims about the modal status of the entity's integration into certain mereological complexes. Because not all dispositional ascriptions make such claims, not all dispositional ascriptions entail functional ascriptions. For example, an artificial organ may be disposed to pump, even if it could not function as a heart, due to its inability to enter into a mereological relation with circulatory systems. Or, to take another example, a glass's fragility is amongst its dispositions: It does not make sense to say that it functions as a fragile object. So, not all dispositions entail causal functions. The two claims, together, gives us **FED** and the negation of **DEF**.

Conclusion

In this chapter, I looked at the relationship that holds between functions and dispositions. There were two aims. They were as follows:

> [3.1] To disambiguate and outline four senses of the term 'function'.
>
> [3.2] To argue against the claim that bearing dispositions entails the occupation of a functional role, and to argue that the occupation of a function role entails the possession of certain dispositions.

First, I disambiguated the term 'function' into four central senses. Next, I looked at two arguments concerning the relationship between functions and dispositions, and showed that both rest on equivocation. Following that, I argued that functional ascriptions entail dispositional ascriptions.

Chapter 4

Functional Norms

Introduction

The ladder is for travelling up and down, so it should have rungs. Glue is for sticking parts together, and thus ought to have a firm hold. The heart is for blood pumping, and thus should pump the blood. These statements concern what entities *ought* to do given what they are *for*, that is, given their *purpose*. Put another way, these claims are teleo-normative.

In the last chapter, I disambiguated the term 'function'. Of notable import were two senses: The causal role sense, and the teleological sense. In this chapter, I am concerned with statements of the above sort, that involve the teleological sense and the normative. Its central task is the explication of the concept of a functional norm, a notion that will be employed later in the thesis. The aims are as follows:

[4.1] To outline the notion of a normative judgement.

[4.2] To distinguish two classes of normative judgement: teleological judgements and deontological judgements.

[4.3] To explicate the concept of a functional norm.

This chapter is composed of two sections. The first is composed of two subsections, the second of three. In section 1, I outline the notion of a normative judgement, and give a general sketch of what makes a judgement normative. In the second, I outline the notion of a teleo-normative judgement, and distinguish functional standards from deontological or 'rule' standards. I then distinguish norms from conventions, and outline the notion of a functional norm. Finally, I conclude.

4.1 Normativity

4.1.1 Normative judgements

Some judgements are normative, others are not. Most would agree with that claim. Nonetheless, there appears to be little agreement on precisely *which* judgements are normative. The presence of normativity is generally uncontested; the same cannot be said for its extension. The dispute is, however, down to ambiguity rather than substantive philosophical disagreement. The term 'normative', like so many of terms of philosophical art, is used in different ways by different authors. In what follows, I clarify what I mean by use of the term 'normative'.

But first, we should briefly clarify the subject of the predicate 'normative'. It is a *judgement*, rather than a *proposition* or *statement* that is said to be normative. Why so? Primarily because there is little agreement about their intentional objects. Whilst what follows will not *hinge* on a cognitivist conception of normative judgements, on which they take *propositional* objects like belief, it will likely appeal to those of a cognitivist ilk. That it does is, to my mind, a positive upshot: *prima facie*, cognitivism is a plausible position to hold. After all, normative judgements, as we shall see, are typically denoted by declaratives, i.e., statements of the form 'that such-and-such is so-and-so'. Declaratives typically express propositional content; a non-cognitivist must meet the challenge of explaining why we speak as though they do, if they do not. In that respect, I am in sympathy with Wedgwood, when he writes:

> [T]he burden of proof should be on someone who claims that the mental state normally expressed by the sincere utterance of a declarative sentence is something other than belief in the propositional content of that sentence. (Wedgwood, 2007, pp. 37-38)

In what follows I will speak of 'judgements', though at times I will slip between the terms 'statement' and 'proposition'. Dissatisfied non-cognitivists may replace as required.

Which judgements are normative? My starting point will be with the following view given by Fumerton. Consider:

> [A] judgment is normative if it is equivalent in meaning to a conjunction of statements which include paradigmatically normative terms. (Fumerton, 2001)

On this view, a judgement inherits its normativity from its constituents. We might put the claim another way: those judgements that are normative are the ones which are expressible with the use of normative terms. Put into the language of conceptual

philosophy, a normative judgement is one that bears, as part of its conceptual constituents, a normative concept. On reflection, however, Fumerton's criterion may be overpermissive. In particular, it is not obvious that a judgement contains a normative term is sufficient for its counting as normative. After all, judgements that *mention* paradigmatically normative terms, in contrast to one's that *use* them, include pragmatically normative terms, but are plausibly not properly regarded as normative. For example, suppose that we take some normative term 'Δ'. The judgement:

> Δ is a normative term.

bears a paradigmatically normative term. But the judgement is plausibly non-normative.[1] The problem cannot be avoided by retreat to conceptual phrasing; replace 'term' with 'concept' and the issue re-arises. Not all discussion of normativity is necessarily normative. If the criterion is to be worth its salt, plausibly we must restrict its extension to those judgements that use normative terms.

Whether this view is correct, I know not. Nonetheless, it will serve well as a starting point: We may identify normative judgements through the identification of paradigmatically normative terms. Which judgements are paradigmatically normative? According to Thomson:

> Normative propositions divide into evaluatives (such as propositions to the effect that such and such is good) and directives (such as propositions to the effect that so and so ought to do this or that). (Thomson, 2010, p. 713)

This passage makes claims about normative propositions, but its point may be recast in terms of judgements. Normative judgements fall into two classes: The first are what she calls the *directives*, which others prefer to call *prescriptives*, and the second are what she call *evaluatives*. I will take each class in turn.

Paradigmatic prescriptives are those judgements expressed by the term 'ought'. Consider:

> Our thinking is rich in what is often called normativity. We think that A ought to be kind to his little brother, that B ought to move his rook, and that C ought to get a hair cut. These are normative judgements. Intuitively, they differ starkly from such nonnormative judgements such as that A kicked his little brother, that B is playing chess, and that C has brown hair. (Thomson, 2008)

> As I shall use the term, these questions about what *ought* to be the case are "normative" questions — indeed they are the paradigmatic examples of normative questions. (Wedgwood, 2007)

[1] For a view on which the judgement may well be counted as normative, see (Gibbard, 2012). For interesting discussion, see (Williamson, forthcoming).

It should be noted, however, that strictly speaking not all statements that contain the term 'ought' are either prescriptives or directives. Prescriptions and directions are *instructions*, but not all judgements that include the term 'ought' are instructions. Thomson's examples serve well as illumination. If I say that C ought to get a hair cut, then I *might* be issuing an instruction, if the judgement is expressed directly to C, but I may not. If, for instance, I callously whisper that C ought to get a haircut in B's ear for the purposes of making fun of C, then I offer no instruction, but my judgement nonetheless contains the term 'ought'. Further examples of members of the prescriptive class include:

1a. The frosting ought to be vanilla.

1b. One ought to eat pudding after the main course.

1c. Doughnuts ought to have holes.

Some would not include non-prescriptive uses of the term 'ought' from the extension of 'normative'. Thus, 'Doughnuts ought to have holes' would not be counted as normative. But as I shall use the term, at least some non-prescriptive uses are to be included under the normative net. To avoid the impression that I count only the instructional variety, I shall refer to prescriptives as 'ought judgements'.

Before we turn to the second subclass of normative judgement, it should be noted that whilst 'ought' is a paradigmatically normative term, it is not the only term that feature in ought judgements, for two reasons. First, the term 'ought' as it features in ought judgements, expresses the same concept as the terms 'should', and 'must', at least in a wide range of contexts.[2] Second, the concept 'ought' comes as part and parcel of a cluster of concepts, all of which are, like certain quantificational concepts, plausibly interdefinable alongside negation. The other members of the cluster which I have in mind are the terms 'forbidden' and 'permissible'. Hence, whilst I will focus on 'ought judgements', we could perhaps just as easily focus on judgements of what is permissible or forbidden.

That deals with the first class of normative judgement: Ought judgements. I now turn to the second, what Thomson called *evaluatives*. Evaluative terms include 'good', 'bad', 'better', 'average', 'worse', and 'awful'. Evaluative judgements include:

2a. Vanilla frosting is good frosting.

2b. Chocolate pudding is better than other kinds of pudding.

2c. A meal with no pudding is a bad meal.

So far, I have now outlined two subclasses of normative judgements, *evaluative* judgements, and what I called *ought* judgements.

[2]Cf. (Broome, 2013, p.11).

4.1.2 Upholding to standards

To call teleological judgements *normative*, at least where we hold a permissive sense of 'teleological', is sometimes met with deserving scepticism. Moreover, to capture evaluatives and prescriptive under the same net may be thought an exercise in mixing specimens. The purpose of this section is to bruite a notion of normativity that will be sufficiently liberal to include functional norms as *bona fide* normative, and to unite both evaluative and prescriptive judgements under a single flag.

That may be thought a tough burden to bear. Is there anything that unifies normative terms, or do they form nothing more than a motley crew? On a fairly standard view, the answer is 'yes'; normative judgements are those that are *non-descriptive*. Put in Ramsey's terms, whilst descriptive statements—to use an unfortunately unilluminating term—*describe* the world, normative judgements *criticise* the world. Put another familiar way, descriptive judgements say what *is* the case, whereas normative judgements do not. Here are some examples:

> A point that is generally acknowledged is that the evaluative is part of the *normative*, where the normative is understood as concerning what we *ought* to do, in contrast with what *is* the case. (Tappolet & Rossi, 2015)
>
> We often think, not just about what *is* the case, but about what *ought to be* the case. (Wedgwood, 2007, p. 17)

Is this the best way we have to distinguish the normative from the non-normative? Not obviously. After all, there is a perfectly respectable sense in which some normative judgements say what 'is the case', and a perfectly respectable sense in which some descriptives do not. For instance:

> 3a. The yard is such that ball games are forbidden,
>
> 3b. There are good people and bad people, that's just the way the world is,

are evidently normative, but appear to make claims about 'how things are'. A similar point is made by Mark Bauer. He writes:

> It is common enough refrain when discussing the normative to talk about the "prescriptive" versus the "descriptive" as if this captured the difference between the normative and the non-normative respectively. The prescriptive is just, however, one species of the normative [...] [t]he normative includes norms that are not rules of prescriptions to act. "Fire exits ought to be unblocked" or "A hammer head's face ought to be $1\frac{1}{4}$ inches diameter" are standards that are not prescriptions [they are] descriptives of, not a prescription for, how things ought to be. The distinction between the

> prescriptive and the descriptive does not, then, suffice to distinguish the normative from the non-normative. (Bauer, 2009, pp. 245-246)

Further examples include those judgements that bear, as constituents, the truth-functional connectives. For instance, the statements:

4a. If you don't wake up early tomorrow, then you may not have pudding,

4b. Either the next ball is red, or I made a bad financial decision,

appear to be (in part at least) normative, despite, again, 'saying how things are'. It may be thought that, strictly speaking, the property of being normative should only be applied to truth-functional *relata*,[3] but that would be a mistake: normative terms may apply to complex judgements taken as a whole. After all, there is nothing senseless about the judgement:

5. It is bad that life presents further hurdles if you are female,

in fact, the judgement seems perfectly respectable. There *is* something bad about the conditional taken as a whole. Moreover, the judgements:

6a. The brownies may have been mouldy.

6b. Meat is not sweet.

do not concern what *is* the case, but rather how things *could be* and how things are *not* respectively.[4] It appears, then, that to merely distinguish the relevant judgements in terms of whether they 'say what is the case', is either trivially false, or requires extensive precisification. Is there another way that we may demarcate the non-normative from the normative, that makes no essential reference to whether the judgement in question 'says how things are'?

There is. The method of distinction I have in mind is similar to that given by Meckler. Consider:

> [A]ll normative expressions are, in one of their senses, descriptive. Sometimes they describe things in general, sometimes a narrower class of items often called "values". But they differ as normative expressions by virtue of their function, which is to lay down "conditions to be met". (Meckler, 1953, p. 579)

[3]If we were to be precise, we would disclude those *relata* that themselves contain truth-functional *relata*. We may think of these as, in some sense, 'atomic'.

[4]Cf. (von Wright, 1963, pp. 9-11.).

On this view, the distinction between a non-normative and a normative judgement is one of linguistic function. When one makes a normative judgement, one's judgement makes essential reference to what we may call *standards*, which may be thought of as *ideals* or *values*. More precisely; they involve 'holding an entity up' to some standard. In the case of ought judgements, one merely holds the entity up to that standard. If I judge that the pudding ought to be chocolate, then I hold the pudding up to some standard. I am judging that when puddings are 'as they should be', puddings are chocolate.

The notion of a standard, as I use it, is non-technical. I offer no principled way to individuate standards. Rather, standards must often be determined on a case-by-case basis, and by appeal to intuition. It will, however, be helpful to say a few words on the topic. First: Standards need not be spatiotemporal particulars, whether idealised or everyday. Nor are they mysterious platonic objects. The standard by which puddings are assessed is not some *ideal pudding*, the best-of-all-possible puddings, though in some cases consideration of idealised particulars may be fruitful in drawing out the standard by which entities are assessed. Moreover, to hold up to a standard does not require that the standard is explicit, known, or recognised. The notion is admittedly vague, but is serviceable and, I contend, well-understood.

The 'holding up to' relation is rather puzzling; it should not be confused with comparison. Comparison involves the determination of similarities, ought judgements are silent on the properties of the entity that is held up to a standard. If the pudding ought to be chocolate, then it ought to be chocolate irrespective of whether it is chocolate or vanilla. There is no comparison made to the standard, the standard is merely imposed. Nevertheless, comparison does enter essentially into evaluative judgements. When one makes an evaluative judgement, on the present view, one engages in a comparative enterprise: one compares the relevant phenomenon to a standard. But to compare against a standard is not just to compare. Rather, to compare against a standard requires also that a preference is given between the comparators. If I judge that the pudding is *good*, then I hold the pudding up to a standard, and invoke positive comparison between the standard and its comparator, i.e., the pudding. If I judge that the pudding is good because it is chocolatey, then I judge that it is comparable to the standard, where satisfaction of the standard may be achieved, in part at least, through being chocolatey.

To surmise: What makes a judgement normative, on the sense I have in mind, is that it 'holds up' an entity to a standard. To count as normative, a judgement need not compare against that standard, though if it does the judgement will be evaluative. In contrast, non-normative judgements do no such thing. If I judge that the ball is red, or that the fire is hot, I neither hold up the ball or the fire to any standard. But when I judge that the ball ought to be red, or that fire is good, then my judgement makes

essential reference to standards.

In this section, I outlined the notion of *normativity*. I outlined two varieties of normative judgement: ought-judgements and evaluatives. I then criticised the standard distinction between the normative and the descriptive, and outlined one way to understand the term 'normative', namely as denoting those judgements that hold up to standards, values, or ideals. Some may be discontent with that account, they may take normativity to require more than mere holding up to standards. Whilst there may be senses on which that holds, those senses are not my concern. When I say normative, I am speaking of those judgements that hold entities up to standards. Whilst that may not be a 'full-blooded' sense of the term, it is normative enough, for present purposes at least.

4.2 Teleological Normativity

Normative judgements are judgements that hold entities up to standards. 'Teleological normativity' denotes a class of normative judgements. We may call these 'teleo-normative judgements'. The conception of normativity sketched above is sufficiently liberal to include those judgements that may be called teleo-normative. What *are* those judgements; what makes them teleo-normative? On the account outlined in what follows, a judgement *j* is teleo-normative just in case *j* holds up an entity to a standard imposed by that entity's design origins. There are three subsections: in the first, I introduce the notion of teleological normativity, in the second I outline the distinction between norms and conventions, and in the third I outline the notion of a functional norm.

4.2.1 Teleological standards

Teleological normativity is any form of normativity, where the normativity is determined by some purpose, aim, or goal. An example of a teleological judgement may be found in the following passage:

> Why should the ladder you're building have sides? Because if it doesn't have sides there will be no place to put the rungs. Why should the ladder you're building have rungs? Because if it doesn't have rungs it won't be able to help anyone climb up and down. Why should the ladder you're building be able to help someone climb up and down? It just should — that's what a ladder is for. If you dont understand why a ladder should be able to help someone climb up and down, then you dont understand what a ladder is! Here we appeal to the *function* of ladders to put a stop to a regress of "Why?" questions. (Silverstein, 2016, p. 214)

Consider the last of these claims, that the ladder should be able to help someone climb up and down. The climbing up and down is the purpose of the ladder. That is what it is to *be* a ladder. It is in virtue of its having that purpose that an ought-judgement holds. The ladder ought to function as an object that allows one to climb up and down, because that is the ladder's purpose. This is a teleo-normative judgement. Because purposes are determined by an entity's origins, so too are the teleo-normative standards to which such entity's are upheld.

According to some, teleo-normative judgements are analytic truths.[5] If one grasps the concept of ladder, then one can, from the meaning of that term alone, come to know that a ladder should allow one to climb up and down. Whether these are genuinely analytic, however, is hard to tell. Determining analyticity is a tricky task, outside the prototypical cases. Others, including Tyler Burge (2010, p. 312), have claimed not that such truths are analytic, but rather that they are knowable *a priori*. If they are analytic, that would be explained. Analytic truths are plausibly all knowable *a priori*, even if not all truths knowable *a priori* are analytic. But like analyticity, a prioricity is difficult to determine. When Burge uses the term, he means '*warranted, but not warranted through sensory material or perception [...] typically warranted through understanding or reason*.' But if Williamson (2013) is correct, to describe a truth as knowable *a priori* is empty—the notion is epistemologically insignificant. Of course, all of this hinges on a broadly cognitivist conception of normative judgements. I do not know whether such judgements are analytic, nor whether they are *a priori*. But they do share close residence. Entities that are classified in terms of their purposes evidently bear their purposes in virtue of that classification. We can accept that, whilst remaining silent on whether that entails analyticity, or knowability by the capacities of reason alone.

When a normative judgement *j* is not teleo-normative, *j* is as I shall say 'rule-normative'. To call a judgement rule-normative is not to classify it as imposed by some individual or institution. Rather, 'rule-normative' is a functional category: Rule normative judgements play a cognitive functional role that teleo-normative judgements cannot. In particular, rule normative judgements can play what may be thought of as a *motivational* role: they can play a special role in items of practical reasoning. If I judge that I ought to be good, then (oddities notwithstanding) I will try to be good. My judging can *motivate* me to act in accord with the judgement's content: it is in that sense that they are 'rules'. In contrast, teleo-normative judgements *cannot* play that functional role. This is, I take it, why arguments from purpose are thought to fail. To judge that one has a purpose is not motivational in this sense. Rather, to feel the normative force of the content, one must first judge *that one ought to act in accord with one's purpose*. Interestingly enough, John Barth put this point excellently, in his short story entitled '*Night-Sea Journey*', in which the narrator is an angst-ridden sperm.

[5]See, for instance, (Foot & Montefiore, 1961) and (Foot, 2001).

Consider:

> Arguments from function and design don't impress me: granted that we can and do swim, that in a manner of speaking our long tails and streamlined heads are 'meant for' swimming; it by no means follows – for me, at least – that we should swim, or otherwise endeavour to 'fulfil our destiny.' (Barth, 1968)

Now, I must insist that the following point be clarified: not all judgements pertaining to purpose may be properly regarded as teleo-normative, at least insofar as I use the term. Rather, some are rule standards. Consider, for example, the former two of Silverstein's normative judgements—that the ladder should have sides, and that it should have rungs. These are not teleological through-and-through. There is nothing necessary about having sides or rungs—some rope ladders have neither. Such ladders may be *worse* than those that have rungs and sides, *vis-á-vis* the ladder's serving its purpose. But this involves a distinct standard, one not imposed by design origins. Rather, such judgements hold the ladder up to an ideal—the ideal of best ladder-performance. That ideal is not determined by the ladder's design origin. A ladder should allow one to climb up and down, that is its *purpose*. Its purpose is not to allow one to climb up and down as best it could, despite that being eminently desirable. Similarly, the heart aims to pump the blood; a given heart could do so better. But its doing so better sets it against a different kind of ideal—the best possible satisfaction of its aim. At best, this tells us that entities can serve their purposes better than they in fact do. True enough, but these are not distinctively functional evaluations—they are functional in the sense that they take *derive aims from purposes*, not functional in the sense that they are individuated by an entity's history. That is to say, to function *properly* is for it to function as it is supposed to. To function *better* is not to function properly. The heart, when functioning as it is supposed to, does not pump as the best of all possible hearts might pump.

We might call judgements of the former kind *quasi*-teleo-normative judgements. Quasi and *bona fide* teleo-normative judgements are sometimes run together. Some authors recognise the distinction, but take them both to be genuinely teleo-normative. Consider Plantinga:

> There is a sort of ambiguity in the notion of working properly. On the one hand, a thing works properly when it works in accordance with its design plan, when it works just the way in which it was designed to work. My radio works properly when there is nothing wrong with it and it works just as its designer designed it to. But what shall we say when it works as it was designed to, all right, but has a very poor design and won't receive stations

> more than 500 yards away? Then it does not work very well, despite its functioning precisely in accord with its design plan. [...] Again, the same distinction holds in the case of an animal or other organism. Perhaps you think the human knee is poorly designed; then you may think that a knee functioning in accord with its design plan is nonetheless not functioning well, even when the way it works does not deviate from its design plan. (Plantinga, 1993, p. 27)

If I am correct, there is no ambiguity here. Working 'properly' is a matter of working in accord with the design plan that explains one's origins; to work 'well' in this other sense has nothing to do with the notion of functioning properly, aside from its role in determining the ideal such judgements employ. Others are less careful to distinguish the two. Burge, for instance, allows his notion of a 'natural norm' to include what I have called the 'quasi' variety. He is, of course, aware that such norms are not, in his own words 'naturalistically reducible'.[6]

Teleo-normativity is not constrained to ought-judgements—there is a teleological evaluative dimension also. We may say that x is functioning *better* or *more properly* than y. Simpler: x may function poorly, y may function excellently, or x and y may function just as well. Such evaluations, like evaluations of all kinds, differ only insofar as they compare the functioning of an entity to its standards. To say that the heart ought to pump the blood is to make explicit the design standard of the heart—to say that a heart functions better is to compare it against that standard. Such standards may be individuated in a coarse-grained, or a fine-grained fashion. We might say that the heart is supposed to pump the blood, but we would do better to say that it is supposed to pump the blood via, say, making squeezing motions. Consider Price:

> The heart moves the blood around the body *by* making squeezing motions; the wing provides lift *by* flapping. The flapping of earlier wings contributed to the workings of their governors, because it is just by flapping that they provided lift for flight. So the fact that this particular pair of wings is present can certainly be explained by the fact that earlier wings flapped [...] it will be perfectly legitimate to say the function of wings is to flap.[7] (Price, 2001, p. 52)

Question: how low do we go? Answer: ask mother nature, or the artefact's designer. It depends upon what was selected or intentionally designed, and what accidentally occurred. Of course, we might also give a more fine-grained individuation by giving a more fine-grained individuation of the bearer of the purpose. Hearts are supposed

[6]Cf. (Burge, 2010).

[7]There are, no doubt, analogies here with the case of intentional action. We say that Jones murdered the town *by* poisoning the water supply, and so on. See (Hornsby, 1980).

to pump the blood—the hearts of certain species are supposed to pump the blood in specific ways. How fish hearts are supposed pump the blood, for example, differs from how human hearts are supposed to do so. The types fish-heart and human-heart, then, are united in their coarse-grained blood pumping purpose, and even in their more fine-grained purposes to pump the blood by making squeezing motions. But they differ in their even finer grained purposes.

Recall that there are three classes of purpose: natural, artificial, and intentional. It would seem to follow that there are three corresponding classes of teleological normativity. Examples of natural teleo-normative judgements include:

N1. The kidneys ought to filter the blood.

N2. The heart ought to pump when the sinus nodes activates.

whilst examples of artificial teleo-normative judgements include:

A1. The big red button ought to release a nuclear weapon.

A2. The brake lever ought to pull the brake wire once pressed.

Now, it may be argued that whilst there are three classes of purpose, there are not three classes of teleological normativity. The proponent may ask that we consider:

I1. Harry ought to have impressed Sally with his jokes.

I2. Archie ought to have hit the target.

It seems evident that claims I1-I2 do not follow from the facts that Harry intended to impress Sally and that Archie intended to hit the target respectively. This is rather puzzling. But the puzzlement may be dispelled by noticing that the disanalogy arises from incorrect individuation of the subject of the relevant obligations. The relevant obligations do not hold on agents, but on their *actions*. For consider:

I1*. Harry's joke ought to have impressed Sally.

I2*. Archie's shot ought to have hit the target.

I1* and I2* now seem less strained. And that should be no surprise: If the origins account is correct then purposes are determined etiologically, and we are concerned not with the origin of the individual, but with the origin of the agent's action. But in any case, I will not be primarily interested in intentional purposes, so we may cast them to one side.

Teleo-normative judgements play a central role in the diagnosis of fault. This will be important, so permit me to end this subsection by tangentially spilling some ink. A vivid example of what I have in mind may be found in the medical sciences, where

the identification of malfunction serves a central role in diagnosis. Diagnosis of a heart condition often requires a standard by which a given heart may be assessed; the working of past tokens of the same type that explain the persistence of the trait. Similar points apply to artefacts; to know why an engine has stopped running, one often enough needs to know how an engine functions, when it functions properly. Moreover, the identification of artefact purposes has other heuristic benefits. I can typically come to know what something *does* by knowing what it was *designed* to do, for artefacts typically do what they are supposed to. Again, purpose is no guarantee of the occupation of causal roles, but it does well enough as an indicator, enough of the time.

The act of holding entities up to teleo-normative standards finds value not only in the identification of malfunction, but also in the identification of *improper* function. When something improperly functions, it occupies what we may call its proper 'coarse-grained' functional role, but not what we may call its 'fine-grained' functional role. Consider a heart so-diseased that it cannot circulate the blood. The heart would be malfunctioning. But consider now a heart also diseased that it does not occupy its selected causal role but, through some sort of strange fluke, is nonetheless able to circulate the blood. This heart would not *mal*function, it would *improperly* function. The distinction is important: the probability that x requires medical treatment conditional on its malfunctioning is higher than its functioning improperly. So long as it does what it is 'supposed to'—in a coarse-grained way—often enough it will do well enough if left alone. If it ain't broke, so the saying goes.

So far in this section, I have outlined the notion of a teleo-normative judgement. Along the way, I contrasted teleo-normative judgements with *rule* normative judgements. Below, in what remains, I will outline the notion of a *functional norm*.

4.2.2 Norms and conventions

A broader recap is due. So far in this chapter, I have outlined the notion of *normativity*, and I have delineated two class of normative judgements: *rule* based normative judgements and *teleological* judgements. In what remains of this chapter, I will outline a notion that will play an important role in the argument given in the chapter that follows: that of a *functional norm*.

Norms are widely discussed in the philosophical literature. Unfortunately, as is too oft, discussion typically proceeds in the absence of a sufficiently clear account of what norms are supposed to be. Much discussion on norms is, unsurprisingly, intimately connected to rule-normative judgements. Here are some examples:

> Now, a norm governing assertion is the rule that needs to be followed in order to offer a proper assertion, much like we have rules that need to be

> followed in order to properly perform in certain competitive sports. Of course [the norm] can be, and frequently is, violated. I may, for instance, assert that pollution is killing our local wildlife when I only suspect that this is the case [...] where an assertion is made in the absence of the [norm satisfaction] the asserter in question is appropriately subject to criticism. (Lackey, 2007, p.594)

> [I]t is natural to suppose that some norms are more intimately connected to the nature of asserting than any norm is to the nature of jumping. One might suppose, for example, that someone who knowingly asserts a falsehood has thereby broken a rule of assertion, much as if he had broken a rule of a game; he has cheated. On this view, the speech act, like a game and unlike the act of jumping, is constituted by rules. (Williamson, 1996, p. 489)

Norms are, on this view, a special sort of rule. The term 'norm', as is not supposed to be synonymous with 'rule'. Rather norms are supposed to be in some sense, and as Williamson notes in the passage above, *constitutive rules*. If so, it may be reasonably thought that the notion of a function norm contains a manifest contradiction. Nonetheless, it is not uncommon for authors to speak of 'functional norms'. We saw that above, where Burge spoke of what he calls 'natural norms', but a more explicit example may be found in the following passage from Peter Graham:

> There are norms that do not require the capacity to represent, think, internalize, or subscribe to a norm. There are norms that are neither prescriptions nor guides. Functional norms are norms in this broader sense. [...] Such norms need not prescribe or guide. No norm tells the heart what to do. The heart does not look up or represent any norm to guide its activity. Functional norms are a broader kind than prescriptive or guiding norms. They do not depend on the aims or intentions of individuals, on being represented or being endorsed. (Graham, 2014, pp. 22-23)

If the notion of a functional norm is to be of service, we need first to cast its extensional net wider. That is the task to which I now turn. What, then, is a norm? As I understand the term, a norm is a *constitutive standard*. The account is more general, for not all standards are rules.

What is it to say that some standard is constitutive? As I will interpret the notion, constitution is not a mereological notion, but rather denotes *essentiality*. At first blush, essences are a class of property. Whether a property is essential to some entity x is determined by the nature of x. x's essential properties are those properties which x has simply in virtue of being the sort of subject that x is. Put another way: Fx holds *qua* x's x-ness. When a property is non-essential, it is said to be *accidental*.

Now, even more than the term 'normativity', the extension of the term 'essence' is highly contested. Purported examples of essential properties would include of Socrates being a man, or of a particular organism being a Silverback Gorilla. On a fairly standard view, essential and accidental properties are distinguished along modal lines. Essential properties are ones that objects bear *necessarily*, accidental properties are ones that objects bear *contingently*:

Essential F is essential to x iff. Fx, and necessarily, x exists only if Fx.

Accidental F is accidental to x iff. Fx, and possibly, x exists and not Fx.

Whilst the left to right direction of *Essence* is fairly uncontroversial, it is highly dubious that the right to left direction holds, largely due to the work of Kit Fine. The basic problem is that not all properties that hold as a matter of necessity are ones which hold in virtue of an entity being the very entity that it is in the relevant sense. The following passage from Fine puts the problem in clear terms:

> Consider Socrates and the singleton containing him. Now although it is plausible to suppose that the singleton essentially contains the man, it is not plausible to suppose that the man essentially belongs to the singleton. There is nothing in the nature of Socrates which demands that there be any sets, let alone the one that contains him. However, the standard accounts of essence in terms of necessity are unable to account for this asymmetry. For under such an account, the singleton essentially containing Socrates will consist in something like its being necessarily the case that the set contains Socrates if the set exists. But if this is true, then it will also be necessarily the case that Socrates belongs to the set if the man exists.[8] (Fine, 1995, p. 241)

I don't intend to delve into this metaphysical swamp here. Nonetheless, for present purposes we can largely jettison these problematic cases, for the relevant essences with which I will be concerned do not hold necessarily in the trivial way that set membership does. We can, to borrow Davidson's terminology,[9] afford 'Spinozistic extravagance'. What would matter is the failure to include *bona fide* norms, that the conditions count too much will not yield substantive problems.

Some standards are essential, in the following sense: it is constitutive of certain entities that they are compared to a certain standard. That is, they are held up to that standard *in virtue* of being the very entity that they are. In contrast, some entities are compared to certain standards not *qua* the entity that they are. Where a standard S

[8]See also (Fine, 1994).
[9]See (Davidson, 1970, p. 212).

holds on ϕ, but only accidentally, the standard is not a norm; with respect to S, ϕ is governed by a *convention*. To classify a standard as a convention is not to denounce that standard as trivial or unimportant. It does, in some sense, entail that the standard is arbitrary. But arbitrariness need not entail triviality. It is a convention on question asking at talks that one is polite and constructive; this is no trivial standard—it is deeply important. Rather, to classify a standard as a convention is to say that it does not govern the activity necessarily, or in virtue of its being the kind of activity that it is. An example will help. Consider promising. Suppose that, in some case c, there are two standards that apply to the institution of promising. They are as follows:

Word-Keeping If S promises S^* to ϕ, then S ought to ϕ.

Spit-Shake If S promises S^* to ϕ, then S ought to spit-handshake with S^*.

The former is essential, the latter accidental. Thus, the former is a norm on promising, the latter is a convention. Why so? Because it is *constitutive* of promising that one is held up to *Word-Keeping*. But is it *not* constitutive of promising that one is held up to *Spit-Shake*. Put another way: if an activity is not governed by *Word-Keeping*, that activity is not one of promising. But if an activity is not governed by *Spit-Shake*, then it could be the activity of promising. In explicit terms:

<u>Norm</u>

v is a norm on ϕ-ing if and only if:

N1. $\Box\, \forall x$ (x ϕ-s only if v is a standard on ϕ-ing).

<u>Convention</u>

v is a convention on ϕ-ing if and only if:

C1. $\Diamond\, \exists x$ (x ϕ-s and v is not a standard on ϕ-ing),

C2. v is a standard on ϕ-ing.

The definition of a norm differs from that of a convention, in that the latter imposes two conditions of satisfaction, whilst the former imposes only one. But that is for simplicity: given the axiom of reflexivity, which holds virtually all modal systems, the norm analogue of C2 will straightforwardly follow.[10]

It is worth noting that the distinction so sketched, in contrast with that found in the quotations given at the start of this section, make no essential reference to the notion of a rule. As such, they are applicable to non-rule based normativity. Pleasingly, they may be successfully applied to teleological norms, as I shall now show.

[10]Here is the axiom: **T**: $\Box\, p \rightarrow p$. See (Zeman, 1973) for some systems in which **T** does not hold.

4.2.3 Functional norms

Whenever x has either a natural or an artificial purpose, there are functional norms that hold on x. We will start with artefacts. Suppose that Dexter designs some artefact a, with the intention that it should mow the lawn once the grass is long. We might call this radical invention the *auto-mower*. Now, to count as an auto-mower, it is neither necessary nor sufficient, recall, that an object *in fact* mows the lawn once the grass is long. Something could *function as* an auto-mower, without thereby *being* an auto-mower. Rather, for x to be an auto-mower is for x to be such that x is *supposed* to mow the lawn. Membership of the kind 'auto-mower' is teleological, not causal. Now, because auto-mowers share the same purpose—that of cutting the lawn when the grass is long—there are functional norms that hold necessarily on auto-mowers. Moreover, anything with that purpose counts as an auto-mower.

Not all of the standards imposed on auto-mowers are essential. Auto-mowers may, for instance, be held up against a standard of durability and frugality; it may be that in Dexter's society, auto-mowers ought to be durable and affordable. But these are not, unlike its functional norms, essential to auto-mowers; in a society sufficiently affluent and fond of single-use items, it may be that auto-mowers should be expensive and dispensable.

Similar points apply to natural purposes. If something is a heart, then it need not function as a heart. Nor does anything that functions as a heart count as a member of the kind. Nonetheless, they all, in virtue of their being hearts, are governed by functional norms. Anything that counts as a heart, in virtue of its being a heart, should pump the blood. I don't, notice, say that bearing such a purpose is sufficient. I say only that all hearts share a common purpose, not that all entities that share that purpose are hearts. Again, it may be that a society upholds biological traits to differing standards: Hair has a purpose, but society holds one's hairstyle up to the standards of fashion. That is not a functional norm, nor is it a norm in any sense; it is a non-functional convention.

To some extent, the remarks above are in keeping with the following claim from Mumford, when he writes:

> Some objects are the objects they are in virtue of the function they perform. One example of such a thing is a thermostat. Something is a thermostat in virtue of it having the function of triggering a switch when a pre-calibrated temperature threshold is passed either from a higher or lower temperature. Any object that has this functional role is a thermostat no matter what its other properties are that realize this functional role. (Mumford, 1998, p. 197)

But Mumford, as before, is using 'function' in its causal sense, and that cannot be

right. Something isn't a thermostat merely because it could function as a thermostat, or even because it *is* functioning as a thermostat. Something is a thermostat when it was designed to function in a certain way, not when it in fact does. A broken thermostat is still a thermostat.

I anticipate the following reply: For the argument to work, it must be that the relevant purposes are essential to the objects in question. Now, in the last chapter it was argued that purposes are *extrinsic* properties. If purposes are extrinsic properties surely, it may be contended, they are not essential? In fact, it seems eminently plausible, at least *prima facie*, that purposes could not be essential. For consider the case discussed: A given heart could have developed in a different species, and in that species it could have served a different purpose – i.e., it may have been a waste disposal organ. Now, if purposes are extrinsic, then they are not essential. If purposes are not essential, then the functional norms that hold in virtue of something's having a certain purpose cannot be essential.

The argument is powerful, but there are several responses available. The first would be to deny the inference from '*F* is extrinsic' to '*F* is inessential'. As Bird notes:

> [T]hese reasons need not be taken to be decisive. Perhaps the membership of some natural kinds is not an intrinsic matter. It may be that the belief that it is intrinsic stems from the view that natural kinds have essences, and that essential properties should be intrinsic. But as we have seen [...] the essential properties of individuals need not be intrinsic; perhaps they need not be for kinds either. (Bird, 2009a)

The examples to which Bird refers are those famously raised by Kripke: the necessity of origins, and of natural kinds.[11]

Let us consider the former. Origins are extrinsic properties—after all, they are relational. Nonetheless: I am not tempted by that response, for I take the argument for origins essentialism to be weak. Certainly, it goes against the grain of metaphysical intuition: it seems perfectly plausible to suppose that *this very* organ could have derived from a distinct origin. Of course, metaphysical intuition is no infallible source of knowledge. Nonetheless, an alternative response would be desirable: is one available?

Fortunately, one is, though is does require some modification to the notion of an essence. In particular, we might adopt a form of *property-essentialism*, rather than what we may call *subject-essentialism*. Authors who defend something like the former include Okasha (2002), Elder (2007), and LaPorte (2004). Here is the distinction. Whilst subject essentialism concerns those properties that an entity bears in virtue of being the very individual that it is, property essentialism concerns those entities a property

[11] For a defence of the former, see McGinn (1976), Forbes (1985, 1986), Noonan (1979, 1981). For the latter, see Kripke (1971, 1980) and Putnam (1975).

applies to in virtue of its being the very property that it is. An example may help. Suppose that x is a green triangle. Amongst x's properties are being green, and being a three-sided polygon. Now, whilst it is *essential* to x's being a triangle that it is a three-sided polygon—it is a three sided polygon *qua* being a triangle, it is *accidental* to x's being a triangle that it is green. It is not green *qua* being a triangle. A triangle is triangular in virtue of its shape, not its colour. Similarly, suppose we take the property 'being red'. It is an essential feature of that property that it is a property of spatially extended surfaces. If F is a property, and F can apply to events, say, then F is not the property of being red. It is an accidental, however, that F applies to chickens, or to Rauschenberg's paintings.

How does this help? Well, whilst it may be that the subject of the property 'being a heart' could have not been a heart, perhaps in virtue of having different origins, it *is* an essential feature of the property 'being a heart' that it applies to entities with a certain kind of origin. After all, for x to count as a heart requires that it has been selected to serve heart functions. So, it is essential to the kind 'heart' that it is governed by functional norms; it need not be essential to any particular organ that it is a heart. If this is the right view to take, then functional norms are not norms that must hold necessarily on *individuals*, but rather of any individual that, as a matter of fact, bears a certain kind of property, in this case a relational property—a certain origin. Whilst a heart could have been other than a heart, conditional on an entity's being a heart, it is governed by functional norms as a matter of necessity. In that sense, hearts are governed by functional norms.

Call this notion a *property* norm. In explicit terms:

<u>Property Norm</u>

ν is a property norm on F if and only if:

N1*. $\Box\ \forall x$ (Fx only if ν is a standard on x).

We could also define the notion of a *property convention*, as follows:

<u>Property Convention</u>

ν is a property convention on F if and only if:

C1*. $\Diamond\ \exists x$ (Fx and ν is not a standard on x),

C2*. ν is a standard on x.

Thus, necessarily, something is a heart just in case it ought to pump the blood. But something is not a heart in virtue, say, of the fact that it ought to make a beating sound, even if there is some sense in which that holds. Whilst it may be that not all property

norms are functional norms, all functional norms are property norms. Something is a functional norm just in case it is both a property norm, *and* is set in the relevant way by the entity's origins. That concludes my outline of the notion of a functional norm.

Conclusion

In this chapter, I had the following aims:

> [4.1] To outline the notion of a normative judgement.
>
> [4.2] To distinguish two classes of normative judgement: teleological judgements and deontological judgements.
>
> [4.3] To explicate the concept of a functional norm.

which are now satisfied. In the first section, I outlined the notion of a normative judgement. I argued against the standard distinction between descriptive and normative judgements, and argued that a normative judgement is one that involves comparison with, or the bringing to light of a standard. In the second section, I gave several examples of teleo-normative judgements, and distinguished them from deontological judgements. In the third section, I explicated the notion of a *functional norm*, which I took to be a kind of *constitutive* standard, which derives from an entity's purpose.

Part II

Doxastic Dispositions

Overt intelligent performances are not clues to the workings of minds; they *are* those workings.

Gilbert Ryle
The Concept of Mind

Introduction to Part II

I turn now to the metaphysics of belief. My primary concern will be the *individuation* of belief content pairs. The term 'individuation' is often used rather ambiguously in philosophy. Here is my usage. According to Lowe, there are two kinds of individuation: Epistemic and metaphysical.[12] Epistemic individuation concerns the way in which we 'carve up' the world. In that sense, individuation is a kind of cognitive ability. In contrast, metaphysical individuation concerns the similarity relations that hold between entities within a given domain. That which individuates x and y, is that which makes x and y identical or distinct, whichever they may be. In what follows, I am concerned with metaphysical, not epistemic individuation.

Now, the question 'what individuates belief-content pairs'? may be broken down into several components. On the one hand, we may ask what individuates *attitudes*, i.e., what makes one attitude a belief as opposed to, say, a desire. I won't be concerned with that question. On the other hand, we may ask what individuates *contents* of distinct beliefs. Suppose I believe that apples are sweet, and you believe that apples are sour. Our beliefs would *differ in content*. Suppose now that I believe apples are sweet, and so too do you (great minds, I dare say!) We would have *sameness of content*. My question is: Given S believes that p and that S^* believes that q, what makes S's belief *a belief that p* as opposed to *a belief that q* and *vice versa*.

What follows will be an attempt to individuate belief-content pairs in purely dispositional terms. There are at least two major hurdles that such an account must face: the *holistic* character of mental dispositions, which makes a pure dispositional individuation appear to succumb to unavoidable circularity, the second is the content externalism in the philosophy of mind, on which the content of one's beliefs are individuated, in part at least, by extrinsic factors. Those who favour the first line of thought may reject a dispositionalist account of belief in favour of a *functionalist* account. Those who favour the second may eschew both positions outright. This part aims to make two tentative conclusions. The first is that the holistic challenge can be accommodated by a dispositional conception of belief. The second is that content externalism can be accommodated by an appeal to *dispositional ideals*.

There are three chapters. In chapter five I outline the attitude of belief and contemporary theorising on the relationship between belief and dispositions. In chapter six, I argue that a dispositional account of mind is better than a functionalist account. In chapter seven, I outline the argument for content externalism, and argue that beliefs are individuated in terms of dispositional ideals.

[12](Lowe, 2003).

Chapter 5

The Attitude of Belief

Introduction

Consider:

1. Joey believes Jane is beautiful.
2. Jane believes Joey is a nice guy.

These statements tell us what Joey and Jane *believe*. They are ascriptions of *belief*. It follows that they are ascriptions that contain, as part of their conceptual content, the concept of belief. If Davidson (1982) is correct, the concept of belief is the mark of a rational animal. Only those who grasp the concept may be said to believe; only those who may be said to believe are candidates for rational thought. That the concept of belief is the precursor to being an agent that believes is a strong claim, no doubt. But even if false, the centrality of the concept should not be shirked. For whether you require the concept of belief to believe, no doubt to count as an agent one must have beliefs. To be an agent is, in part at least, to believe. This chapter introduces the attitude, and contemporary theorising on the dispositional aspect of belief. There are three aims, which are as follows:

[5.1] To outline the attitude of belief.

[5.2] To defend the propositional requirement on belief.

[5.3] To outline contemporary theorising on the relationship between dispositions and beliefs.

Here is the plan. There are two sections. In the first section, I outline and defend what I call the 'propositional requirement on belief' against a recent argument from Robert Audi (2013). This requirement states that S believes x only if x is a proposition. In the second section, I outline a preliminary dispositional account of belief, which

I call *Simple Behaviourism*. Finally, I offer some defence of the view, and reject three purported differences between dispositions and beliefs given by Armstrong (1973).

5.1 The Propositional Requirement

Belief is one of the propositional attitudes. 'Propositional attitude' is a Russellian phrase. Attitudes are expressible as dyadic relations which hold between *agents*, and what are called *contents*. Both talk of 'attitudes' and 'contents' is metaphorical and deliberately vague, and thus should not be taken at face value. Attitudes are the kind of 'stance', or 'posture' that is taken towards a content.[1] Other examples include desiring, intending, deciding, and fearing. Contents are what an attitude is directed towards or 'about'. This 'aboutness' is what Brentano called intentionality.[2] We call what the attitude is about its intentional object. On a fairly standard (though in no way uncontested) view, attitudes may be propositional or they may be qualitative. These roughly coincide with Kant's distinction between *concepts* and *percepts*.[3]

Qualitative contents are, roughly and uninformatively, those qualitative properties of perceptual experiences, such as the 'red-ness' one undergoes when seeing a ripe tomato, or the burning sensation one feels when touching a hot plate. Belief does not (at least not characteristically) take perceptual objects. Belief is a propositional attitude. To say that belief is a propositional attitude is to say that beliefs take propositions as their intentional objects. All propositions are expressible in declarative terms, i.e., by statements of the form 'that such-and-such is so-and-so'.

What are propositions? A vexed question. On an early Russellian view, propositions are constituted from *facts*, what Armstrong calls *states of affairs*.[4] Alternatively, propositions may be thought of as abstract objects. Plausibly, the term 'proposition' is ambiguous, taking distinct meanings in distinct theoretical contexts. For present purposes, we may bruit a rough view: Propositions are the expressed 'meaning' of declaratives. Two declaratives express the same proposition just in case they are identical in meaning. Two declaratives express different propositions just in case they differ in meaning.

Unfortunately the view so bruited is ambiguous, for our chosen *definiens* harbours ambiguity of its own: We may individuate 'meaning' in either a *coarse* or a *fine*-grained fashion. Where meaning is individuated in a fine or coarse-grained fashion, we have respective coarse-grained and fine-grained individuation of propositions. The distinction will become important, so I will say a few words about it.

[1] For use of these latter terms, see (McDowell, 1994).
[2] (Brentano, 1874).
[3] (Kant & Muller, 1896).
[4] See (Russell, 1919), and (Armstrong, 1997).

The distinction between fine and coarse propositional individuation may be explained with the use of the notion of a *concept*. Philosophical concepts are not the concepts of psychology; they are not 'in the head'. Rather, concepts are what Frege called a 'presentational mode' or 'sense'.[5] We may call this, following contemporary idiom, conceptual content. For present purposes, we may take a concept to be the *constituent* of a (coarse-grained) proposition, though again talk of constitution may be mere metaphor.

Examples will clarify. Suppose Smith utters:

3a. Hesperus is bright.

whilst Jones utters:

3b. Alpha Centauri is bright.

What Jones and Smith mean is different. Smith means that the planet Venus is bright, Jones means that the closest star system to Earth is bright. Their meaning differs in both senses. 3a and 3b differ in their coarse-grained propositional content. And it follows that they differ in their fine-grained content too: Coarse-grained difference in meaning implies fine-grained difference in meaning.

But the converse fails: Difference in fine-grained propositional content does not imply difference in coarse-grained propositional content. For example, the statements:

4a. Hesperus is bright.

4b. Phosphorus is bright.

express the same coarse-grained proposition, but are constituted from non-identical concepts, i.e., the concepts of Hesperus and Phosphorus respectively. So 4a and 4b can be said to express distinct fine-grained concepts.

It is important, especially for what follows, that reference and sense are kept distinct, for its conflation masks epistemologically important aspects of the propositional attitudes. One may know Venus under the concept of Hesperus, but not under the concept of Phosphorus. Hence, ascriptions of identity between referents known under distinct conceptual modes may be informative. When I speak of propositions, I will (unless stated otherwise) be speaking of the coarse-grained variety. To distinguish the two, I will denote propositions individuated in a coarse-grained way with single sets of angle brackets, i.e., <A course-grained proposition is denoted>, and fine-grained propositions (that is, propositions specified by their conceptual content), with double sets of angle brackets, i.e., <<A fine-grained proposition is denoted>>.

[5](Frege 1892, 1942).

If standard theorising on the attitude of belief is correct, then conceptualisation under presentational modes is no contingent affair. As a matter of necessity, if belief takes an object, then the object is so conceptualised. Call this the *conceptual requirement on belief*:

CRB Necessarily, if S believes x, then x is mediated by a conceptual mode.

The requirement, as formulated, is silent with respect to the propositional requirement to be defended. If beliefs take non-propositional objects, they may still be conceptualised as a matter of necessity. I take it that **CRB** is plausible enough. Given it won't be the focus of what follows, I will assume it holds.

Because the intentional objects of beliefs are conceptualised under presentational modes, believing may, to use Carnap's (1956) terms, be ascribed under an *intensional* context. That belief is intensional (*s*) should not be confused with the claim that it takes intentional (*t*) objects. Intentionality is a property of the attitudes—their 'aboutness'. Intensionality, in contrast, is a property not of the attitudes themselves, but of the logical contexts in which they sometimes appear. For Davidson, intentionality is one of the defining characteristics of the attitudes. He wrote:

> We may call those verbs mental that express propositional attitudes like believing, intending, desiring, hoping, knowing, perceiving, noticing, remembering, and so on. Such verbs are characterized by the fact that they sometimes feature in sentences with subjects that refer to persons, and are completed by embedded sentences in which the usual rules of substitution appear to break down. (Davidson, 1978)

When a context is intensional, it is non-extensional.[6] In alternative Quinean terms, intensional contexts are *referentially opaque*; extensional contexts are *referentially transparent*.[7] Extensional contexts are referentially transparent, in that such contexts are not closed under extensional substitution and existential generalisation. That is, subsentential expressions with the same extension may be substituted *salva veritate*. In referentially opaque contexts, in contrast, no such substitution rule holds. An example may serve to clarify. In transparent contexts, from a statement such as:

5. Hesperus is bright.

and a corresponding identity statements, such as

[6]Strictly speaking, 'extensional' is ambiguous. It can mean that substitution of coreferential singular terms is truth-preserving, or it could mean that substitution of coextensive predicates is truth-preserving. See (Mackie, 1974, p. 250).

[7]See (Quine, 1960). For discussion, see (Fodor, 1979).

6. Hesperus is Phosphorus.

one may derive

7. Phosphorus is bright.

However, and in contrast, it does *not* follow from:

8. *S* wrote Hesperus is bright on the Great Wall of China.

and 6 that

9. *S* wrote Phosphorus is bright on the Great Wall of China.

so long as the statements are read under an intensional context, *vis-á-vis* the actual words written.[8] Analogously, belief ascriptions may be read under intensional contexts. There is a context under which it does not follow from:

10. *S* believes <Hesperus is bright>.

and 6 that

11. *S* believes <<Phosphorus is bright>>.

So far, I have outlined the attitude of belief. That deals with [5.1]. I then outlined the conceptual requirement on belief, and showed that the requirement makes belief intensional. In what remains of this section, I will defend what I shall call the *propositional requirement on belief*:

PRB Necessarily, if *S* believes *x*, then *x* is a proposition.

Put in equivalent terms: If some attitude does not take a proposition as its intentional object, then it is not an attitude of belief. For any given possible content a belief may take, that content is propositional.

Why should we think that beliefs are necessarily directed towards propositions? Linguistic considerations give one line of support: We certainly *talk* as though they are. Propositions are denoted by declaratives, and belief ascriptions make essential reference to embedded declaratives. We say that Joey believes *that* Jane is beautiful, that Jane believes *that* Joey is a nice guy. If we make the object some non-propositional object, as in:

12. Joey believes beautiful.

13. Jane believes Joey nice guy.

[8] That is to say, the speaker should be using the relevant terms, rather than merely mentioning them.

the resulting sentences are ill-formed. That gives us *prima facie* reason to accept **PRB**. But some deny that picture. Audi (2013) argues that whilst beliefs are necessarily conceptual, they are not necessarily propositional. Whilst **CRB** holds, **PRB** does not. In what remains of this section, I outline and reject two of Audi's argument against the propositional requirement on belief.

Here is the first argument. Whilst many non-declarative phrases make belief statements ill-formed, some do not. In particular, what he calls *predicative* beliefs i.e., beliefs denoted by statements of the form:

14. Joey believes his chances *to be* slim.
15. Jane believes Joey *to be* a nice guy.

Statements such as 14 and 15 ascribe genuine beliefs, and the beliefs do not have propositions as their objects. Why so? Well, because the beliefs are not expressed by declaratives; there is no 'that' clause. Rather, 'believes' is followed by an *infinitive*, i.e., the verb 'to be'. Because there is no declarative, what follows the infinitive cannot express a proposition. So there are beliefs that have non-propositional objects.[9] Thus, **PRB** is false.

Now, the natural inclination is to say that the distinction is false; 14 and 15 refer to propositional beliefs under a predicative guise. The distinction only holds at the notational level. It may be thought, for instance, that 14 simply is equivalent to:

14*. Joey believes <his chances are slim>.

But Audi maintains this would be mistaken. Predicative beliefs are not propositional beliefs under a predicative guise, they are different in *kind* to propositional beliefs. Why so? Because propositional and predicative beliefs have different properties, and are thus (by Leibniz's law) distinct kinds of belief. In what follows, I outline two differences Audi gives. I will reject both.

The first concerns conceptual content. It is not that predicative beliefs do not have conceptual content, but rather they differ in the *extent to which* they have conceptual content. As he notes, to deny that beliefs must have propositional objects is not *'to deny that believing is essentially conceptual'*. Thus, he accepts **CRB**.[10] He maintains that whilst beliefs that have propositions as objects are *doubly* conceptual, their predicative counterparts are only *singly* conceptual:

[9] He also considers another kind of statement, statements of the form: '*x* looks to *S to be y*', though most of his attention is on the form given. Analogous points, however, apply.

[10] In fact, it is not patent where Audi stands on **CRB**. He seems to waver on the point. Consider: 'We might, then, speak of a purely (non-conceptually) predicative use of 'believe'.' It is not entirely clear what it would mean for a belief to *not* be mediated via a conceptual mode. For present purposes, I have taken Audi to accept (or remain neutral on, at least) **CRB**. The issue I wish to focus on is his denial of **PRB**.

> The point is that whereas propositional believing is *doubly* conceptual—requiring that the believer conceptualise both what the belief is about and what is predicated of it—predicative believing, in minimal cases, is only *singly* conceptual, requiring conceptualisation only of the predicated property. (Audi, 2013, p. 33)

What exactly does Audi mean? What is it for x to be *doubly*, rather than *singly* conceptual? The view is something like the following. In predicative believing, one need not conceptualise both the subject *and* the predicate. In propositional belief, one must conceptualise both. Why should we accept that? To see why, consider the following case:

> *Shiny Glass* *S* spots a shiny object out of the corner of their eye. Unbeknownst to *S*, it is a well polished glass.

In *Shiny Glass*, *S* believes the glass *to be* shiny, but does not believe <the glass is shiny>. After all, *S* does not believe <the shiny thing is a glass>. Rather, *S* is simply aware that *something* is shiny. Thus, *S*'s belief has the concept of shiny, but not the concept of glass as part of its content. *S* does not conceptualise the subject, but only the object. *Shiny Glass* shows that we conceptualise the subject only contingently when it comes to predicative belief. In his own words:

> With predicative belief, then [...] there is no particular notion, no specific conceptual "handle", that must yield the subject of any proposition I believe about the object: in order to believe the rock *to be* craggy, I do not have to believe '*that the rock* is craggy', '*that the stone* is craggy', '*that the surface before me* is craggy', or any such thing. (Audi, 2013, p. 34) (Italics in original.)

Now, in reply one could argue that our beliefs can bear what we might call 'conceptually thin' content concerning the subject. For instance, it may be conceptualised under the existential operator: *S* may believe there exists a shiny thing. The predicative belief, then, might be *de dicto*. So, whilst *S* does not believe <<the glass is shiny>>, it may still be that *S* believes <<there is something shiny>>. Thus, the object is conceptualised as an *existing entity*. Whilst it is thinly conceptualised as such, it is nonetheless conceptualised.

It is quick to see, however, that this reply won't quite do the trick. For existential content of the above sort fails to adequately capture the content of the belief. After all, *S* may have already formed a belief that *something* in the world is shiny; it seems that *S* has formed a belief that differs in content when spotting the glass. The problem is particularly pressing, for in the present case the subject is not properly regarded

as conceptualised; there is nothing about the glass that features in the content of the belief.

But we may enrol another kind of conceptually thin content to jettison the reply: *demonstrative* content. Whilst it may be that *S* does not conceptualise the subject under any particular mode, *S* may well conceptualise the glass as '*that* shiny thing' or 'thus is shiny'.[11] This is problematic for Audi, as now the relevant predicative belief may be translated into a propositional notation, as follows:

16. *S* believes <*that* is shiny>.

In response, Audi may resist the notion of demonstrative content—he would not be the first. But that gives rise to an alternative objection: We will have failed to distinguish propositional from predicative belief, for *if* there are plausibly cases of predicative belief in which we do not conceptualise the subject, then there are plausibly cases of propositional belief in which the same holds. Namely, those declaratives that appear to make reference to demonstrative content. After all, one can believe <that is shiny>, as in 16, whilst the 'that' represents a demonstrative. Or, to take a fairly standard kind of case, Harry may believe <*that* is watching him>, without believing anything about the particular that is doing the watching. The first difference, then, is merely apparent. Either all beliefs involve conceptualisation of the belief's object, or else propositional beliefs, just like predicative beliefs, may not conceptualise the subject.

I turn now to the second purported difference, which concerns the extent to which the beliefs in question are *truth-apt*. He writes:

> Propositional beliefs are (1) true or false, depending on whether *p*—say 'that the font is weighted'—is true or false, [...] [u]nlike propositional beliefs, predicative (objectual) beliefs have a significant degree of indefiniteness in virtue of which it can be misleading to call them true or false. They are accurate or inaccurate, depending on whether what one believes of the object (such as that it is rectangular) is or is not *true of* it. (Ibid, p. 32)

In support, he considers the following kind of case:

> *Noisey Fan* Samantha, a young child, attributes the property 'noise maker' to a fan. She believes the fan to be making noise. However, Samantha is too young to know what a fan is.

[11] For more on demonstrative content, see McDowell (1994a, 1998), Kelly (2001), Peacocke (2001), Tye (2006), and Wright (2003).

His contention is that whilst Samantha's belief is to an extent, 'correct', or 'right', that it lacks the required specificity to be correct or right in virtue of its being true. The possession of a semantic value requires sharp determinacy, which Samantha's belief lacks. Thus, Samantha's belief is not truth-apt. He writes:

> [S]he truly believes *it* to be making noise. She is, then, *right about it*. But this holds even if she has no specific concept of what it is that is making the noise. (Ibid, 31)

It follows that her belief does not express a proposition, for all propositions are truth-apt. Thus, **PRB** is, it seems again, false.

But we should ask: Why are we to think that Samantha's belief cannot be true? That much is not clear. The reason Audi cites is that it is not determinate enough to take a semantic content. He appears to be suggesting as follows: Because there is nothing that Samantha conceptualises the object under, she cannot have true or false belief about it. But this is patently false, so long as we allow demonstrative content. Interestingly, this is exactly what Audi *does* in the above passage: He attributes truth to a belief—the belief <*it* is making noise>. If the 'it' denotes demonstrative content, then its truth conditions are as clear as day. The belief is true just in case the referent of the demonstrative bears the attributed property. So Audi owes us a stronger account of *why* the belief is indeterminate, and in what sense it is.

An alternative argument might run as follows. The content of a predicative belief cannot be true, for the content is presented under a non-declarative form. Thus, Samantha's belief may be *correct*, in the sense that other non-declarative speech acts (i.e., 'Get out of my house and never come back!') may be correct. But the belief cannot, due to its non-declarative form, be true or false. So, there is a difference between the two: The content of propositional beliefs must be true or false predicative beliefs need not. We should note, however, that this view is viciously circular, at least *vis-á-vis* **PRB**. To say that predicative beliefs are devoid of declarative content is itself to reject **PRB**; it thus cannot be used against the requirement. And in any case, that would require that such beliefs are *never* true. But that is patently false. Georgia may truly see the grass to be green, Sally may truly see Harry to be the useless chump that he is.

So far, I have outlined and rejected the first motivation Audi gives for the rejection of **PRB**; namely that there are *predicative* beliefs, which are beliefs that do not take propositions as their intentional objects. I looked at two ways that predicative beliefs are purported to differ. I argued both differences were merely apparent. In the absence of a difference in properties, appeals to Leibniz' law must be revoked. That deals with his first argument. In what remains, I outline Audi's second argument against **PRB**.

The second argument involves a parallel with perception. He writes:

> Notice that just as we may speak of perceiving *that* and perceiving *to be* — which imply corresponding kinds of beliefs, we may speak of believing that something is so and, by contrast, believing a thing to be such-and-such. (31)

It is worth briefly drawing out two claims here. Put schematically, they are as follows:

> (α) For all cases c, S sees <x is such-and-such> in c only if S believes <x is such-and-such> in c.

> (β) For all cases c, S sees 'x *to be* such-and-such' in c only if S believes 'x *to be* such-and-such' in c.

Both may be doubted. It may be argued, for instance, that there is some case c, such that S sees a stick *to be* bent in c, without believing the stick *to be* bent in c. For example, if one is aware of the illusion. Even if seeing is knowing, it is not obvious that knowing implies believing.[12] But for present purposes let us grant Audi the schemata's validity. Audi's claim is as follows. Suppose that in a case c, S does not believe that p. He claims that:

> (γ) Possibly, there is a case c^* in which S sees 'x *to be* such-and-such' and not <x is such-and-such>.

where c^* is alike in all respects to c except with respect to S's seeing. It will follow, given (β), that S believes x *to be* such-and-such, but it will not, at least from the schemata above, follow that S believes <x is such-and-such> in c. He gives the following case in support:

> *Planetarium* Millie believes herself to be in a planetarium. She is incorrect; she is in an open-air theatre. She sees the moon to be bright. However, Millie does not see *that* the moon is bright.[13]

Why is it correct to say, of Millie, that she sees the moon *to be* bright, without seeing *that* the moon is bright? The purported answer is as follows. Millie does not believe that the moon is full, for she falsely believes that what she is seeing is not the moon. As he notes, Millie *'sees the moon and its fullness while falsely thinking that [she] is in a planetarium, and so does not believe either that the moon is full or that it is the moon [she] sees'*.

[12] Cf. (Williamson, 2000).

[13] In the original, Audi speaks of Millie seeing the moon *to be* full. I modified the case to 'bright', as it is unclear that Millie will see the moon *to be* full if she does not believe herself to be looking at the moon, but she will believe herself to be looking at something bright, irrespective of whether it is the moon.

That negates the consequent of the relevant instance of (β). The falsity of the antecedent then follows: Millie does not believe <The moon is bright>. Nonetheless, Millie believes the moon *to be bright*. She does not know it is the moon see believes to be bright, but she nonetheless believes the moon to be bright. So, it follows that:

17. Millie believes the moon *to be* bright, and Millie does not believe <the moon is bright>.

which gives us the negation of **PRB**; there are cases of predicative, non-propositional belief.

Should we accept the argument? We should not. For the argument that Millie does not believe <the moon is bright> rests on an intensional reading of the belief ascription, whilst the argument that Millie does believe the moon *to be* bright rests on an extensional reading. Put in the same context, the case no longer serves as a counterexample. For whilst it is true that Millie (extensionally) sees the moon *to be* bright, we may rerun the same point: Millie does not believe it is the moon she sees, for she believes herself to be in a planetarium. So, whilst she believes:

18. The moon to be bright.

read extensionally, she does not believe 18 read intensionally. Rather, on an intensional reading, she is perhaps better characterised as believing:

19. The computer-generated image to be bright.

but her belief in 19 is now plausibly just a matter of her believing the fine-grained proposition:

20. <<The computer-generated image is bright>>.

It is commonplace to attribute beliefs to agents in an extensional context. The attribution may be true, even if the content of the belief, the mode under which the content is presented, differs for the agent in question. So Audi's case may be interpreted along those lines. Millie may not believe that she is looking at the moon, but she may nonetheless believe <The moon is bright>, or the moon *to be* bright. If so, we have no reason to suppose that seeing the moon *to be* bright does not imply believing <the moon is bright>. Audi's argument rests on a fallacious contextual shift. Thus, we have no reason to reject **PRB**.

In this section, I looked at two arguments given by Audi to reject what I called the 'propositional requirement on belief'. Both revolved around the notion of predicative belief. The first was that there are sufficient differences between the two kinds of belief, the second was that, by analogy with perception, there are cases of believing in the predicative sense, which do not entail believing in the propositional sense. I argued that both fail. In the absence of a compelling argument, we should accept the propositional requirement on belief. That deals with aim [5.2].

5.2 Belief as Dispositional

Above I defended the view that belief is necessarily a propositional attitude. The defence rested essentially on the intentionality of the attitude—its featuring in contexts in which the logical rules of substitution break down, and the notion of demonstrative content. I turn now to the third and final aim of this chapter, to outline the relationship between belief and dispositions.

5.2.1 Simple behaviourism

Belief is a dispositional state. Whilst philosophers are well known for their disagreement, philosophers of all kinds have come to share the view that belief is, in some sense, essentially dispositional. Of course, beliefs are not *any old* disposition—rather they are a kind of *behavioural* disposition. Here are some historical endorsements, first from Quine:

> For all the liveliness and fluctuation of beliefs, believing is not an activity. It is not like scansion or long division. [...] Nor is it a fit or a mood, like joy or grief or astonishment. It is not something that we feel while it lasts. Rather, believing is a disposition that can linger latent and unobserved. It is a disposition to respond in certain ways when the appropriate issues arise. (Quine, 1970, p. 10)

and second from Ryle:

> Certainly, to believe that the ice is dangerously thin is to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people's assertions to that effect, in objecting to statements to the contrary, in drawing consequences to the contrary, and so forth. But it is also to be prone to skate warily, to shudder, to dwell in imagination on possible disaster and to warn other skaters. It is a propensity not only to make certain theoretical moves, as well as to have certain feelings. (Ryle, 1949, p. 135)

Both Ryle and Quine come to the same conclusion, though the theoretical backdrop of their claims could not differ more starkly. On Quine's view, dispositions were 'placeholders' for scientifically respectable concepts; on Ryle's, to attribute existence to dispositions—and thereby states of belief—was none other than to commit gross category error. Dispositional ascriptions did not serve the function of reference. Rather, they served as 'inference tickets'; warrants from assertions to assertions.

On this view, which we may think of as the *simple* dispositional account, beliefs are to be straightforwardly identified with behavioural dispositions. To possess a belief

is to be in some dispositional state. Moreover, dispositional properties can be said to individuate belief-content pairs. What distinguishes beliefs that p from, say, beliefs that q are those dispositions agents who hold the belief bears. What makes two beliefs that p the same is that agents in those states share certain dispositional properties. We can put these two claims explicitly as follows:

Simple Behaviourism

1. *Possession Condition*

> For all propositions p, there exists a unique behavioural dispositional property D, such that for all agents S, S believes that p iff. DS.

2. *Individuation Conditions*

> 2.1 *Sameness* Two beliefs, Bp, Bp* have the same content just in case they bestow the same behavioural dispositional properties on their bearers.
>
> 2.2 *Difference* Two beliefs, Bp, Bp* have distinct content just in case they bestow the distinct behavioural dispositional properties on their bearers.

What is a behavioural disposition? As I understand it, behaviour may be regarded as any in principle observable movement of the agent in question, where that movement is, to use an inadequete turn of phrase 'under one's control'. Thus, non-intentional breathing is not a behaviour; intentional breathing is. It will be helpful, for what follows, to distinguish two kinds of behaviour: *verbal* behaviour, which includes sayings, writings, and other forms of communication, and *non-verbal* behaviour, which includes any kind of behaviour not counted as verbal.

If belief is a behavioural disposition, then quite clearly we should not interpret belief to be a *canonical* disposition, but rather a *conventional* disposition. Perhaps this is what Armstrong had in mind when, in the preceding chapter, he argued that beliefs differ from dispositions in that they are not individuated with respect to certain stimuli and/or manifestations. Why so? The answer is not hard to come by. Take Smith's belief with the content <the volcano is erupting>. There are a wide range of canonical dispositions that are indicative of holding that belief. Crucially, they differ drastically in their stimuli and manifestations. For instance, if Smith will be disposed to upon being asked 'what should we do?', reply 'get under the tables!', and upon seeing pillars of smoke, to run for cover, and so on. So states of belief must be 'reduced down' to different canonical behavioural dispositions, however that is to be achieved.

But importantly, it will not follow that people *will* behave in those ways, at least given the view sketched in part I—not even counterfactually. All that is required is that they *could* be caused to do so. Now, it may be thought that this would severely diminish the explanatory power of the account. After all, if all that ascriptions of belief entail are nomological possibilities, how could we *possibly* get the kind of explanatory power and predictive success from the theoretical framework that we in fact do? This is a strong objection. The behaviourist *may* reply that the dispositions have to be 'sufficiently stable', whatever that amounts to. Thus, it must be that one *could easily*, say, run upon seeing pillars of smoke, or buy flowers upon seeing them if one believes that is one's lover's desire. We do speak of strengths of belief, so this is not an altogether implausible route to follow. Nevertheless, a simpler reply is readily available. Namely: The kind of causal relations relevant to dispositional ascriptions are typically robust. Thus, in the absence of interferences with the causal basis, the relevant causal relations will typically satisfy the relevant counterfactuals. So whilst all that is demanded is a possibility, it is a *causal* possibility, and it is in virtue of the casual requirement that mental states may bear explanatory force in behavioural *explanans*.

Now, if it is true that dispositions are non-identical to their causal bases, and the causal bases are, to put it crudely 'bits of brains', then it will follow that dispositions are non-identical to bits of brains. Nonetheless, it remains open that they are higher-order properties of bits of brains—the claim is only that so long as states of mind are dispositionally classified, no identity relation will hold between mind and body. Rather, on the causal contingent account, beliefs would be, as Goodman feared, modal properties of agents.

What might motivate the simple view? Why should we think, for example, that to believe <The ice is dangerously thin> just is to be disposed to exhibit a range of behaviour, i.e., careful treading, warning peers, and so on. In what follows I support the account. I will give a range of *prima facie* motivations for the account. First, we have four *behavioural* reasons. They are as follows:

> *B-Reason 1* If S believes p, then S will typically be disposed to behave in certain ways. For example, if Billie believes that there are bad men out to get him, then Billie will typically be disposed, under certain circumstances, to draw the blinds, to check electrical devices for bugs, and to monitor phone conversations with care.
>
> *B-Reason 2* If S and S^* have the *same* beliefs, then they will typically be disposed to behave in the same way. If Billie *and* Willie believe that there are bad men out to get them, then Billie and Willie will typically *both* be disposed to, under similar circumstances, draw the blinds, check for bugs, and so on.

B-Reason 3 If *S* and *S* differ* with respect to whether they believe that *p*, then they will typically be disposed to behave in *different* ways. If Willie believes that there are bad men out to get him, but Billie does not, then whilst Willie will typically be disposed to draw the blinds, and check for bugs, in certain conditions, Billie will not.

B-Reason 4 If *S* changes their beliefs, then *S* will typically change their behaviour. If Willie believes that bad men are out to get him, but then comes to be convinced by Billie that he is just being paranoid, then he will typically no longer be disposed to draw the blinds, etc.

Second, we have two *epistemological* reasons. They are as follows:

E-Reason 1 We can come to know what agents believe, by observing their behavioural manifestations. Billie can know that Willie believes there are bad men out to get him by watching him draw the blinds, check for bugs, and so on. Similarly, we can know what agents do *not* believe via the same process. If the vicar lives a life of hedonism and debauchery, we may come to doubt their faith.

E-Reason 2 We can predict people's behaviour by knowing their beliefs. If Billie knows that Willie believes there are bad men out to get him, then he can successfully predict that he will behave in a paranoid fashion.

None of these are decisive. My point is not to defend the view fully, but to outline why one might be tempted to hold it. The reasons given at least *appear* to present strong *prima facie* evidence for *Simple Behaviourism*. Moreover, they present stronger evidence for the view, virtually platitudinous in contemporary theorising, that the concepts of mind form a *folk explanatory framework*. We employ the conceptual apparatus of mind to explain why so-and-so did such-and-such, and to predict what so-and-so will do given that certain conditions obtain.

A further positive upshot of the simple account is its ability to demarcate beliefs that differ in their presentational modes. For it may be thought that beliefs which express the same proposition, but present under distinct modes, are identical with respect to certain 'core' behavioural dispositions, but non-identical with respect to certain 'peripheral' behavioural dispositions, such as certain verbal dispositions. Take for example the belief <<Hesperus is bright>> and the belief <<Phosphorus is bright>>. One who believes << Hesperus is bright>>, but not <<Phosphorus is bright>>, will be disposed to assent to requests such as, 'could you tell me whether Hesperus is bright?' but not to 'could you tell me whether Phosphorus is bright?' But believers who hold just one will be united, perhaps, in their disposition to point to Venus, if asked to point to a bright star.

Further support for *Simple Behaviourism* may be found by considering several ways in which beliefs are structurally similar to other dispositional concepts. In particular, as noted above by Quine, beliefs like dispositions 'lie latent'. Consider a lump of sugar. The solubility of the lump is amongst its dispositional properties, but it does not reveal itself at all times, but only once its stimulus conditions have been satisfied; in this case, when the lump is placed in water. At that time, the disposition comes out of hiding—It manifests. Analogously, a belief may lit latent, and need not manifest in behaviour at any given time at which it is possessed. Rather, it need only manifest when certain stimulus condition obtain. Consider Audi:

> It is probably uncontroversial that the property of believing [...] is dispositional rather than occurrent. [...] Take believing that I am a conscientious citizen. This is, in part, being disposed to say that I am one, under conditions that elicit that sort of verbal manifestation of my belief, such as your asking me whether I intend to vote. Yet I have this belief in dreamless sleep, just as sugar can be soluble while in a solid, unaltered lump. By contrast, to have an *occurrent* property (at least for mental properties) is to be doing, undergoing, or experiencing something. (Audi, 2013, pp. 30-31)

Audi makes use here of the dispositional/occurrent dichotomy. Beliefs are not *occurrent*, on this view, in the sense that to believe p, one's belief need not be manifesting in any sense. Rather confusingly, this same distinction has been employed to denote a quite different dichotomy. Examples may be found in the following passages, the first from Armstrong:

> Beliefs are states of the mind which, so far from us being currently conscious of, we need not even know that we possess. (Other people, or we ourselves at a later date, may postulate their existence in order to explain some feature of our observed conduct or our mental life). Nevertheless, a belief can be a content of consciousness. It can be 'before our mind'. (Armstrong, 1973, p. 21)

the second from Mumford:

> Occurrent and dispositional states of belief can be distinguished. Occurrent beliefs are mental events, such as John's belief at 3 o'clock on Thursday that he is being watched. Dispositional beliefs are more enduring states that can be ascribed over longer periods of time and need not be currently entertained for the ascription to be true. (Mumford, 1998, p. 7)

the third from Cummins:

> There is a familiar distinction of common sense and philosophy between beliefs that one has but is not currently aware of—"dormant beliefs" if you will—and beliefs that are "currently present to consciousness." (Cummins, 1983, p. 77)

The dichotomy Mumford, Armstrong, and Cummins have in mind is distinct from that carved above. It is more akin to Hume's distinction that employs the same terms,[14]. Call a belief that is occurrent in this way 'Hume-occurrent'. Call a belief that is occurrent in the previous sense 'Audi-occurrent'. Hume-occurrence differs from Audi-occurrence, for whilst the former implies the latter, the converse fails. Hume-occurrence is *just one way* that a belief may be Audi-occurrent, for a belief's being 'before the mind' is, at most, only one kind of way in which a belief may manifest.

Here is a case to elucidate. Suppose that Sally believes that chocolate cake is bad for her health. She may believe this, even if she does not spend *all* of her time reflecting on the proposition. We might think, in contrast, that it 'sits in her mind' in some sense and perhaps only comes to the forefront when she has perceptual experiences of chocolate cake. She only Hume-occurrently believes that chocolate cake is bad for her health when she sees a tempting slice. But beliefs may manifest in her behaviour in all sorts of ways that arguably *do not* require the belief to be 'before the mind'. Put another way: A belief may manifest in behaviour, in the absence of the believer's conscious reflection on its content.

For example:

> *Good Catch* Katy knocks a glass she believes to be precious from the cupboard. Without thinking, Katy catches the glass before it smashes on the marble surface.

In *Good Catch*, Katy may be said to manifest her belief that the glass is valuable. But she may not consciously attend to that belief; it may be said to exhibit in her behaviour without being accessible to reflective thought. We might, then, take the conscious/unconscious distinction to be a subset of the manifestations the belief gives rise to. Not all manifestations of dispositions involve conscious reflection, but all conscious reflections on beliefs are doxastic manifestations.

Whilst the distinction is good, I would like to suggest that certain purported implications of the latency of belief are wrong-headed to draw. In particular, it appears that the distinction has resulted in the attribution of far more beliefs to agents than they in fact hold. One clear example may be found in the following passage from Lycan:

[14]See (Hume, 1739, Bk I, Part I, Sec. 7).

> At this moment (even as I write), my wife Mary believes that my tie looks like a prize from the country fair coconut shy. This harsh observation occurred to her a few seconds ago, and she has just voiced it. [...] My wife also believes a number of other things, or so we might routinely suppose: that she is less than eighteen feet tall, that 10,329 < 10,328, and that snow in Stockholm does not instantaneously turn bright orange when it hits the ground. (Lycan, 1986, p. 61)

Lycan's contention here is that Mary *explicitly* believes that her husband's tie looks like a prize from the country fair coconut shy, but merely tacitly believes the proposition concerning Swedish snow. We should note that there are two ways this case may be read, that concern which propositions one counts as tacitly believing on Lycan's account.[15] To see that, consider Mary's belief <Mary is less than eighteen feet tall>. On the one hand, Lycan may be read as suggesting that Mary has in fact formed the belief, perhaps due to actively considering whether or not her own height is more or less than eighteen feet. On the other, let us suppose that Mary counts as unconsciously holding the belief on the basis of simply knowing that she is under six foot tall, and that six feet is shorter than eighteen feet.

The example suggests the latter. But this permits the possession of an overwhelming number of beliefs, and thus Lycan's account appears to commit us to the view that we have far more beliefs than we might commonly suppose. For Mary will be counted as having infinite corresponding beliefs, where she believes that she is not over n feet tall, where the value of $n > 18$. This should give us pause. We do not obviously have infinite beliefs. It is plausible that most of our beliefs are not occurrent at any given time, perhaps even that there are some beliefs we have never, and perhaps will never, bring to conscious attention. But to commit to that should not require commitment to our holding an infinite number of beliefs.

How, then, might we explain Lycan's example away? Now following Audi, I suggest that Mary does *not* believe <Mary is less than eighteen feet tall>, but rather she is *merely disposed to believe* <Mary is less than eighteen feet tall>. As he notes:

> Antecedent belief of the propositions in question, believing them before being asked whether we do, is also the readiest explanation of why we answer the questions affirmatively without having to think about them. These considerations incline many people to attribute to us far more beliefs than, in my judgment, we have. Antecedent belief may be the readiest explanation of our spontaneous answers, but it is not the best explanation. I contend that, here, what may seem to be antecedently held but as yet unarticulated

[15] Cummins (1983, p. 7) gives a similar example: The belief that crocodiles do not wear silk pajamas.

> dispositional beliefs are really something quite different: dispositions to believe. (Audi, 1994)

The basic view is as follows. Mary may be quick to answer questions, such as whether she is less than eighteen feet tall, but this need not imply that she in fact believes that she is. There are, in contrast, some beliefs that we can form quickly, and without significant cognitive effort.

An opponent may object: But if beliefs are dispositional, does Mary not have the relevant dispositions here? If she were to be asked whether she was less than eighteen feet, for instance, she would respond positively. But this objection rests on a confusion. For Mary does *not* have the disposition in such a case; rather she is disposed *to gain the disposition*, given certain stimuli. Put another way: Mary has a *pre*disposition to enter into a dispositional state. For analogy, darkening glasses have the disposition to enter into a kind of dispositional state; upon being hit by sunlight, they are disposed to alter their properties in such a way as to be disposed to reflect light differently. But they do not bear the latter disposition until they have undergone such a change. They are *disposed* to alter their dispositional properties. So too for Mary; she does not *believe* the proposition, but she is disposed to do so. She is disposed to enter a dispositional state—The state of belief.

5.2.2 Armstrong's three differences

So far in this section, I have detailed several lines of support for what I called *Simple Behaviourism*, which takes belief possession to be coextensive with the possession of certain behavioural dispositions. I turn now to my third aim: To outline and reject three purported differences between dispositions and states of belief. All three are given by Armstrong (1973). The first concerns the nature of dispositional stimulus conditions, the second dispositional manifestations, and the third the recombinant character of content. Each will be taken in turn.

Difference 1. The first Armstrong considers concerns dispositional stimulus conditions. He writes:

> One point of distinction between dispositions such as brittleness, is that the concept of the former involves the notion of an initiating cause of a certain sort which triggers off the manifestation. The brittle glass is brittle because it breaks *when hit*. A piece of sugar is soluble because it dissolves *when placed in water*. But the concept of belief seems to involve no notion of a class of initiating causes which in turn bring about the manifestation of expression of the belief. No doubt initiating causes will always be present when the belief is manifested. But they play no special role in the concept of belief. (Armstrong, 1973, p. 16)

The purported difference is as follows: (1) prototypical dispositional concepts are individuated in terms of their stimulus conditions, whereas (2) the concept of belief is not. I do not disagree with the latter, but I do disagree with the former. For take Armstrong's own example: It is not just *hitting* that counts as the stimulus for brittleness. It could be striking, twisting, dropping on (or throwing at) a hard surface. It could even be a sufficiently loud voice, appropriately pitched. It may be argued that the various stimuli above may be classified in terms of the exertion of some kind of force. But that, it should be noted, cannot be part of the *meaning* of the dispositional concept; it was *discovered* that all relevant stimuli admitted of such classification, it was not part of the *concept* of brittleness. So, there appears no reason to suppose that this gives us a genuine distinction between the concept of belief, and other dispositional concepts.

Difference 2. The second concerns not dispositional stimulus conditions, but rather dispositional manifestations. Consider:

> If brittleness is manifested, it can be manifested in only one sort of way: the brittle object breaking if stuck. But there is no one such way that a belief that the earth is flat must manifest itself, if it does manifest itself. For instance, the manifestations may not take the form of outer or inner assent. (Armstrong, 1973, p. 17)

Again, this claim is not obviously true, not on the belief front, but on the brittleness front. For brittleness can manifest in a wide range of ways: A brittle glass may crack, shatter, splinter, and so on. Or take another dispositional concept: Fragility. Kneecaps, fine china teapots, and old parchments are fragile, but they manifest their fragility in a wide range of ways. So beliefs are not *un*like, but rather *very much like* dispositions in that regard.

Difference 3. The third difference may be found in the following passage:

> In the case of beliefs, as opposed to dispositions like brittleness, it seems that the states involved must have a certain internal structure. Suppose that A believes (i) the cat is on the mat; (ii) the cat is asleep; (iii) the cat is black. These three beliefs, although all different, involve a common element. Now if we take beliefs to be states of the believer, must we not take it that these states have an internal structure such that to common elements in the thing believed correspond common elements in the state which is the belief? [...] How otherwise could beliefs with different content give rise to different manifestations or expressions? (Armstrong, 1973, p. 18)

Armstrong's phrasing relies upon the identity theory, which we have already rejected, but it may be re-cast in ontologically neutral terms. Armstrong may demand

instead that the causal bases bear an 'internal structure' that explains the similarity in their manifestations. Nevertheless, the argument fails. Here is why. First, note that the passage makes the following three claims:

> 21. The difference in the relevant manifestations of beliefs that differ in content can only be explained by differences in the internal structure of their causal bases.
>
> 22. Where beliefs have similar contents, their causal bases must be similar in some respect.
>
> 23. Dispositions like brittleness need not have a certain internal structure.

21 seems plausible enough. But what about 22 and 23? First, it is not obvious what 23 is supposed to amount to: Does he mean distinct objects that are brittle? To show there is a genuine difference here, Armstrong needs to give us other dispositions that have a common element, where those dispositions do not require an 'internal structure'.[16]

But even granted that one can be found, 22 appears straightforwardly false. Put another way: There is an answer to his rhetorical question. As Hurley puts it, Armstrong's claim rests on something akin to a vehicle/content confusion. Consider:

> The assumption that the processes that support true thought must have a classical architecture (even if they are implemented by a connectionist network) imposes a requirement of *causal systematicity* on thought. On its face, this looks like a vehicle/content confusion. (**?**, p. 1)

There is no requirement that the structure of the vehicles of content be isomorphic with the purported structure of the psychological attitudes. Here is an example to show why. Consider the following case from Dretske:

> *Basketball Play* "Let this dime on the table be Oscar Robertson, let this nickel (heads uppermost) be Kareem Abdul-Jabbar, and let this nickel (tails uppermost) be the opposing center. These pieces of popcorn are the other players, and this glass is the basket. With this bit of stage setting I can now, by moving coins and popcorn around on the table, represent the positions and movements of these players. I can use these objects to describe a basketball play I once witnessed." (Dretske, 1988, pp. 52-53)

[16] Incidentally, it is not obvious why Armstrong accepts 23., given his acceptance of a type-type identity theory.

Suppose that we design a machine that represents the players in *Basketball Play*, and moves pieces of popcorn and dimes to represent the players on court. Does it follow that each of the movements it represents that are similar must realised similarly in machine? For instance, let the court's perimeter be denoted by $ABCD$, one corner per letter. Suppose that x moves from A to B, and y from C to D. There are geometrically similar properties between the movements. But the machine may track movements from A to C using a distinct mechanism from the way in which it tracks movements from B to D. And so it seems to follow that there need be no similarity in the structure between that which represents phenomenon that are similar content-wise. And that something like this is going on with belief, I contend, is at least conceptually possible. So it is far from obvious that this supports the view that beliefs differ from standard cases of dispositions.[17]

In this section, I took a brief first look at the claim that belief is a dispositional state. I outlined what I called *Simple Behaviourism*, and then gave it support. I then outlined and rejected three purported differences, raised by Armstrong, between beliefs and other dispositions. That deals with the final aim, [5.3].

Conclusion

In this chapter, I had the following aims:

> [5.1] To outline the attitude of belief.
>
> [5.2] To defend the propositional requirement on belief.
>
> [5.3] To outline contemporary theorising on the relationship between dispositions and beliefs.

which have now been satisfied. In the first section, I outlined the attitude of belief, and defended the 'propositional requirement on belief' from an argument given by Robert Audi. In the second section, I provided a preliminary account of the view that belief is dispositional. Finally, I offered defence of the account, and rejected three differences between dispositions and beliefs given by Armstrong.

[17] For similar discussion, see (Fodor, 1987), (Fodor & Pylyshyn, 1988), and (Aizawa, 2003).

Chapter 6

Belief as Dispositional

Introduction

According to what I called Simple Behaviourism, beliefs are behavioural dispositions. Nowadays, dispositional accounts of belief—at least of that form—are highly unpopular. If beliefs are not behavioural dispositions, what are they? Whilst there is hardly unanimous agreement, orthodoxy dictates that if we are to be dispositionalist about states of belief, then we ought to be *functionalists*. As we shall see, whilst that may require that beliefs have certain dispositional properties, there is no straightforward identification between beliefs and those dispositions.

Whilst the doctrine is met with widespread scepticism, there is a sense in which functionalism has become the default model of mind. And that is bad news, I say. And I say that is bad news because functionalism in the philosophy of mind is not just false—it presents a radically inaccurate conception of mental states. My gripe concerns the functionalist's commitment to spatiotemporal individuation of mental phenomena. To be individuated functionally is to be individuated in terms of causal roles. And that requires spatiotemporal existence. And the very notion that mental states are spatiotemporal has, I believe, misled many philosophers of mind.

In this chapter I am going to defend a dispositional account of belief. To say that S believes p, on the view to be developed, is akin to saying S is flammable, fragile, or (perhaps less surprisingly) irascible. The account will be close, though non-equivalent, to Simple Behaviourism. I call it *Complex Behaviourism*. The basic thrust is as follows: The arguments against a dispositional conception of belief are *weaker* than many have assumed, and the arguments against functionalism are *stronger* than many have assumed.

Here are the aims of the chapter:

[6.1] To outline two kinds of 'functional analysis'.

> [6.2] To outline Functionalism in the philosophy of mind, and to argue that it is more plausible than Simple Behaviourism.
>
> [6.3] To outline and support Complex Behaviourism, and to argue that it is more plausible than Functionalism.

Here is the structure. There are two sections. In the first, I deal with [6.1] and [6.2]. I outline two kinds of 'functional analysis', and outline functionalism in the philosophy of mind, showing that it is more plausible than Simple Behaviourism. In the second section, I deal with [6.3]. I outline an argument given by Helen Steward (1997) against the spatiotemporality of mental states, and an alternative novel argument that employs the notion of plural realisation discussed in Chapter 2. Finally, I outline and defend complex behaviourism.

6.1 Functionalism

In this section, I deal with aims [6.1] and [6.2]. There are two subsections, one for each respective aim. In the first, I outline and distinguish two kinds of 'functional analysis'. I call them F-analysis and D-analysis. In the second, I outline functionalism in the philosophy of mind, and show that it relies upon the notion that mind can be F-analysed.

6.1.1 Functional analysis

Suppose you were to stumble upon some mysterious long-lost artefact. You investigate the artefact, and come to learn that it exhibits some kind of behaviour—perhaps it glows when the sun strikes its flattest face. An inquisitive mind may ask: How does it work? And one way of answering that question is to perform what Cummins calls a 'functional analysis'. Painted with a broad brush:

> Functional analysis consists in analysing a disposition into a number of less problematic dispositions such that programmed manifestation of these analyzing dispositions amounts to a manifestation of the analyzed disposition. (Cummins, 1983, p. 28)

According to Cummins, functional analyses takes two forms. For reasons that should become obvious, I will reserve the term 'functional analysis' (or F-analysis) for the first. The second, also for reasons that should become obvious, I will call 'dispositional analysis' (or D-analysis). Let us take each in turn.

F-analysis

F-analysis involves a two-step process. The first step Cummins calls 'system analysis'. System analyses are mereological: To provide a system analysis of x is to break x down into its component parts. The second stage involves the determination of dispositions those parts bear that, when manifesting in unison, give rise to the relevant dispositions that are to be explained. For instance, if we wanted to know how a has the disposition to glow once sun strikes its face, one may deconstruct a, and attempt to understand how its parts work together to give rise to that disposition. He gives another illustrative case:

> Assembly-line production provides a transparent example of what I mean. Production is broken down into a number of distinct tasks. Each point on the line is responsible for a certain task, and it is the function of the workers/machines at that point to complete that task. If the line has the capacity to produce the product, it has it in virtue of the fact that the workers/machines have the capacities to perform their designated tasks, and in virtue of the fact that when these tasks are performed in a certain organized way—according to a certain program—the finished product results. (Cummins, 1975, p. 760)

It is worth noting that it is not only entities naturally construed as systems that admit of functional analysis. A glass' fragility is functionally analysable, but a glass is not naturally construed as a complex system. Suppose we take a particular glass that is fragile; it will be fragile in virtue of its molecules being bonded in a certain fashion. Now, when the disposition manifests (i.e., the glass smashes) the molecules, given their bonding, are disposed to behave in certain ways. When the bonds break a chain reaction is set off which results in the breaking of further molecular bonds. It is this manifestation which is causally responsible for the shattering of the glass. So we can explain the fragility of the glass by breaking down the glass into its molecular components, and citing the dispositions of those components.

Once the relevant parts have been isolated, and their dispositional properties have been determined, the functional analysis is complete. Now, as presented, functional analysis is just a method of explaining how an individual entity has certain dispositional properties, but we may classify entities in functional terms also. For instance, we could take two production lines, and classify any part that bears a certain kind of dispositional property in terms of its role in the production line taken as a whole. For example, we could classify all entities that take a specific task on the production line as 'labellers'. The membership of such classifications is determined by occupation—or in some cases the possible occupation—of causal roles.

Whilst functional analyses are in principle possible, and do occur in the sciences and elsewhere, in the next chapter I am going to argue that many of the purported functional classifications typically do not proceed in terms of the actual or possible occupation of causal roles, but rather in terms of 'functions' read teleologically. Membership of the class of hearts, for instance, or of the classifications relevant to schematic diagrams in electronics are not individuated in purely causal terms, but in terms of purposes.

D-analysis

To constitute an F-*analysans*, x must constitute a spatiotemporal part. And that is because of the first step: Mereological analyses of systems are spatiotemporal analyses. D-analysis differs from F-analysis in that regard: it involves no system analysis. Nevertheless, there is a sense in which D-analysis bears a mereological component. To perform a D-analysis of x, one simply explains a complex disposition in terms of the possession of two or more simpler dispositions of x. Cummins mentions as examples a cook's disposition to bake cakes being 'broken down' into the other dispositions of the 'whole cook', and the dispositions relevant to the multiplication of 27 and 32 into the dispositions to multiply 2 and 7, and to add 5 and 1. Similar point may apply to elasticity. To be elastic, perhaps, is to deform reversibly under stress: This requires both the disposition to deform, and the disposition to reverse its deformation. In that sense, elasticity may be thought to be a complex disposition.

As we shall see, it appears to have been almost univocally assumed that functionalism in the philosophy of mind involves F-analysis. But as I shall argue, it is D-analysis that is relevant to the theory of mind, if either at all. And that is no trivial difference: The two kinds of analysis differ radically in kind. One difference that will become important in what follows concerns the entities such analyses attribute dispositional properties to. In an F-analysis, the dispositions are attributed to *parts* of the analysed entity. In a D-analysis, in contrast, the dispositions are possessed by the functionally analysed entity taken as a whole.

So far, I have dealt with [6.1]. I now turn to [6.2], namely to outline functionalism in the philosophy of mind, and to show that it is a better theory than Simple Behaviourism.

6.1.2 Functionalism

Functionalism in the philosophy of mind subsumes the mental under a more general theoretical kind. Crudely put, functionalism characterises the mind as a kind of complex system, wherein beliefs, desires, and rest of the attitudes occupy functional roles that, taken together, result in the myriads of observable human behaviours. In that sense, functionalism is starkly contrasted with substance dualism. Mental states are not mental in virtue of their being composed from 'mental stuff', they are mental in

virtue of their having certain causes and effects. Because the occupation of causal roles is often enough substance-neutral, so too is functionalism. Mental states are mental in virtue of *what they do*, not *what they are*.

Nevertheless, the term 'functionalism' when used in the philosophy of mind is somewhat of a family resemblance term; it denotes a range of theories that are similar in certain respects, disparate in others. As Block (1978) notes, this is primarily because such theories are:

> the products of a number of rather different projects: attempts to reformulate logical behaviourism to avoid objections, attempts to exploit mind-machine analogies, attempts to apply empirical psychology to philosophy of mind, and attempts to argue for—or against—mental-neurological identity theses. (Block, 1978, p. 261)

We can distinguish those functionalist theories that concern mental state *concepts*, and those that concern mental states *themselves*. The former we call *analytic* functionalism, the latter we may call *ontological* functionalism. The difference is not trivial: A proponent of the former may reject the advances of the cognitive sciences as an advancement in our understanding of mental structure. Like Lewis, we may simply take the concept of mind to be a folk hypothesis, like so many other folk hypotheses, irredeemably useful, but nonetheless false. In what follows, I will be concerned with analytic functionalism. I will not be primarily concerned with the relationship of the concepts of folk psychology to those of the cognitive sciences.[1] From now on, the term 'functionalism' should be read as 'analytic functionalism'.

Proponents include Lewis (1966). Armstrong (1968, 1980), Putnam (1960, 1967), Fodor (1968), and Block and Fodor (1972). Central to the view is that mental concepts are causal concepts. Armstrong puts it well:

> The concept of a mental state is something that is, characteristically, the cause of certain effects and the effect of certain causes. What sort of effects and what sort of causes? The effects [...] will be certain patterns of behaviour of the person in that state [...] The causes of mental states will be objects and events in the person's environment. (Armstrong, 1981, p. 22)

Why should we think that mental concepts are causal? I will provide three motivational reasons in favour of the view. Here they are.

1. Causal Ascriptions First of all, we often speak as though mental phenomena cause behavioural events, and as though non-mental events cause mental events. I'll leave the examples to Rundle:

[1] For interesting discussion, see (Wedgwood, 2006).

> Much of what we say about the mental appears to be shot through with causal idioms. Bad news is said to disappoint or depress, a person's face to remind us of another, an unusual happening to intrigue or excite us; a chance remark may make us think of something we were meant to do, and a thought may set someone off smiling or chuckling; a sudden realization that a person means ill may send a chill of fear up the spine, and fear may in turn cause us to tremble or stammer. (Rundle, 1997, p. 1)

It seems to follow that mental concepts are causal concepts.

2. Explanatory Legitimacy On a similar note, further motivation may be found by considering the fact that mental states often enter (apparently irreplaceably) into the explanations of events. For example, to the explanation-seeking question:

> **Q** Why did Jones shoot Smith?

we may offer the following *explanans*:

> **E** Because Jones desired to obtain enough money to pay for his daughter's operation, and believed both that Smith was a millionaire and that he was set to inherit Smith's fortune.

If mental explanation is a species of causal explanation, then the explanatory force of **E** relative to **Q** is legitimised.[2] Beliefs and desires on this view explain behaviour because they are *responsible for that behaviour*, in the same way that lower level dispositions, or disposition-laden parts may explain the possession of more complex, or otherwise higher-order dispositional properties.

3. Distal Causation Finally, support may be found through consideration of cases involving the discovery of distal causes of behaviour. You observe S ϕ-ing, and naturally explain S's ϕ-ing with reference to S's believing that p, desiring that q, and so on. But then you somehow discover that the causal source of S's ϕ-ing was in fact S^*, who had cruelly wired S up to a radio-controlled device. Now, in such a case we would cease to attribute the relevant mental states to S, and attribute them to S^*, albeit with slight variation in content to allow for indexicals and so on. Put another way: When we discover the causation to be distal, we *also* take the mental states to be distal. That would be explained if mental state concepts were a kind of causal concept.

That concludes my defence. Now, it is important to note that whilst functionalists will agree that mental state concepts are causal concepts, they may disagree on their extension. We can distinguish at least two positions that may be held on what mental states *are*. There are *role* functionalists, and *realiser* functionalists. One is a *realiser*

[2]This is particularly true if, like Lewis, we hold that all explanations of events are a matter of providing information about the event's causal antecedents. See (Lewis, 1986a).

functionalist just in case one holds that mental states are *identical to* the occupiers of some specified causal role. One is a *role* functionalist, in contrast, if one identifies the mental state with some second-order property: being such that one has a part that occupies that causal role.[3]

For familiar reasons,[4] the role functionalist cannot help herself to the considerations above, as the relevant higher-order properties are relational, and thus causally inefficacious. In what follows, I am going to argue against functionalism, but my gripe—or at least not my *present* gripe—is not with the role-functionalist. Thus, from now on, when I say 'functionalism', I mean not only analytic functionalism, but analytic *realiser* functionalism.

If functionalism in the philosophy of mind is true, and mental states are causal concepts, then mind would more likely be the product of F-analyses rather than D-analyses. Why so? Well, because dispositions are not causes. Causal bases are causes, and causal bases are non identical to dispositions.[5] And certainly, this fits best with how philosophers have described functionalism in the philosophy of mind. Fodor will serve as a vivid example:

> What came of it was a new account of the type/token relation for psychological states: psychological-state tokens were to be assigned to psychological-state types solely by reference to their causal relations to proximal stimuli ('inputs'), to proximal responses ('outputs'), and to one another. The advertising claimed two notable virtues for this theory: first, it was compatible with physicalism in that it permitted tokenings of psychological states to be identical to tokenings of physical states (and thus to enjoy whatever causal properties physical states are supposed to have). (Fodor, 1985, p. 82)

On this view, mental states are *parts of agents*, parts that bear certain dispositional properties, which when manifesting together give rise to the higher-level behaviour of agents. Beliefs and desires are the spatiotemporal occupants of causal roles: Parts of a cognitive system, parts which, when causally interacting, given rise to the plethora of behaviour human beings exhibit.[6]

Now, in the second half of this chapter, I am going to argue that this is radically mistaken, at least with respect to states of belief. I will leave open whether sensational states—pains, emotions, and so on—are F-analysed. But at best it is D-analysis *not*

[3]In fact, this is only a first approximation. For the canonical version, see (Shoemaker, 1981).

[4]See (Kim, 1989, 1998).

[5]Some do hold that dispositions are causes, though those who do are typically identity theorists. For interesting discussion, see (Squires, 1970), (Stevenson, 1969), and (Armstrong, 1969).

[6]It should be noted that the relevant notion of 'system' need not be construed in purely mechanistic terms. Mental parts may be analogue, like pulleys and levers, but if Fodor is correct, then mental states are most likely realised computationally.

F-analysis that is relevant to belief. And the upshot is that beliefs and desires are dispositions *not* disposition-laden parts. But for now, I want to briefly show why it is tempting to suppose that mind admits of an F-analysis.

For a working definition, let us define functionalism about belief. As before, we will give both possession and individuation conditions:

Functionalism

1. *Possession Condition*

> For all propositions p, there exists a unique functional role R such that for all agents S, S believes that p if and only if S has a part X that occupies R.

2. *Individuation Conditions*

> 2.1 *Sameness* Two beliefs, Bp, Bp* have the same content just in case they occupy the same causal role.
>
> 2.2 *Difference* Two beliefs, Bp, Bp* have the distinct content just in case they occupy distinct causal roles.

In what remains of this section, I am going to argue that Functionalism is more plausible that Simple Behaviourism which, recall, was defined as follows:

Simple Behaviourism

1. *Possession Condition*

> For all propositions p, there exists a unique behavioural dispositional property D, such that for all agents S, S believes that p iff. DS.

2. *Individuation Conditions*

> 2.1 *Sameness* Two beliefs, Bp, Bp* have the same content just in case they bestow the same behavioural dispositional properties on their bearers.
>
> 2.2 *Difference* Two beliefs, Bp, Bp* have distinct content just in case they bestow the distinct behavioural dispositional properties on their bearers.

Now then, wherein lies Functionalism's theoretical advantage? I will offer two reasons to accept Functionalism over Simple Behaviourism. Here they are:

A. Inter-Mental Causation Functionalism is a better theory than Simple Behaviourism because it is able to accommodate the phenomenon of inter-mental causation. According to Simple Behaviourism, the dispositional properties are behavioural. I was rather liberal on what 'behavioural' meant—verbal and non-verbal behaviour was to be counted. Now, the problem is that the characteristic effects/causes of mental states are often enough the entering and/or exiting of other mental states. Inference provides a transparent example. To believe that no prime number greater than 5 ends in a 5 is, in part, to infer that n is not prime, upon discovering that n is both greater than and ends in a 5. Similarly, the belief that wealthy people have more satisfying lives may result in the desire to be wealthy.

Simple Behaviourism cannot obviously accommodate this insight, as the entering and/or exiting of mental states are not behavioural events. In contrast, Functionalism can. Mental parts may cause and/or be caused by the creation and/or destruction of other mental parts. Moreover, Functionalism provides an excellent account of inferences. Inference is a kind of causal process, where mental parts engage in the reconstruction of one's cognitive architecture.

B. Dispositional Holism Functionalism is also a better theory than Simple Behaviourism because it is able to account for the holistic nature of mental dispositions, and of the holistic nature of mental explanation. To make the idea clear, consider again the following two reasons offered in support of Simple Behaviourism in the previous chapter:

> *B-Reason 1* If S believes p, then S will typically be disposed to behave in certain ways. For example, if Billie believes that there are bad men out to get him, then Billie will typically be disposed, under certain circumstances, to draw the blinds, to check electrical devices for bugs, and to monitor phone conversations with care.
>
> *B-Reason 2* If S and S^* have the *same* beliefs, then they will typically be disposed to behave in the same way. If Billie *and* Willie believe that there are bad men out to get them, then Billie and Willie will typically *both* be disposed to, under similar circumstances, draw the blinds, check for bugs, and so on.

Both of these are false. To show that, it is sufficient to establish:

> *Dispositional Distinctness* Possibly, there exist two agents S and S^*, such that S and S^* both believe that p, and S's belief that p disposes S to ϕ, and S^*'s belief that p disposes S to ψ, where $\phi \neq \psi$.

If Dispositional Distinctness is true, Simple Behaviourism is false. And Dispositional Distinctness is true because mental dispositions are holistic in the following sense:

> *Holistic Thesis* Where S believes that p, the behavioural dispositions S bears are determined, in part, by S's other mental states.

To see, all we must do it consider cases in which agents both believe p, but differ with respect to their other mental states. For instance, suppose that Willie believes there are bad men out to get him, but also believes that they are *deaf*. Well, Willie may be disposed to draw the blinds, but not to check the phone for taps. Now, if Billie believes that there are bad men out to get him, but also believes that they are *blind*, then he will be disposed to check the phone for taps, but not to draw the blinds. Similarly, if Willie desires to be caught, but Billie does not, then only Billy will be disposed to take precautions.[7]

It follows that for any given belief-content pair there is not a unique dispositional property an agent must bear to hold that belief, *for the dispositions an agent bears is determined by an agent's total mental state, not their mental states taken alone*. Similar points may be found in the following passages:

> A belief is not a simple behavioral disposition. At best, it is a disposition to behave in certain ways given certain desires. On the other hand, a desire is at best a disposition to act in certain ways given certain beliefs. This means that there can be no noncircular way to give a purely behavioristic analysis of belief or desire. There is no way to translate a simple statement about belief or desire, without loss of meaning, into a statement that speaks only of purely behavioral dispositions. (Harman, 1973, p. 11)

> To appreciate the virtues of this procedure one should recall the circularities that have plagued attempts to give behavioural definitions of mental states. Someone who believes that it is raining may be disposed to take an umbrella when he goes out—but only if he wants to keep dry. And someone who wants to keep dry may be disposed to take an umbrella—but only if he believes it is raining. It appears that to give a behavioural definition of either of these mental states (the belief, or the want), one would have to mention the other; so there appears to be no way of formulating a noncircular dispositional definition of both, or a purely behavioural definition of either. (Shoemaker, 1981, p. 93)

[7] For the famous cases, see (Putnam, 1963). For discussion see (Gibbs, 1969).

> Pains are responsible for certain kinds of behavior—but only in the context of our beliefs, desires, ideological attitudes, and so forth. From the statement '*X* has a pain' by itself no behavioral statement follows—not even a behavioral statement with a 'normally' or a 'probably' in it. (Putnam, 1963, p. 30)
>
> [T]he corresponding concepts *must be introduced together or not at all.* What falls under mental concepts will be a complex and interlocking set of causal factors, which together are responsible for the "minded" behaviour of men and the higher animals. (Armstrong, 1981, p. 24)

For Armstrong, there is nothing particularly surprising about the holistic character of mental concepts. He writes:

> Correlative or mutually implicated concepts are common enough: for instance, the concepts of husband and wife or the concepts of soldier and army. No husbands without wives or wives without husbands. No soldiers without an army, no army without soldiers. (Armstrong, 1981, p. 24)

However, it is not obvious that the functionalist should take the concepts of mind to be correlative or implicative, in the way that the concepts of husband and wife, or soldier and army may be. Rather, they should be taken to be 'correlative' in the way that parts of functionally classified systems are. Belief stands to desire as brakes stand to engine, not as husband stands to wife.

Functionalism is able to accommodate the holistic thesis. For functional classification *just is* a way of explaining the complex behaviour of a system, through determining the causal interactions of the system's parts. Change the parts, and the behaviours are liable to change. This is nothing mysterious: All it tells us is that the higher-level causal powers of our beliefs and desires are dependent upon the other beliefs and desires (i.e., parts with distinct causal properties) that the system bears. Analogously, I can explain how the production line creates its produce, but in doing so I must explain how the parts of the production line interact to bring about the disposition to, say, produce an item once an order is placed.

Similarly, to explain the behaviour of an agent one must break the agent down into their 'mental parts'—the interaction of those parts gives rise to items of practical reasoning, which in turn give rise to behaviour. To explain why Jones went to the cupboard by citing his desire for coffee and belief that there is coffee in the cupboard, then, is explanatorily analogous to explaining why the blood flows by citing the presence of the heart and the veins, or to explain the production of an object by citing the presence of certain machines on the assembly line. Just as by knowing the make-up of

someone's circulatory system one can predict and explain its behaviour, by knowing what *S* believes, desires, or intends one can predict and explain *S*'s acts.

That deals with [6.2]. I turn now to the second section of this chapter, in which I will deal with [6.3].

6.2 Complex Behaviourism

Functionalism may be a better theory than Simple Behaviourism, but there exists a dispositional account of belief that is better yet. The purpose of this section is to outline and defend that account. I call it *Complex Behaviourism*.

There are three subsections. In the first, I follow Helen Steward in criticising the coherence of the notion of a spatiotemporal state of belief. In the second, employing the argument against token-token identification of dispositions and their causal bases developed in Chapter 2, I argue that functionalism in the philosophy of mind is false. In the third, I argue that a purely dispositional account of mind, *contra* the argument given in the last section, is perfectly able to accommodate both inter-mental causation, and the holistic character of mental dispositions. I then argue that Complex Behaviourism is a better theory than Functionalism.

6.2.1 Token states

As outlined above, on a functionalist account, at least insofar as mental states are the occupants of causal roles, mental states are spatiotemporal parts. Now, amongst the guiding assumptions of functionalist theorising on the nature of mind is the view that mental states are essentially *causal* entities. And the reason why mental states were supposed to be essentially causal lies in their featuring in apparently causal explanations of action. In what follows, I am going to outline two reasons to be sceptical of the view that mental states are the occupants of causal roles. The first concerns the use of the term 'because' in mental explanation. The second concerns the notion of a token state.

A causal because?

A strong motivation for taking mental states to be causally efficacious is that mental explanations of action typically use the 'because *x*' locution, and it is typically assumed that this must denote some kind of causal 'because'. For instance, Rundle writes:

> Such a conception is the starting-point for much contemporary philosophy of mind and action, and it is a starting-point that to many appears unquestionable: if 'because' is not self-evidently causal in this use, at least there

> appear to be no alternative readings clamouring for our attention. (Rundle, 1997, p. vii)

Rundle is certainly correct that this assumption is taken as a starting point in the philosophy of mind. Just consider the following passages from Millikan and Dretske:

> For beliefs, normally a prominent part of one's reasons for acting (desire being another prominent part), are special kinds of representations. Beliefs are those representations whose causal role in the production of output is determined by their meaning or content—by the way they represent what they represent. (Dretske, 1988, p. 53)

> In the peculiar case of human belief and desire, part of the functional role concerns use of these representations in inference, prior to their eventual effects on action. (Millikan, 2004, p. 79)

Nevertheless, it is in fact quite odd to speak of mental states causing events. Statements of the form 'my belief that p caused me to ϕ' are awkwardly phrased. Rather, we would typically say that one ϕ-ed *because* one believed that p. There are exceptions. It is more natural to say: 'A horrible pain caused me to flinch'. But in general, even in the case of desire, it is less awkward to use the 'because x', rather than the 'caused by x' locution.

Moreover, Steward has argued that there is an alternative available. For it is not *only* causally efficacious spatiotemporal entities that enter into causal explanations that involve the 'because of...' locution. The ball fell, smashing a glass. Why did the ball fall? That it was close to the edge has explanatory force, but that it was close to the edge does not denote a causally efficacious individual that moved the ball. Or suppose Sally failed to win the competition: That there were many talented applicants may constitute an *explanans*, though there being many talented applicants is not a causal efficacious individual. Put in Steward's terms: It is not only *causally efficacious entities* that causally explain. *Conditions*—which for present purposes we may construe as sets of cases or centred worlds[8]—causally explain too. No doubt, there is *some* connection between causation and conditions. That the ball was close to the edge tells us that more cases of pushing cause the ball's falling. Lewis, for instance, would no doubt say that they point to possible causal chains—they give information about the various possible causal chains that could have occurred, conditional on the cited *explanans*.[9]

Importantly, dispositional properties can enter into causal explanations as conditions.[10] The ball fell and smashed the glass. Why did the glass smash? That it

[8]See 'Reply to Jackson' in (Williamson, 2009).

[9]See (Lewis, 1986a). Of course, we can bicker about the extension of 'causal explanation'. If your net is not cast as wide, just replace terms where required.

[10]Cf. (Rundle, 1997, p. 230).

was fragile helps us to understand how. Or take another case: A fire rages all night long, burning the building to its foundations. The fire's raging all night long may be explained by its parts being highly flammable. The flammability does not *cause* the building to burn. The cause would be the ill-stubbed cigarette, or whatever else caused the fire. The flammability is a condition, not a causally efficacious spatiotemporal entity. The upshot is that Rundle's question has an answer: an alternative reading of 'because' is available. And so the fact that we use the 'because x' locution is no evidence that mental states are causally efficacious entities. And the oddity of explicitly employing causal language in mental explanations is evidence that they are not.

Stative tokens

According to Steward, another reason that philosophers of mind have been seduced into mistakingly believing that mental states are causally efficacious spatiotemporal entities revolves around the dual imposition of (1) the type/token distinction and (2) the 'state of...' locution. Now, it may be thought that this is none other than bad philosophy—mistaking linguistic for metaphysical considerations. But hold tight. For Steward is well worth her salt. After all—Williamson's (2000) theory of knowledge takes Steward's insight part of its foundation. Put in his terms: Agents are in mental states, not *vice versa*.

Think about it like this: Stative ascriptions typically make use of phrases of the form 'state of x'. As Steward notes, this locution has two uses: (1) where x takes as its possible values *the entity in the state*. Examples include: The state of Jones, the state of play, the state of affairs, the state of the water molecules, and so on. But sometimes the ascription involves a use (2) where x takes as its possible values *state kinds*. Examples include the state of coldness, the state of orbit, the state of disrepair, and so on. So far so good. However, close reflection reveals that the 'state of' locution has a wide application. We speak of states of green, states of play, states of affairs, states of the brain, and of course states of mind. As Robinson has pointed out, it appears probable the locution is but a 'stylistic variant' on predication.[11] Put another way, we can use stative terminology to simply state how things are. If I say the room is in a red state, I am in some sense saying 'how the room is'. In a sense, this is merely equivalent to saying 'the room is red'.

Steward's contention is that the term 'state' has had misleading consequences in the philosophy of mind. She writes:

> [T]he 'state of...' form of expression has nevertheless been very powerful, as stylistic variants go. For it has several features which have enabled philosophers to formulate various theses concerning states which might

[11] Cf. (Robinson, 1990).

> not have seemed nearly so attractive had they had to restrict themselves to other modes of expression. (Steward, 1997, p. 117)

'Objection, objection!'

'Pray tell.'

'Well, Steward has assumed that authors are using the term 'state' in some specific way. But we should not worry about stative terminology: It only serves as a placeholder. Just look here at what Armstrong says:

> I attach no special importance to the word 'state' . For instance, it is not meant to rule out 'process' or 'event'. (Armstrong, 1968, p. 82)

So isn't all of this missing the point?'

It is not. Firstly, the term 'state' *is* used in a specific way by authors, even Armstrong. He may claim otherwise, but consider the following passage from the *very same book*:

> When I have a desire to go out for a drink, I am in a certain mental state (as distinct from a process or event)... (Armstrong, 1968, p. 152)

But secondly, and more importantly, the problem concerns the application of the type/token distinction to stative ascriptions, not the use of stative terminology more generally. Here is the rub. Steward contends that when we use the 'state of...' locution, we are liable to distinguish *token* states from *types* of state. But in fact, whilst there is nothing wrong with the notion of a token state, *there are no entities that are both token states and spatiotemporal particulars*.

It will be helpful to consider what I take to be a fairly paradigmatic stative ascription: States of orbit. You and I can both be in orbit at the same time. We can be both *in the same* orbit at the same time, and in *different* orbits at the same or different times, though of course at no time could we be in the same place in the same orbit. We can count token orbits. Your orbit may be further away, it may be on a different axis. But none of these orbits is a spatiotemporal entity.

Now, it is tempting to attempt to identify the state of orbit with an *event*—say the event that starts upon your entering the relevant belt, and that ends upon your exiting that belt. But this would be a mistake. To see why, suppose Armstrong (Neil, not David) is propelled in a rocket at t_1 and enters an orbit O around Earth. At t_2, Armstrong exits orbit, and lands on Earth. At t_3, Armstrong is propelled identically to before, and ends up in O. Let me ask: Is Armstrong in the same orbit? I say he is! Notice also what Armstrong would likely report: He would say that he had been in the *same orbit twice*, not that he had been in *two of the same orbit*.

For another example, consider a pump-action shotgun's 'loaded' state. That state can be shared: Other pump-action shotguns can be in the same loaded state. But there

are no token 'loadeds'. There are token loadings and token events during which a gun is loaded. But these are not tokens of the loaded state. None of this is to say that there are no token spatiotemporal beliefs, just that we should not be seduced into thinking they are in virtue of the fact that we can speak of token states of mind.

Nevertheless, as I shall now argue, there is at least one compelling reason for supposing that there are no spatiotemporal beliefs. And that reason is as follows: Mental types are more akin to types with no spatiotemporal tokens than to types with spatiotemporal tokens. More precisely, when a type has spatiotemporal tokens, and can be attributed to agents, where those agents do not overlap in their parts, there will exist at least two tokens. In contrast, when a type has no spatiotemporal tokens and can be attributed to agents, where those agents do not overlap in parts, there will *not* exist at least two tokens. And belief is more akin to the latter kind than it is to the former.

Take the type 'pimple'. There are spatiotemporal pimples. Pimples can be counted: I may say 'there are four pimples in the room'. Alas, one can possess pimple-parts. Because token pimples are spatiotemporal, if you and I have a pimple, there must be at least two token pimples: Your pimple and my pimple. Similar points apply to 'heart'. So long as we share no parts, then if you have a heart and I have a heart, then there are two hearts. Compare 'pimple' and 'heart' to the type 'red'. Ferraris are not tokens of the type red. Ferraris are tokens of the type 'red object'. Red does not have spatiotemporal tokens: Token 'reds' include crimsons, burgundies, scarlets, and so on. Suppose you and I are both red. Does it follow that there exist two token reds? The answer is 'no'. You and I may be both the same shade, even if we have no parts in common. Or consider 'states of orbit'. You and I are in orbit and share not parts: Does it follow that there is more than one orbit? Again, not at all: We may both be in the same orbit, perhaps at different points on its trajectory.[12]

Now, 'belief' is unlike 'pimple' and 'heart', and more like 'orbit' and 'red'. And the reason is that 'token' beliefs are not spatiotemporal entities that 'lie in our head', just as 'being red' or 'being in orbit' are not entities that count amongst our spatiotemporal parts. Rather, 'tokens' of the type belief are typically beliefs *that p* or *that q*. If you and I share belief, then we both believe that p. The similarity relation concerns content, not the possession of spatiotemporal particulars. You and I both believe that p—how many beliefs do we know are in the room? This is a strange question, and it is a strange question because we do not typically count beliefs spatiotemporally. If you believe that p and q we can say you hold two beliefs. But the 'two' quantifies contents, not attitudes.

[12]Of course, terms can be ambiguous between distinct types. Consider the type 'animal'. A token animal may be a species of animal, i.e., 'lion', 'tiger', or 'bear'. In contrast, a token of the type 'animal' may be a spatiotemporal organism. If I say 'get that animal out of my house' I would typically have a spatiotemporal token in mind, though I may have in mind all animals of a certain species.

The upshot of all of this is that speaking of 'token beliefs' is highly misleading, because the mere fact that we can apply the type/token distinction does not imply—although tempts us to suppose—that there must be spatiotemporal entities that constitute the 'tokens' of certain 'types'. Not all type/token relations are *abstracta/concreta* relations. And reflection on the ascription of belief—in particular ascriptions that make use of 'belief' as a count noun—suggests that there are no spatiotemporal token beliefs. These considerations are far from decisive. But they give strong reason to suppose that mental states cannot occupy causal roles as functionalists require.

6.2.2 Back to plural realisation

Above I argued that the notion of a spatiotemporal token state of mind is of doubtful coherence. I will now provide an alternative argument for the view that mental states are not identical to the physical occupants of causal roles. It may be thought that much has already been shown. After all, have I not already shown that the identity theory is false? And does it not, then, follow that mental states cannot be identical to states of the brain?

It does not. Whilst *some* functionalists, such as Armstrong (1968), endorse a three place identity relation between disposition, causal basis, and state of mind, the realiser functionalist may merely token-identify the mental state and the occupant, whilst allowing the disposition to be non-identical to both. All I have shown is that *dispositions* are non-identical to their causal bases, not that mental states are non-identical to their putative realisers. Nevertheless, as I shall now show, the argument may be straightforwardly re-deployed.

Here we go. If functionalism holds true, then there must exist some unique spatiotemporal part x, such that x occupies some specific causal role R. But it is conceptually possible that there exists an agent in some mental state M, where there is no unique token property that occupies R. And that is because mental states, even if they are not dispositions, can be plurally realised. Here is an example:

> *Amphibian* An amphibious creature x has two systems that regulate its behaviour. The first system, 'SYSTEM-1', operates when x is on land, and the second, 'SYSTEM-2', operates when x is in the depths. But in shallow waters, both systems operate, which results in identical behaviour to when only one system is operating.

The reader should anticipate the *reductio* to be run. First, we consider a case in which x is on land:

> *Land Itch* The creature x believes that it has an itchy chin, and as a result

> scratches its own chin. Because x is on land, the realiser of the relevant causal chain is 'SYSTEM-2'.

In *Land Itch*, x's itch is realised by 'SYSTEM-2'. On a functionalist account of mind, it follows that x's belief that x has an itchy chin is identical to 'SYSTEM-1'. Yet again, the problem should now be obvious. We can re-run the argument for another case, merely switching the location of x. Consider, for instance:

> *Sea Itch* The creature x believes that it has an itchy chin, and as a result scratches its own chin. Because x is on land, the realiser of the relevant causal chain is 'SYSTEM-1'.

In *Sea Itch*, x's itch is realised by 'SYSTEM-2'. On a functionalist account of mind, it follows that x's belief that x has an itchy chin is identical to 'SYSTEM-2'. By the transitivity of identity, 'SYSTEM-1' = 'SYSTEM-2'. But *ex hypothesi*, 'SYSTEM-1' ≠ 'SYSTEM-2'. We have generated a contradiction. Functionalism in the philosophy of mind is false.

6.2.3 Accommodating holism

In what remains, I will support a dispositional account of belief. A dispositional account is consistent with belief's featuring in mental explanations. As was argued above, dispositions enter into explanations as conditions.

Here is the account:

Complex Behaviourism

1. *Possession Condition* For all propositions p there exists a unique dispositional property D, such that for all agents S, S believes that p iff. DS.

2. *Individuation Conditions*

   2.1 *Sameness* Two beliefs, Bp, Bp* have the same content just in case they bestow the same dispositional properties on their bearers.

   2.2 *Difference* Two beliefs, Bp, Bp* have distinct content just in case they bestow the distinct dispositional properties on their bearers.

Complex Behaviourism differs from Simple Behaviourism only insofar as it does not require that the relevant dispositional properties are behavioural. As such, it allows the entering and exiting of mental states to count. And as has already been

noted, there is nothing inconsistent about the notion of a dispositional manifestation amounting to the entering into/exiting of another dispositional state. When we looked at the distinction between believing and being disposed to believe, we considered the example of darkening glasses which have the disposition to enter into a kind of dispositional state; upon being hit by sunlight, they are disposed to alter their properties in such a way as to be disposed to reflect light differently. *Mutatis mutandis* for belief.

Unlike the functionalist, the complex behaviourist does not conceive of mental states as spatiotemporal particulars. But complex behaviourism then has an onerous burden: To account for the Holistic Thesis. Functionalists have a story to tell as to why the relevant dispositions are determined holistically. Surely, it may be thought, the Complex Behaviourist is at a loss in that respect?

Nevertheless, as I shall now argue, Complex Behaviourism is perfectly able to accommodate the Holistic Thesis. How so? The argument does not attempt to account for the holistic character of psychological explanation by appeal to a single consideration, but rather by appeal to many. The Holistic Thesis holds for several reasons—any search for a panacea will be stillborn. In what remains I offer two ways in which Complex Behaviourism can accommodate the holistic thesis. Here they are.

1. Masking and Mimicking Many of the cases employed to vindicate *Holistic Thesis* can, I contend, be dealt with by the dual notions of masking and mimicking. And this is because sometimes an agent's dispositions are masked or mimicked by other mental states. For example, hope that one wins may mimic the dispositions of a belief that one has won, a desire that one win may mask dispositions relevant to the belief that one has not. Similar points apply to desires. A desire for coffee may mask the desire for getting to work on time: A desire for cigarettes may mask the desire to quit.

This seems relevant to the case given above. If Willie believes there are bad man out to get him, he will be disposed to draw the blinds. But he will not be so disposed if he believes them to be blind. And that is because the belief that the men are blind masks the disposition to draw the blinds. In that sense, certain beliefs can mask the relevant canonical dispositions of others.

2. Stimulus Conditions Other cases can be explained by noting that manifestations can constitute, or at least entail the obtaining of stimulus conditions. This is particularly vivid in cases involving the 'interaction' of desire and belief. On a fairly standard account, to desire that p is to be disposed to act in accord with how one believes that they can make it the case that p.[13] For instance, to desire coffee is to be disposed to act in the way one believes will make it the case that one has coffee. Now, the basic thought is that desires are dispositional, and their manifestations may partly constitute the stimulus conditions of states of belief. To believe that the ice is dangerously thin is

[13] An articulation of this view may be found in (Smith, 1987, 1994).

to be disposed to tread with care, but only if one desires to live. And that is because without desires, beliefs cannot manifest in action.

For all that has been said, there may be reasons to accept a functionalist, as opposed to a purely dispositional conception of mind. I am not at present concerned with offering a full defence of Complex Behaviourism, but I will briefly finish this chapter with two ways in which it is theoretically preferable to Functionalism.

First, the theory is more parsimonious. According to a functionalist account, mental states are (1) spatiotemporal entities that (2) occupy a certain causal role (i.e., bear certain dispositional and integrational properties). But on the complex account, mental states are simply dispositions of agents. There is no need to permit spatiotemporal mental entities into our ontology. No doubt, behaviour *will* be the result of the causal interaction of causally efficacious entities. But those will be physical states—states already permitted into our ontology.

Secondly—and this is important—we avoid the overdetermination issues that vexed materialist theories of mind. If the physical determines behaviour, and so too does the mental, then we appear to have causal overdetermination. Now, on an identity theory this is no problem, for the overdetermination is illusory (the physical determination *is* the mental determination). But we have seen that the identity theory is no good. Of course, there are other approaches one may take. But if we are prepared to accept that mental states are dispositions—as opposed to causally efficacious spatiotemporal entities—then the problem of causal overdetermination fades away.

Finally, a dispositional account seems to give a better fit for linguistic reasons. As we saw above, we typically do not count beliefs as we count spatiotemporal parts. That would be explained if beliefs were dispositions, as opposed to disposition-laden parts. And that is because dispositions in general do not admit of token counting in this sense. It would sound odd to hear 'there are three fragilities/flammabilities in the room'. Rather, one would say 'there are three fragile/flammable objects'. Analogously: 'there are thee beliefs that *p*' in the room sounds off. 'There are three believers that *p*' is more appropriate.

In this section, I supported a dispositional account of belief over a functional account. That deals with [6.3].

Conclusion

This chapter had the following aims:

> [6.1] To outline two kinds of 'functional analysis'.
>
> [6.2] To outline functionalism in the philosophy of mind, and to argue that it is more plausible than *Simple Behaviourism*.

> [6.3] To outline and support *Complex Behaviourism*, and to argue that it is more plausible than functionalism.

which have now been satisfied. The upshot of the above is that a purely dispositional account of belief is at least a serious contender with functionalist analogous, and in at least some respects yields a better account.

Chapter 7

Dispositional Ideals

Introduction

In the last chapter, I argued that a purely dispositional theory of belief is preferable to a view on which beliefs are functionally individuated. On what I called Complex Behaviourism, belief-content pairs are individuated in solely dispositional terms. To share belief is to share dispositional properties. To bear distinct beliefs is to bear distinct dispositional properties. This all sounds hunky-dory: Distinct belief-content pairs enter into distinct psychological explanations, and that would be accounted for by their constituting distinct dispositional properties. The belief that *spiders* are out to get you will result in different behaviour, given different causes, from the belief that *mafiosi* are out to get you. In the latter case, one may be caused to enter a state of fear and run for the hills upon seeing men dressed in expensive Italian suits. One will do no such thing in the case of the belief about spiders.

But even if Complex Behaviourism is preferable to a functional account of belief, it does not avoid a functionalism's greatest challenge, which may be put thus: Content Externalism is both true, and inconsistent with a purely dispositional individuation of belief-content pairs. I call this *the externalist challenge*. Not all are convinced that the challenge is legitimate. Lucky for them! They may rest with a dispositional account. But for those of us who feel its force, content externalism cannot be shirked: Solace must be found. This final chapter aims to seek recourse from the externalist challenge.

As we shall see, there is no way out: to survive, Complex Behaviourism must be modified. Nevertheless, the modification I will suggest retains its spirit. My modificatory approach involves a broader methodological claim about the purpose of the concepts of folk psychology. It has been presumed that folk psychology is primarily—like many a theory of the natural sciences—a predictive/explanatory tool. But as I shall argue, whilst to a large extent that is correct, philosophers have greatly misinterpreted the way in which the explanatory and predictive power of folk psychology is achieved.

On the view to be sketched, folk psychology is first and foremost a *regulatory* theoretical construct. The purpose of folk psychology is to regulate human conduct: It is *not* a hypothetical construct of the inner workings of agents. Folk psychology does have predictive and explanatory power— but it does so in virtue of its being regulatory.

Here is the plan. There are two sections. In the first, I outline the externalist challenge, and give the argument for the inconsistency of Complex Behaviourism and Content Externalism. I then reject three attempts to resolve the inconsistency. In the second, I outline the view that certain theoretical constructs are regulatory, and outline and defend a regulatory theory of content.

The aims are as follows:

> [7.1] To outline the argument for content externalism, and for its inconsistency with *Complex Behaviourism*.
>
> [7.2] To reject a number of context-relative responses to the problem.
>
> [7.3] To outline a regulatory theory of belief-content individuation.

Section 1 deals with [7.1] and [7.2]. Section 2 deals with [7.3].

7.1 The Externalist Challenge

Are dispositional and externalist accounts of belief in tension? According to a fairly standard line of thought, the answer is 'yes'. The aim of this section is to outline both externalism and that tension, and to sketch and reject a number of context-relative reformulations to avoid the problem. Thus, it deals with aims [7.1] and [7.2]. There are three subsections. In the first, I outline Content Externalism. In the second I outline the argument for the inconsistency of Content Externalism and Complex Behaviourism. In the third, I sketch and reject several 'context-relative' attempts to avoid the externalist challenge.

7.1.1 Content externalism

According to

> *Content Externalism* Where S believes that p, the value taken by p is determined by factors external to S.

Put somewhat ambiguously, and employing Putnam's trademark witticism: 'Meaning ain't in the head'. The notion of 'external' may be presicified thus: Allowing an abundant reading, the property 'being such that one believes that p', where p denotes

a unique content is extrinsic. That is to say: Intrinsic duplicates may differ in whether they bear the property of believing that *p*. I call this *the externalist challenge*.

The arguments in its favour are compelling. They employ the case method. In order to work, the following features must obtain: There must be two cases, both involving intrinsic duplicates, but in which intuitively those duplicates differ with respect to the content of their beliefs. We change just the external world, and the contents change themselves.

The first cases were given by Putnam (1975). No need to re-invent the wheel. They are, pretty much in their original form, as follows:

> *Earthy* Earthy lives on planet Earth. On Earth, 'water' refers to the natural kind substance 'H_20'. A waiter pours a glass of H_20 into Earthy's glass. Earthy subsequently forms the belief 'there is water in my glass'.
>
> *Twinny* Twinny lives on Twin-Earth, which is identical to Earth except that every molecule of 'H_20' on Earth is replaced with a molecule of 'XYZ', a distinct natural kind substance. The residents of Twin-Earth call XYZ 'water'. XYZ behaves *exactly* how water behaves with respect to human usage, i.e., it flows, quenches thirst, dissolves salt, and so on. A waiter pours a glass of XYZ into Twinny's glass. Twinny subsequently forms the belief 'there is water in my glass'.

The basic thought runs as follows. Both Earthy and Twinny are (1) intrinsic duplicates but (2) form beliefs that differ in content. This may be seen with reference to the relevant transparent belief ascriptions: Whilst Earthy believes that H_20 is in his glass, Twinny believes that XYZ is in his glass. Of course, neither agent conceptualises their belief under those presentational modes. Nonetheless, so long as we construe concepts as the constituents of propositions, the presentational modes that they conceptualise under do differ. If asked what they are drinking, they may utter the same words. But the *concepts* they express will be distinct. Why so? The answer is straightforward: The truth conditions of the propositions they express are distinct. Earthy's utterance will be true only if there is H_20 in the glass, Twinny's true only if there is XYZ in the glass. Even those sceptical of truth conditional semantics must concede the point—Difference in meaning does not entail difference in extension, but difference in extension entails difference in meaning.[1] The propositional contents of their mental attitudes differ; their intrinsic properties are identical. *Content Externalism* straightforwardly follows.

It should be briefly noted that, strictly speaking, these original cases fail, as they rely on an implausible conjecture: That two agents could be intrinsic duplicates across the cited worlds. After all, we humans are partly composed of H_20. In the world of XYZ,

[1] Cf. (Armstrong, 2004, p. 31)

our intrinsic properties must be distinct, thus the agents are not *bona fide* duplicates. But similar cases may be constructed involving kinds of which we are not composed. In a world in which there are no tigers but 'schmigers', similar arguments may be run.[2]

Moreover, whilst Putnam's view is often dubbed 'natural kind externalism', the point need make no essential reference to natural kinds. Non-natural kinds may be employed to yield similar cases. For example, we could let there be a world in which teddy bears are filled with some non-natural substance 'fluff', and a distinct world in which everything is identical bar 'fluff' being macro-identical, but compositionally distinct. Believers that 'fluff' is soft could constitute twin-earth agents, despite the referents of their beliefs being non-natural kinds. In fact, all that is required is that there are two kinds k_1 and k_2, such that for all relevant purposes k_1 and k_2 cannot be discriminated by the agent given their intrinsic properties.

The second set of cases were given by Tyler Burge. In contrast to Putnam's, in Burge's we alter not ontological features, but rather semantic features. More precisely, the cases contrast with respect to the socially determined meaning of lexical items. Consider:

> *Arty* Arty has developed a pain in his thigh. Arty thus believes that he has 'arthritis'. In Arty's world, the word 'arthritis' means 'inflammation of the joints'.
>
> *Schmarty* Schmarty has developed a pain in his thigh. Schmarty thus believes that he has 'arthritis'. In Schmarty's world, the word 'arthritis' means 'inflammation of the joints or pain in the thigh'.

The thought runs thus: Arty and Schmarty are intrinsic duplicates, but they differ with respect to what they mean when they express propositions with the term 'arthritis'. Arty means what we mean by Arthritis, Schmarty expresses the disjunctive concept. Again, difference in semantic value is raised to prove difference in meaning: Arty's belief expressed by an utterance such as 'I have developed arthritis' is true, Schmarty's false. From difference in truth-conditions follows difference in meaning.

On closer reflection, however, Burge's cases are not obviously problematic. Why might they be thought to fail? Well, we may reject that the truth values of the expressed propositions differ. For instance, it may be argued that the difference concerns whether they express the proposition with those lexical items that society has deemed correct.[3] Put another way: there is difference in *sentence* meaning, but no difference in *speaker* meaning. One can express a proposition using the wrong terms, though one may fail to convey information by doing so. Nevertheless, to those who do hold that the cases are effective, analogous remarks to those that follow will hold.

[2]See also Burge's 'aluminium' case. (Burge, 2007).
[3]Cf. (Crane, 2001, Ch. 4) and (Segal, 2004).

I have now outlined content externalism and the arguments for the view. That deals with [6.1]. I now turn to the second task of this section: To outline the argument for the inconsistency of Complex Behaviourism and Content Externalism in the philosophy of mind.

7.1.2 The externalist challenge

If Content Externalism is true, then Complex Behaviourism is false. Contrapositively, if Complex Behaviourism is true, then Content Externalism is false. In non-conditional terms: Complex Behaviourism and Content Externalism are inconsistent positions in the philosophy of mind. At best, one may be upheld. At least, one must be revoked. Call this

> *The Inconsistency Thesis* Content Externalism and Complex Behaviourism are inconsistent.

Should we accept the thesis? We should. Here is the argument. The first premise is as follows:

> *Premise 1* Two agents are intrinsic duplicates only if they are causal duplicates.

It may be denied that our twins are genuinely causal duplicates. After all—one may contend—given the environment in which they find themselves, their causal powers are distinct. Given where Twinny is, he cannot, say, cause H_20 to flow from the taps, whilst Earthy can. But that is not the relevant notion of 'causal duplicates'. Rather, causal duplicates enter into the same causal relations *in all cases*. As Fodor puts it:

> Roughly, our biceps have the same causal powers if the following is true: For any thing x and any context C, if you can lift x in C, then so can I; and if I can lift x in C, then so can you. (Fodor, 1987, p. 35)

Generalised to agents, and construing contexts as cases, we might say that two agents S and S^* are causal duplicates just in case: for all cases c in which S can enter into some causal relation R, S^* can enter into R in c, and *vice versa*.[4]

Here is the second premise:

> *Premise 2* Two agents are causal duplicates only if they are dispositional duplicates.

[4]If we accept transworld identity, then a more precise formulation would restrict the accessible worlds to those in which both agents are not present.

I leave it open whether dispositional duplicates must be causal duplicates. On my view, the entailment fails to hold: Two entities may have the same dispositional properties, despite those properties being realised by causally efficacious properties that are causally distinct. But the converse (which we require) holds true: Where two agents are causal duplicates, they must be dispositional duplicates. Why so? Well, suppose that it were false: There can be causal duplicates that are not dispositional duplicates. Then ask yourself: In virtue of what would the entities be dispositionally distinct? To count as dispositionally distinct, there must be some difference in the causal properties of the bases of the relevant dispositions. But if there are differences of that sort, the two entities must be causally distinct. But *ex hypothesi*, they are not causally distinct. Assuming the second premise's negation permits derivation of a contradiction. So we should accept the second premise.

Now, given transitivity, we may derive from the first two premises:

> *Premise 3* (from 1, 2) Two agents are intrinsic duplicates only if they are dispositional duplicates.

We then need just two more premises:

> *Premise 4* If Content Externalism is true, then intrinsic duplicates can be in distinct mental states.
>
> *Premise 5* If Complex Behaviourism is true, then intrinsic duplicates must be in identical mental states.

Premise 4 is supported with the contrast cases discussed in the previous section, and premise 5 follows from premise 3 and the definition of Complex Behaviourism, which, recall, was:

> *Complex Behaviourism*
>
> 1. *Possession Condition* For all propositions p there exists a unique dispositional property D, such that for all agents S, S believes that p iff. DS.
>
> 2. *Individuation Conditions*
>
>> 2.1 *Sameness* Two beliefs, Bp, Bp* have the same content just in case they bestow the same dispositional properties on their bearers.
>>
>> 2.2 *Difference* Two beliefs, Bp, Bp* have distinct content just in case they bestow distinct dispositional properties on their bearers.

We now merely need to assume that both Complex Behaviourism and Content Externalism are true:

Assumption 1 (for reductio) Content Externalism is true.

Assumption 2 (for reductio) Complex Behaviourism is true.

and we are in a position to generate a contradiction: The antecedents of premise 4 and premise 5 will be true, and thus the consequents follow. Applying conjunction introduction, we may generate a contradiction:

Conclusion Intrinsic duplicates must be in identical mental states, and intrinsic duplicates can be in distinct mental states.

Content Externalism or Complex Behaviourism must be denied.

So far, I have outlined the argument for the inconsistency of Complex Behaviourism and Content Externalism in the philosophy of mind. That deals with [7.1]. I now turn to the final aim of this section: To outline and reject attempts to avoid the conclusion by appeal to context-relativity.

7.1.3 Context relativity

If folk psychology does not individuate belief-content pairs in dispositional terms, where does that leave *Complex Behaviourism*? Defunct, no doubt.[5] In this section, I will consider and reject three alternative approaches. They are united by a common theme: they adopt what we may call a *half-blooded* account, on which attitudes are individuated functionally/dispositionally, but on which attitude-content pairs are not, or at least not wholly. Beliefs are all *beliefs* and not *desires* in virtue of their dispositional/functional properties. But the same cannot be said for *beliefs that p* and *beliefs that q*. Here are two endorsements:

> If, however, *that's* what you want Pyschofunctionalism for, then all you need is the claim that *being a belief* is a matter of having the right connections to inputs, outputs, and other mental states. What you *don't* need — and what the philosophical motivations for Pychofunctionalism therefore do not underwrite — is the much stronger claim that being the belief *that P*, being a belief that has a certain *content*, is a matter of having the right connections to inputs, outputs, and other mental states. (Fodor, 1987, p. 69)

> Functionalists think of mental states in terms of their causal roles. So presumably they have in mind some set of generalizations linking mental states to other mental states and to cognitive inputs and outputs. But if there are such generalizations, surely they will 'quantify over contents'. [...] The

[5] To deny that attitude-content pairs are individuated in dispositional terms should not be confused with a rejection of dispositional theories of *conceptual grasp*, such as that advocated by Peacocke (1992). Conceptual grasp may be the disposition to enter certain dispositional states, for instance.

> key here is to recognise that functionalists needn't think of 'content clauses' like 'that the earth is round', 'that p', 'that q', etc., as doing anything more than *labelling* causal roles. [...] That is, instead of thinking of content clauses as picking out beliefs as the beliefs-that-represent-such-and-such, they can think of them as picking out beliefs *directly* as beliefs-with-such-and-such-causal-roles. (Papineau, 1987, pp. 48-49)

But such a move leaves a hole that must be patched: What if not dispositions individuate belief-content pairs? A tempting line of thought makes use of the notion of *semantic evaluation*. To be semantically evaluable is to take a semantic value, to take a semantic value—at least on a Fodorian view—is to bear a certain kind of relational property. The relational property is both *binary* and *exclusive* in the following sense: There are two such values a given content may bear, and for each content, that content takes at least and most one at any given time.

Semantic values themselves do not individuate contents, at most it is the conditions under which the semantic values hold on a given content. In the case of belief, these are truth conditions. A belief that p is not a belief that q, because the conditions under which p is true/false are distinct from those conditions under which q is true/false. What distinguishes, say, a belief that a tiger growls, from a belief that a schmiger growls, are not the dispositions of agents that are in those states, but rather the conditions under which the attitudes are true.

However, semantic individuation faces well-known objections. For one, it struggles with distinct contents that take their semantic values necessarily.[6] Two tautologies may be reasonably taken to be semantically distinct, despite being identical with respect to their semantic values in any possible case. Mathematical truths may differ in content, but mathematical contents are either necessarily true or necessarily false. Some may follow Dummett in individuating semantic values in terms of *proof* or Peacocke in terms of *acceptability*, rather than *truth*; others may follow Chalmers in taking the hyper-intensional turn.[7] But even if we can avoid this semantic swamp, a more worrying issue remains: The account gets things back-to-front. A content takes its semantic value *in virtue of its being the very content that it is*, not *vice versa*.

The standard approach to wriggle free from the externalist challenge is to invoke some kind of *context-relativity*. In what follows, I will outline and reject three attempts to avoid the externalist challenge via appeal to context relativity.

The first is given by Tyler Burge. He tells us unabashedly that from the point of view of physics, twin-earth counterparts are causal duplicates:

> I assume with Fodor that individuals with the same brain states will make

[6] Cf. (Soames, 1992, p. 23).
[7] See (Chalmers, 2002) and (Chalmers, 2011).

> the same movements. I also assume that there are no gaps (into which mental events might swoop) among the events described by the physical sciences. I further assume that if causal chains as described in the physical sciences ('physical causation') did not occur, causal chains as described by psychology ('psychological causation') would not occur. (Burge, 2007, p. 319)

But crucially, the causal powers of agents in identical brain states are *not* identical from the point of view of psychology. For psychology individuates its entities in terms of possessed causal powers *within a given context*. This should not surprise us, he contends. For context-relative causal individuation is *business as usual*, naturalistically speaking. For analogy, he gives two examples. I will briefly detail the first, for it is straightforward to see that it fails. The example concerns the kind 'heart'. He writes:

> Similarly in biology, organs in the body are typed because of their function in the bodily environment that surrounds them. Something is a heart because its organic function is to pump blood in a circulatory system that extends beyond the surfaces of the heart. One can imagine an organ in a different sort of body with a totally different function (it might pump waste for example). The causal powers attributed to such an organ by biology would be different from those attributed to a heart. Such an organ would not be a heart, but it might be chemically and structurally homologous to a heart. The biological kind heart does not supervene on the chemical structures of material that constitutes hearts. (Ibid, p. 323)

The thought is that 'belief' is like 'heart', in that it is individuated in part by its integration into some environment. What matters are the causal powers the agent bears *in the environment in which they are embedded*, not in environments in which they are *not* embedded. The first question we should ask, however, is whether the example given is convincing. One would certainly be excused for supposing that hearts are causally classified, for we are often sloppy in the definition of categories. For example, the OED defines the term 'heart' as follows:

> *Heart*
>
> A hollow muscular organ that pumps the blood through the circulatory system by rhythmic contraction and dilation. In vertebrates there may be up to four chambers (as in humans), with two atria and two ventricles.

But a scrupulous reader will anticipate what follows: Burge has committed the fallacy of equivocation. Hearts are not classified in terms of their causal powers, but

in terms of the exercised causal powers of previous tokens of the type that explain the persistence of the trait. The OED is incorrect—not all hearts pump the blood; some are diseased. The upshot of all this is that the individuation of biological traits gives us no reason to suppose that context-relative classification is business as usual.

Nevertheless, he offers a second example, one that will not be so readily dispensed with. He considers the geological kind 'plate'. Consider:

> If there were no sliding of land masses across the face of the earth, land masses would not have been typed as plates. Moreover, the causal powers associated with these land masses would have been differently described by geology. (Ibid)

Agreed! Where x bears causally relevant relations to y, then holding fixed that relation the causal powers of x may differ from the causal powers of x when that relation is not fixed. My brakes cause the vehicle to decelerate once integrated in the vehicle; but outside the vehicle they have no such power. This leads him to conclude that:

> A land mass with substantially the same non-relational physical features could—because of its different relations to the environment—be of a different geological kind. Thus geological kinds do not supervene on the kinds of masses that are described by physics and that constitute the geological entities. (Ibid)

So far so good. When geology classifies entities, it classifies them not just with respect to their causal powers, but their causal powers *within a given context*. Something that is a plate may not have been a plate, and it may not have been so for it could have failed to bear certain relational causal powers. Drawing on the account of functions given, we might say that it would not have been a plate had it not *functioned* as a plate – i.e., been integrated in a complex in which it would bear a/some complex relative disposition(s).

But with those remarks in place, it is clear that the case is not sufficiently analogous. For a plate differs in its classification across contexts *because it changes its functional organisation*, which our twin-earth agents' cases do not. To see that, notice that if one were to drop a plate into some other context, then it may cease to be a plate. But drop a believer that p in a different context, and they do not cease to be a believer that p! The environmental relations are of relevance because the plate is classified as *part of the Earth taken as a whole*. But mental states are not parts of their environments in that sense. So the fact that geology individuates its kinds that way says nothing about the individuation of mental states. Mental states are not sufficiently similar to plates. The response fails.

The second I will consider is found in the rewarding work of John Heil. The response arises from what he calls the 'Dart-Tossing' of mental states, on which believing, desiring, and so on, are *projective* states. When one forms a belief, one 'projects out' into the world, in a way similar to how a dart thrower projects a dart to the board. This sounds strange—better to take it from the horse's mouth:

> Suppose the intentional character of states of mind were tied to their dispositionality: what makes a thought about a tree a thought about a tree is the difference it makes to the dispositionalities of the thinker [...] The thought's dispositionality takes advantage of the built in projective character of dispositions. [...] What of an agent's environment? Return to Twin Earth. An inhabitant of Twin Earth who entertains thoughts he would express by uttering sentences containing "water" is entertaining thought about *XYZ*, not thoughts about water. [...] What the thought concerns, however can depend on what context [sic!] in something like the way in which what "here" or "now" designates depends on the location of the speaker or the time of utterance. [...] The thoughts "project to" twin-water, not water, because twin water, is on the scene. (Heil, 2004, pp. 288-289)

Now, part of Heil's motivation for supposing this will work is his commitment to a surprisingly popular account of intentionality (in Brentano's sense). Other proponents include Armstrong (1968), Martin and Pfeifer (1986), Molnar (2003) and Molnar and Place (1996). The basic thought runs as follows: Mental states are *about* the world in that sense, they 'point out into' the world. Dispositions also 'point out into' the world, in particular towards their manifestations. So intentionality just is a feature of dispositional states: The 'aboutness' of mental states is the 'aboutness' of dispositions. The basic thought is that what determines the content of a thought is what is 'out there' in the world. He writes:

> On a dispositional account of intentionality, the projective character of thought—its ofness, or forness, or aboutness—stems from its dispositional nature. [...] The thought "projects to" twin-water, not water, because twin water, and not water, is on the scene. (Heil, 2004, p. 289)

as does Molnar:

> Powers, or dispositions, are properties *for* some behaviour, usually of their bearers. These properties have an object towards which they are oriented or directed. The objects of powers are usually called 'manifestations', a name that carries an epistemological loading. (Molnar, 2003, p. 61)

On this view, then, the context is determined by the projective character of thought. Now, the notion that the projective character of thought is in some sense a dispositional feature is at best confused, at worst false. If you are itching to know why, I defer to the powerful arguments to be found in Bird (Ch 5., 2007). But even granted that implausible assumption, a major problem instantly re-arises: How do we determine the scope of the throw? How far must water be to be outside one's 'mental range'? A dart travels some distance: What it hits and what it can hit is determined by the force exerted by the tosser. But evidently when we form beliefs we do not send out 'mental rays' into the environment. So, we need to answer: How do we determine the 'projective range' of mental states? Perhaps Heil will reply that the scope is determined, in some sense, by the manifestations of the relevant dispositions. Perhaps, for instance, Earthy will be disposed to search for water on Earth, but not on Twin-Earth. But could Earthy not search for water on a long-lost planet? And if so, would the range not then be extended to planets that contain XYZ? Without a suitable account of the nature of the mysterious notion of mental projection, Heil's view should be revoked.

The final view I will reject has been developed by Jerry Fodor. It makes use of the notorious notion of 'broad content', which should be contrasted with 'narrow content'. Narrow contents are individuated in terms of their causal powers alone. Broad contents, in contrast, are functions (read mathematically) from narrow contents to contexts. As he tells us:

> The 'broad content' of a thought, by contrast, is what you can semantically evaluate; it's what you get when you specify a narrow content *and fix a context*. (Fodor, 1987, p. 48)

Earthy and Twinny differ in their broad content, but they do not differ in their narrow content. So functionalism is inconsistent with externalism only insofar as we classify states broadly, but it is in the broad classification that lies the rot. A legitimate classification of mental states would *not* classify this way, and thus a legitimate theory of mind would not be inconsistent with content externalism.

The appeal to narrow content, however, does not come without strife. The most obvious question that we may raise is: How are the functional *relata* determined? If a mental state is a function between attitude-content pairs and contexts, what determines the values of the relevant outputs? Put another way, what *anchors* the content to a context? Fodor has a better answer than Heil. He writes:

> I don't want to worry, just now, about the problem of how to articulate [context]. Some story about constraints on the causal relations between H_2O tokenings and water-thought tokenings (and between XYZ tokenings and 'water'-thought tokenings) would be the obvious proposal; but it doesn't

> matter much for the purposes now at hand . Because we do know this : Short of a miracle, it must be true that if an organism shares the neurophysical constitution of my Twin and [is in the same context], it follows that its thoughts and my Twin's thoughts share their truth conditions. (Fodor, 1987, p. 48)

Nevertheless, even setting that aside, Fodor's account comes with several worries. First, it is notoriously difficult to specify what 'narrow content' is. Fodor claims that we should not expect to be able to individuate narrow contents. But secondly, and more problematically, it simply cannot be the case that beliefs about water and beliefs about twin-water are identical in their causal properties. For it is possible—as we now do—to have beliefs about *both* water *and* twin-water. And they can, in some individuals, yield distinct manifestations. For instance, I am disposed to say that Twinny, but not Earthy, believes that twin-water is wet. So belief-content pairs cannot be functions from narrow contents to contexts, if we are to take the narrow content of the two beliefs to be identical, for one may hold both beliefs in virtue of bearing what appears to be *two distinct narrow contents*. I cannot have *two of the same narrow contents*, at least on a Fodorian view, where those two contents are realised by entities with distinct causal properties. So narrow content, even setting aside the troubles concerning its individuation, fails to overcome the externalist challenge.

That deals with the final aim of this section, [7.2]. I now turn to the final aim of the thesis, [7.3], namely to outline a regulatory theory of content.

7.2 Dispositional Ideals

In what remains, I sketch a novel account of the individuation of belief-content pairs. I use the term 'sketch' with care: The account is rough around the edges. This is an idea to be developed. It does not constitute a finished model. The basic line of thought is to some extent in keeping with the following remark made by John McDowell:

> To make sense of the idea of a mental state's or episode's being directed towards the world, in the way in which, say, a belief or judgement is, we need to put the state or episode in a normative context. A belief or judgement whose content (as we say) is that things are thus and so—must be a posture or stance that is *correctly* or *incorrectly* adopted according to whether or not things are indeed thus and so. [...] The relation between mind and world is normative, then, in this sense: thinking that aims at judgement, or at the fixation of belief, is answerable to the world—to how things are—for whether or not it is correctly executed. (McDowell, 1994, pp. xi-xii)

I do not know what McDowell's overall view amounts to, so I don't know how similar our views are. Nonetheless, the theory to be sketched *does* take the individuation of belief-content pairs to be a distinctively normative enterprise, permitting a wide reading of 'normative'. But certainly unlike McDowell, the normativity I am concerned with does not concern the fact that one's belief ought to correspond to how the world is. Rather, it concerns what one ought to be disposed to *do*, given that one believes.

What remains is composed of two subsections. In the first, I argue that certain classifications, in particular 'proper functional' classifications are what I shall call *regulatory*. In the second, I argue that folk psychology employs regulatory classifications.

7.2.1 Classification and regulation

In *Knowledge and the State of Nature*, Edward Craig attempted to illuminate the concept of knowledge by asking what the concept is *for*, that is—what the *purpose* of ascribing knowledge is. Unlike Craig, I will refrain from employing an explicative approach. But I will ask a similar question: What is the purpose of folk psychology? Why do we ascribe beliefs, desires, pains, memories, and so on?

It has been univocally assumed that belief/desire psychology is primarily a tool used to *predict and explain behaviour*. For some, it does so by providing an accurate theoretical reconstruction of our cognitive apparatus. For instance, Fodor writes:

> Here are some things it seems safe to assume about science: We want science to give causal explanations of such things (events, whatever) in nature as can be causally explained. Giving such explanations essentially involves projecting and confirming causal generalizations. And causal generalizations subsume the things they apply to in virtue of the causal properties of the things they apply to. Of course, in short, what you need in order to do science is a taxonomic apparatus that distinguishes between things insofar as they have *different* causal powers, and that groups things together insofar as they have the *same* causal properties [...] [I]f you're interested in causal explanation, it would be mad to distinguish between [Earthy's] *mental states* and [Twinny's]; their mental states have identical causal powers. (Fodor, 1987, pp. 33-34)

Let us draw out the general thrust of this passage. He makes the following claims: (1) naturalistic taxonomy is distinctively causal, and (2) folk psychology taxonomises in accord with the natural sciences. That seems reasonably innocuous, right?

Wrong. I won't engage in bickering over the extension of 'naturalistic taxonomy', all I will claim is that either (1) or (2) is false. Either not all naturalistic taxonomy

classifies in terms of causal powers, or else folk psychology does not taxonomise in accord with the natural sciences. The basic thought is that folk psychology is more akin to certain classificatory schemes that do not carve along causal joints. My guiding examples will be frameworks that classify entities teleologically.

Proper functional classification serves as the prototype and norm. As I shall now argue, where entities are classified in teleological terms, the primary purpose of that classification is often *regulatory*. What do I mean 'regulatory'? Regulation, as I am using the term, is a normative phenomenon. To regulate x is to uphold x against a standard—to assess it relative to that standard, and importantly to be *guided by* that standard. To regulate, in short, is to *control* by rules. For example, to regulate a clock is to uphold it to a standard—the agreed upon time—and to control the clock in order to make it accord with that standard, i.e., to make it tell the *correct time*.

Consider the classifications of biological traits. No doubt, in the biological sciences these classifications are intimately connected to explanation. But such classifications have distinct purposes in distinct disciplines. In the medical sciences the purpose of proper functional classification is regulatory—medical doctors are not typically interested in the origins of your heart, rather they are concerned with maintaining the functioning of your body, and in doing so preserving life. Analogous points apply to proper functional classifications in the objects of artifice. The mechanic typically wants to know how the car is *supposed to work*, in order to diagnose malfunction. Schematic diagrams in electronics are not *descriptions of systems—they are prescriptions*: They tell one *how one ought to build an item of electronic equipment*.

In a sense, proper functional classification may be seen as a form of *idealisation*, though it differs radically from the standardly discussed cases of idealisation in the sciences. Idealisation typically serves the purpose of noise elimination, of 'honing focus' onto those salient features that a theorist is concerned with. This kind of idealisation, in contrast, is normative. 'Heart' is an idealised category, for to be a heart is to be regulated against an ideal. To be a bicycle is not to operate in a certain way—broken bicycles are bicycles. Rather, to be a bicycle is to be upheld to a dispositional standard.

To classify against an ideal is not to assume that anything could act in accord with that ideal, not perfectly at least. Of course, an idealisation that is impossible to achieve on at least some occasions, or to come close to achieving on some occasions, would be a useless ideal. As Kornblith writes:

> An appropriate human ideal must in some ways be responsive to human capacities. Ideals are meant to play some role in guiding action, and an ideal that took no account of human limitations would thereby lose its capacity to play a constructive action-guiding role. At the same time, our ideals

> cannot be so closely tied to what particular individuals are capable of that we fail to recognize that some individuals at some times are incapable of performing in ideal ways. There is a large middle ground here, and it is here that reasonable ideals are to be found. (Kornblith, 2001, p. 238)

A set of brakes may be upheld to an ideal—the functional ideal of stopping the vehicle upon being activated. But they may not be capable of doing so *across all environments*. In fact, one of the primary uses of upholding to teleological ideals is precisely to attune an entity to its environment. Move to Russia, and one may wish to invest in snow-tyres, as they will function better than other kinds of tyre.

So far, I have argued that some classificatory schemes—in particular proper functional schemes—serve a regulatory, rather than a predictive/explanatory purpose. In what remains of this thesis, I am going to argue that folk psychology is regulatory in this sense.

7.2.2 Dispositional ideals

I will now support

> *The Methodological Claim* Folk psychology is a regulatory conceptual scheme.

I will then argue that this allows for a dispositional individuation of belief.

According to the methodological claim, folk psychology is a regulatory theoretical construct. Why so? Here is my support: Sometimes, our beliefs and actions come apart. And when they come apart, we criticise and ascribe obligations. And when we are criticised for behaving as we ought not, we modify our behaviour. Jimmy asserts his belief that there is water south of the village. Upon becoming thirsty he wanders north, returning empty handed. One would expect, if they were to be informed, that the other members of Jimmy's community would criticise his behaviour. If you believed there was water to the south—why didn't you walk south? That's what believers that water is south are supposed to do! Similar points apply to the regulation of children's behaviour. You say you love your parents—*act like it*. Or consider the literature surrounding implicit bias. We may believe that men and women are equal, but our actions do not conform to the relevant norms—we treat them as though we did not. This does not tell us that we harbour sexist beliefs. It tells us at most that we do not act in accord with our feminist beliefs. You believe that men and women are equal—*act like it*. Prescriptions of this sort would be senseless if belief just were a behavioural disposition. One need not act in accord with one's beliefs to believe. But one will be criticised if one fails to do so. And that is, I suggest, good evidence that folk psychology is regulatory.

Before we turn to the second claim, I wish to consider a potential objection: If folk psychology is not primarily an explanatory/predictive theory, why have so many brilliant individuals thought that it is? The following passage from Craig comes to mind:

> To illustrate the point with an example of one which cannot without qualms be thought of in this way, we might suggest the wish to explain, in some fashion, the behaviour of one's fellows, or the wish to understand them in a way which makes them the same sort of being as oneself. (It might be thought, and has been suggested to me, that this idea could help us to see the concept of knowledge as some sort of theoretical construct, useful for explaining why other members of our community behave as they do.) But just how widespread this concern with explanation is, in particular whether it is widespread enough to fit our present bill, is very hard to say—thinking in these terms might just be a reflection of our contemporary obsession with the methods of the natural sciences. (Craig, 1990, p. 4)

Nevertheless, the question may be pushed further. It would be *highly surprising*, one may claim, if folk psychology were not primarily an explanatory/predictive theory. 'After all', it may be contended, 'that is what folk psychology does best!' At least, so says Jerry Fodor:

> Commonsense psychology works so well it disappears. It's like those mythical Rolls Royce cars whose engines are sealed when they leave the factory; only it's better because it isn't mythical. (Fodor, 1987, p. 3)

But Fodor is radically mistaken: Commonsense psychology works so well it disappears, *only in the cases in which it works*. And the cases in which it works are ones in which everything is 'as it ought to be'. To those who do not bear the relevant capacities—for those who are, for instance, suffering from certain psychological conditions—folk psychology is about as useful as guesswork. But when everything is going as it should, the conceptual scheme has predictive and explanatory power. Put another way—folk psychological explanation and prediction is predicated on an overarching assumption: That the relevant case is as it ought to be. Where the case is as it ought to be, where agents are able to regulate their behaviour according to the relevant norms, commonsense psychology may be employed to predict and explain their behaviour. But in lieu of those conditions holding, folk psychology is so useless it fades away.

Similar points apply to other regulatory schemes. I can explain how Harry got to Sally so quickly by citing that he rode a bicycle. But the given explanation relies on the bicycle operating properly. Similarly, I might explain how a digestive system works,

and in doing so I would assume it is not operating deviantly. It should be noted, however, that 'as it ought to be' is not solely an intrinsic matter. To explain that Harry got to Sally so quickly by citing his riding a bicycle has less force during a snowstorm than a summer's day. To operate properly, bicycles require certain conditions to obtain, just as biological traits do. The haemoglobin enter into an explanation of how oxygen is transported throughout the body, but not when the altitude is sufficiently high. Similarly, perceptual capacities cannot enter into explanations of valid representation where the environment is overly deceptive.

The upshot is that we need not deny that folk psychology has an important explanatory/predictive aspect. But we must accept that the explanatory and predictive force is predicated on auspicious circumstances obtaining. In fact, we may even tentatively suggest that the primary function of folk psychological ascriptions *is to instantiate conformity, so as to allow oneself and others to be predicted and explained*. Mind reading is predicated on a shared behavioural scheme. How else could we predict and explain the behaviour of complete strangers? How else could strangers predict and explain our own behaviour? Dispositional rogues and inauspicious circumstances result in a breakdown of predictive and explanatory success. And that's bad news, because it is deeply important that our behaviour can be predicted and explained.

That deals with the methodological claim. We are now in a position to support

> *The Individuation Claim* Belief-content pairs are individuated not by possessed dispositions, but rather by dispositional ideals.

On this view, *belief is individuated dispositionally, but by dispositional ideals rather than actually possessed dispositions*. In some respects, the account bears similarities to Lewis' account of linguistic convention, on which linguistic convention arises to solve what he calls *decision problems*.[8] He compares his view to one expressed by Hume in the following passage:

> a general sense of common interest; which sense all the members of the society express to one another, and which induces them to regulate their conduct by certain rules. (Hume, 1739, III.ii.2.)

My contention is, in short, as follows: By characterising ourselves and others as believers, desirers, and so on and so forth, we regulate our conduct by certain rules. And those rules amounts to *idealised dispositions*—dispositions an ideal agent would bear. To believe is *in part* to bear a dispositional property, but *also* to be upheld to a dispositional ideal. Explicitly put:

[8]See (Lewis, 1969).

Dispositional Ideals

1. *Possession Condition*

> For all propositions p there exists a unique dispositional ideal D_i, such that for all agents S, S enters a state of belief that p iff. S enters some dispositional state D, and is thereby upheld to D_i.

2. *Individuation Conditions*

> 2.1 *Sameness* Two beliefs, Bp, Bp* have the same content just in case they uphold their bearers to the same dispositional ideal D_i.
>
> 2.2 *Difference* Two beliefs, Bp, Bp* have distinct content just in case they uphold their bearers to distinct dispositional ideals D_i, D_i*.

Here is the view. There are certain conditions under which we enter dispositional states. For instance, when Doxy undergoes a perceptual experience of a bar of Toblerone, she is disposed to enter into a kind of dispositional state. Now, for any case c, if S enters into a novel cognitive disposition in c, then there exists some dispositional ideal such that S ought to have entered in c. Doxy should have become disposed to say 'yes, there is Toblerone there!' when asked, or to break off and gobble a chunk if overtaken by desire. It is this dispositional ideal that determines the content of the state, *not the actual dispositional state Doxy enters into*. To be upheld to the norm requires that one has entered a dispositional state, and so to believe that p requires that one is in *some* dispositional state or other. The point is that the content of the state is not—or at least not wholly—determined by the nature of that state, but by the dispositional standard to which one is upheld in virtue of entering that state.

Now, notice that this amounts to what I have called a *property norm*, which was defined as follows

<u>Property Norm</u>

ν is a property norm on F if and only if:

> N1*. $\Box$ (Fx only if ν is a standard on x).

What are the norms of which I speak? Roughly, to bear those dispositions the ideal agent would bear. Unlike the ideal agent of formal epistemology, however, our ideal agent is not logically omniscient. Rather, the ideal agent in this context simply behaves in accordance with the truisms and platitudes of folk psychology: with one important caveat—the ideal agent has no (or at least very few) inadequate powers of discrimination. An example may help to clarify. Suppose we take two agents, A

and B, and place them in a room. A and B are mental duplicates, except that where A believes *orange juice* (but not apple juice) is sweet, B believes *apple juice* (but not orange juice) is sweet. In the room there is a freshly squeezed glass of each. Next, they are asked to point to a glass that contains sweet liquid. What will our agents do? Obvious: A will point to the orange juice, B to the apple juice. That's what folk psychology predicts, but of course *they may not*. Rather, they *should*, given their beliefs. B may mistake orange juice for apple juice, or A *vice versa*, as they may lack the appropriate powers to discriminate the juice of oranges from the juice of apples.[9] But the dispositions relevant to the individuation of their beliefs are not those that arise under cases involving inadequate capacities of discrimination. An agent that accords with the ideal—ideal as that agent may be—is not hindered by such worries: they see oranges where there are oranges, apples where there are apples.

On the present view, the norms hold on *S* in virtue of *S*'s entering into a dispositional state in a certain context. It may be argued that these are property norms, but not functional norms. For recall that, on the view sketched, a functional norm is property norm set by etiological features of the relevant entity. It may be thought: surely *S*'s origins have *nothing* to do with what they believe? And if so, dispositional ideals of the relevant sort cannot be functional norms.

There are two replies that I can think of. The first would be to deny the objection. It may just be that, given your being the sort of entity that you are, you ought to enter into certain kinds of dispositional states in certain contexts. Similarly, ants given the kind of entity they are, should, upon finding a morsel of food, be disposed to leave scent trails. This is not, however, altogether satisfying. For it does *not* seem to be the agent's origins that matter, but rather the *context in which the belief is formed*, and whilst that does include backwards-reaching features, in particular in the case of historical reference, it is *not* the case that it is *S*'s origins that do the explanatory work.

The second, then, which if forced I would accept, would be to say that the dispositional ideal holds on the *property itself*. Thus, the property-norm would be higher order. The dispositional state *ought to be* a certain way. On this view, then, to say that *S* is upheld to an ideal is at best elliptical: to be precise we should say that *S*'s disposition is upheld to the relevant ideal. At first blush, that is dissatisfying, for the notion of a second-order norm seems dubious. But reflection reveals that it is perhaps more plausible than one may initially suppose. For analogy, consider an insect that enters a certain 'stance' when it detects a nearby predator, which puts it in some dispositional state (perhaps it becomes disposed to release venom when approached). We could perhaps call this the insect's 'offensive' stance. Now, it may be that not all of that kind of insect, when they enter into the offensive stance, end up bearing the disposition at issue. Perhaps, for instance, the venom does not release in certain individuals. Now,

[9] They may, for example, have achromatopsia (total colour blindness).

we can still identify 'the offensive' stance, and say that the stance *ought* to be a certain way. I.e., the dispositional state ought to have a certain character. So it doesn't seem implausible that property norms can hold on states. Moreover, what determines the way the state ought to be *is historical*, and can enter into etiological explanations of the state's formation. If I know that Jones believes *water* is wet, then I come to know about the causal origins of the relevant dispositional state. I am told that S believes that water is wet: by learning the standard to which S's dispositional property is upheld, I learn about how S came to bear that property.

It may be wondered: is this not proper functionalism in disguise? After all, if I am appealing to functional norms, have I not essentially individuated belief in terms of its *proper function*? Not obviously. For even proper functionalists endorse that commitment I have eschewed: for x to have a proper function, x must be able to function in such-and-such a way, and that requires that x be a spatiotemporal part. Consider Millikan:

> To take human beliefs to be intentional icons is to postulate that beliefs (explicit occurrent ones, at least) correspond to something physiological—neural structures, energy-transfer patterns, or whatever—these psychological devices having their own jobs to do. The performance of these jobs, when coupled with the performance of other jobs by devices that cooperate with beliefs, leads to the performance of further jobs... to be a belief involves having certain kinds of proper functions, and it is physiological structures or activities that *have* these proper functions. (Millikan, 1984, p. 138)

On my view, the proper functions are not purposes of agents' parts, but are rather ideals that hold on agents' dispositions. So if proper functionalism is committed to the possession of entities that occupy causal roles, then the account is not proper functional.

S believes that p only if S bears a dispositional property that is upheld to the relevant dispositional ideal. Why should we accept the view? I will offer two reasons for accepting the account. They are as follows. Firstly, it accounts for the difference in content between Twin-Earth counterparts. Secondly, it accounts for the force of folk psychological predictions and explanations. I will take each in turn.

The account yields the correct verdict across twin-earth cases. Both Earthy and Twinny are dispositional duplicates: They are in distinct mental states not because of their dispositions differing, but because *they are upheld to distinct dispositional ideals*. And the reason is that the states are formed in distinct contexts. Put another way: dispositional ideals account for both the difference, and the sameness of belief-content pairs. If the account yields the correct verdict in twin-earth cases, then such agents are upheld to distinct dispositional ideals. That raises the question: in what way do the dispositional ideals of our troublesome duo differ?

Well, suppose that both Earthy and Twinny are transported to a planet foreign to both. We place Earthy at the North pole, and Twinny at the South pole, and for purposes of idealisation assume that the environments are identical in all relevant regards. Now, let us suppose that the rivers are filled with XYZ. In this case, the manifestations of Twinny's dispositions will accord with her ideal: Upon desiring XYZ, she will move to XYZ and quench her thirst. But Earthy's manifestations will *not*—upon experiencing desire for H_20, Earthy will be taken to some other substance—XYZ. We could put this in alternative terms. Whilst it will be *entirely accidental that Earthy's intention to quench her thirst is satisfied, it will be entirely non-accidental that Twinny's intention is satisfied*. And that is because to be non-accidental in the relevant sense requires satisfying one's dispositional ideals. And the dispositional ideals of our agents differ. Analogously, it will be entirely accidental if a waste disposal unit functions as a heart, though it may well do so. What determines whether one believes is not what dispositions one holds, but rather the conditions under which the manifestations of one's dispositions will result in the accidental and/or nonaccidental achievement of one's aims. And that is determined by the ideals to which one is upheld. This is good news, not only because we get the right verdict, but because we avoid the problem raised against Fodor: We no longer require that all believers that *p* are, *ceteris paribus*, identical with respect to their dispositional properties. To believe is to be in some dispositional state or other, but it need not have some unique causal basis, as Fodor requires. Thus, one can bear water and twin-water beliefs without the pains of contradiction. One can believe that women are equal to men without acting like it, just as, most unfortunately, we should expect.

That deals with the first motivation. What about the second? How does the view account for the force of folk psychological explanation and prediction? Mental states can enter into psychological explanations, for to be in a certain state requires being upheld to a certain kind of ideal, and to be upheld to that kind of ideal implies that one entered a dispositional state in some case *c*, such that *c* is a member of the set of cases in which agents are upheld to the relevant norm. To be counted as a believer is to be counted amongst a wide range of agents who enter into states in similar circumstances. And typically, agents in those circumstances—or at least the subset in which everything is 'as it ought to be'—are disposed to behave in accord with those norms. Thus, *in virtue of the regulatory practice occurring, the behaviour of agents can be predicted and explained*.

This kind of explanation differs dramatically from the type functionalists are liable to give. On a functionalist account, to explain with folk psychological concepts is to provide details about the spatiotemporal causes of the relevant event's occurrence. But on the present view—on which mental states are conditions, not the occupants of causal roles—the explanations hold due to probabilistic relations that hold between being in

that condition and the event's occurrence. To the explanation seeking question:

Q. Why did Jones shoot Smith?

The explanatory force of

E. Jones believed that Smith was trying to kill him.

derives not from the fact that Jones' belief *caused* his finger to pull the trigger, but from the fact that agents who enter dispositional states that are upheld to the same norms are more likely to behave in that way, so long as everything is 'as it ought to be'. In that sense, folk psychological explanations and predictions are more general and more idealised than functionalists have supposed.

A lingering worry remains, though. What anchors the norms, i.e., dispositional ideals to agents? Evidently, we will need some account to hold, if the view is to bear up to scrutiny. But like Fodor, I answer: The context, and determining the nature of *that* is not my problem. We *know*, by reflection on cases, what an agent ought to believe. If in doubt, just take an agent in a case and ask: What *do* they believe? Your folk intuitions will do the work for you.

Folk psychology is not a purely descriptive theoretical construct—it is a prescriptive tool. To count as a believer is not to bear a specific dispositional profile—though typically one will believe that p only if one does, roughly at least. Rather, to count as a believer is to be *upheld against a certain kind of dispositional ideal*. Those dispositional ideals are a kind of functional norm. Belief is a dispositional state. But the content of that belief does not hinge on the dispositional nature of that state. These remarks as mentioned before, do not constitute a finished model, nor have I shown they are true. Rather, they form but a sketch of how one might individuate belief in dispositional terms. Whether this is viable remains to be seen. The devil will be, no doubt, in the details. But that is a task for another time.

In this section, I outlined and defended a regulatory theory of belief-content individuation. That deals with [7.3]. The account is admittedly sketchy. But if true, we can individuate belief in dispositional terms. The lesson is: To do so, we must make use of dispositional ideals, not actually possessed dispositional properties.

Conclusion

In this chapter, I had the following aims:

[7.1] To outline the argument for content externalism, and for its inconsistency with *Complex Behaviourism*.

[7.2] To reject a number of context-relative responses to the problem.

[7.3] To outline a regulatory theory of belief-content individuation.

which have now been satisfied.

References

Aizawa, Kenneth. 2003. Cognitive Architecture: The Structure of Cognitive Representations. *Pages 172–189 of:* Stich, Stephen P., & Warfield, Ted A. (eds), *The Blackwell Guide to Philosophy of Mind*. Blackwell.

Alan, Baker. 2003. Quantitative Parsimony and Explanatory Power. *British Journal for the Philosophy of Science*, **54**(2), 245–259.

Alston, William P. 1971. Dispositions, Occurrences, and Ontology. *Pages 359–88 of:* Tuomela, R. (ed), *Dispositions*. Reidel.

Armstrong, David M. 1968. *A Materialist Theory of the Mind*. Routledge.

Armstrong, David M. 1969. Dispositions Are Causes. *Analysis*, **30**(1), 23–26.

Armstrong, David M. 1973. *Belief, Truth and Knowledge*. Cambridge University Press.

Armstrong, David M. 1981. The Causal Theory of the Mind. *In:* Chalmers, David J. (ed), *The Nature of Mind and Other Essays*. Cornell University Press.

Armstrong, David M. 1989. *A Combinatorial Theory of Possibility*. Cambridge University Press.

Armstrong, David M. 1997. *A World of States of Affairs*. Cambridge University Press.

Armstrong, David M. 2004. *Truth and Truthmakers*. Cambridge University Press.

Audi, Paul. 2013. How to Rule Out Disjunctive Properties. *Noûs*, **47**(4), 748–766.

Audi, Robert. 1994. Dispositional Beliefs and Dispositions to Believe. *Noûs*, **28**(4), 419–434.

Barth, John. 1968. *Lost in the Funhouse*. Doubleday.

Bauer, Mark. 2009. Normativity Without Artifice. *Philosophical Studies*, **144**(2), 239–259.

Bennett, Jonathan. 1988. *Events and Their Names*. Hackett.

Bird, Alexander. 1998. Dispositions and Antidotes. *Philosophical Quarterly*, **48**(191), 227–234.

Bird, Alexander. 2007. *Nature's Metaphysics: Laws and Properties*. Oxford University Press.

Bird, Alexander. 2009a. Essences and Natural Kinds. *Pages 497–506 of:* Poidevin, Robin Le (ed), *The Routledge Companion to Metaphysics*. Routledge.

Bird, Alexander. 2009b. Structural Properties Revisited. *Pages 215–41 of:* Handfield, Toby (ed), *Dispositions and Causes*. Clarendon Press.

Bird, Alexander. 2016. Overpowering: How the Powers Ontology Has Overreached Itself. *Mind*, **125**(498), 341–383.

Bird, Alexander. 2018. I—Fundamental Powers, Evolved Powers, and Mental Powers. *Aristotelian Society Supplementary Volume*, **92**(1), 247–275.

Block, Ned. 1978. Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science*, **9**, 261–325.

Block, Ned, & Fodor, Jerry A. 1972. What Psychological States Are Not. *Philosophical Review*, **81**(April), 159–81.

Bonevac, Daniel, Dever, Joshua, & Sosa, David. 2011. The Counterexample Fallacy. *Mind*, **120**(480), 1143–1158.

Brenner, Andrew. 2017. Simplicity as a Criterion of Theory Choice in Metaphysics. *Philosophical Studies*, **174**(11), 2687–2707.

Brentano, Franz. 1874. *Psychology From an Empirical Standpoint*. Routledge.

Broome, John. 2013. *Rationality Through Reasoning*. Wiley-Blackwell.

Buller, David J. 1998. Etiological Theories of Function: A Geographical Survey. *Biology and Philosophy*, **13**(4), 505–527.

Burge, Tyler. 2003. Perceptual Entitlement. *Philosophy and Phenomenological Research*, **67**(3), 503–48.

Burge, Tyler. 2007. *Foundations of Mind*. Oxford University Press.

Burge, Tyler. 2010. *Origins of Objectivity*. Oxford University Press.

Carnap, Rudolf. 1936. Testability and Meaning. *Philosophy of Science*, **3**(4), 419–471.

Carnap, Rudolf. 1956. The Methodological Character of Theoretical Concepts. *Minnesota Studies in the Philosophy of Science*, **1**(1), 38–76.

Chalmers, David J. 2002. On Sense and Intension. *Philosophical Perspectives*, **16**(s16), 135–82.

Chalmers, David J. 2011. Propositions and Attitude Ascriptions: A Fregean Account. *Noûs*, **45**(4), 595–639.

Choi, Sungho. 2008. Dispositional Properties and Counterfactual Conditionals. *Mind*, **117**(468), 795–841.

Contessa, Gabriele. 2013. Dispositions and Interferences. *Philosophical Studies*, **165**(2), 401–419.

Cowling, Sam. 2013. Ideological Parsimony. *Synthese*, **190**(17), 3889–3908.

Craig, Edward. 1990. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford University Press.

Crane, Tim. 1996. *Dispositions: A Debate*. New York: Routledge.

Crane, Tim. 2001. *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford University Press.

Crane, Tim. 2003. *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representation*. Routledge.

Cummins, Robert C. 1975. Functional Analysis. *Journal of Philosophy*, **72**(November), 741–64.

Cummins, Robert C. 1983. *The Nature of Psychological Explanation*. MIT Press.

Davidson, Donald. 1969a. The Individuation of Events. *Pages 216–34 of:* Rescher, Nicholas (ed), *Essays in Honor of Carl G. Hempel*. Reidel.

Davidson, Donald. 1969b. On Events and Event-Descriptions. *Pages 74–84 of:* Margolis, Joseph (ed), *Fact and Existence*. Blackwell.

Davidson, Donald. 1970. Mental Events. *Pages 207–224 of:* Foster, L., & Swanson, J. W. (eds), *Essays on Actions and Events*. Clarendon Press.

Davidson, Donald. 1978. Intending. *Philosophy of History and Action*, **11**, 41–60.

Davidson, Donald. 1982. Rational Animals. *Dialectica*, **36**(4), 317–28.

Davidson, Donald. 1987. Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, **60**(3), 441–458.

Donnellan, Keith S. 1966. Substitution and Reference. *Journal of Philosophy*, **63**(21), 685–688.

Donnellan, Keith S. 1970. Proper Names and Identifying Descriptions. *Synthese*, **21**(3-4), 335–358.

Donnellan, Keith S. 1974. Speaking of Nothing. *Philosophical Review*, **83**(1), 3–31.

Dretske, Fred. 1988. *Explaining Behaviour: Reasons in a World of Causes*. MIT Press.

Elder, Crawford L. 2007. Conventionalism and the World as Bare Sense-Data. *Australasian Journal of Philosophy*, **85**(2), 261–275.

Evans, Gareth. 1973. The Causal Theory of Names. *Aristotelian Society Supplementary Volume*, **47**(1), 187–208.

Fara, Michael. 2005. Dispositions and Habituals. *Noûs*, **39**(1), 43–82.

Fine, Kit. 1994. Senses of Essence. *In:* Sinnott-Armstrong, Walter, Raffman, Diana, & Asher, Nicholas (eds), *Modality, Morality and Belief. Essays in Honor of Ruth Barcan Marcus*. Cambridge University Press.

Fine, Kit. 1995. The Logic of Essence. *Journal of Philosophical Logic*, **24**(3), 241–273.

Fisher, Justin C. 2013. Dispositions, Conditionals and Auspicious Circumstances. *Philosophical Studies*, **164**(2), 443–464.

Fodor, J. A. 1985. Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum. *Mind*, **94**(373), 76–100.

Fodor, Janet D. 1979. *The Linguistic Description of Opaque Contexts*. Garland.

Fodor, Jerry A. 1968. *Psychological Explanation: An Introduction To The Philosophy Of Psychology*. Ny: Random House.

Fodor, Jerry A. 1975. *The Language of Thought*. Harvard University Press.

Fodor, Jerry A. 1981. *Representations: Philosophical Essays on the Foundations of Cognitive Science*. MIT Press.

Fodor, Jerry A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.

Fodor, Jerry A. 1990. *A Theory of Content and Other Essays*. MIT Press.

Fodor, Jerry A., & Pylyshyn, Zenon W. 1988. Connectionism and Cognitive Architecture. *Cognition*, **28**(1-2), 3–71.

Foot, Philippa. 2001. *Natural Goodness*. Oxford University Press.

Foot, Philippa, & Montefiore, Alan. 1961. Goodness and Choice. *Aristotelian Society Supplementary Volume*, **35**(1), 45–80.

Forbes, Graeme. 1985. The Metaphysics of Modality. *Mind*, **95**(377), 135–138.

Forbes, Graeme. 1986. In Defense of Absolute Essentialism. *Midwest Studies in Philosophy*, **11**(1), 3–31.

Frege, Gottlob. 1892. Über Sinn Und Bedeutung. *Zeitschrift für Philosophie Und Philosophische Kritik*, **100**(1), 25–50.

Frege, Gottlob. 1948. Sense and Reference. *Philosophical Review*, **57**(3), 209–230.

Friedman, Michael. 1974. Explanation and Scientific Understanding. *Journal of Philosophy*, **71**(1), 5–19.

Fumerton, Richard. 2001. Epistemic Justification and Normativity. *In:* Steup, M (ed), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford University Press.

Gibbard, Allan. 2012. *Meaning and Normativity*. Oxford University Press UK.

Gibbs, Benjamin. 1969. Putnam on Brains and Behaviour. *Analysis*, **30**(December), 53–55.

Godfrey-Smith, Peter. 1993. Functions: Consensus Without Unity. *Pacific Philosophical Quarterly*, **74**(3), 196–208.

Godfrey-Smith, Peter. 1994. A Modern History Theory of Functions. *Noûs*, **28**(3), 344–362.

Godfrey-Smith, Peter. 2012. *Complexity and the Function of Mind in Nature*. Cambridge University Press.

Goodman, Nelson. 1954. *Fact, Fiction & Forecast*. University of London.

Goodman, Nelson. 1961. Safety, Strength, Simplicity. *Philosophy of Science*, **28**(2), 150–151.

Graham, Peter J. 2014. Functions, Warrant, History. *Pages 15–35 of:* Fairweather, Abrol, & Flanagan, Owen (eds), *Naturalizing Epistemic Virtue*. Cambridge University Press.

Greco, John. 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.

Grünbaum, Adolf. 2007. Is Simplicity Evidence of Truth? *Royal Institute of Philosophy Supplement*, **82**(61), 179–189.

Hanke, David. 2004. Teleology: The Explanation That Bedevils Biology. *Pages 143–155 of:* Cornwell, John (ed), *Explanations: Styles of Explanation in Science*. Oxford University Press.

Hardin, Clyde L. 1988. *Color for Philosophers: Unweaving the Rainbow*. Hackett.

Harman, Gilbert. 1973. *Thought*. Princeton University Press.

Hawthorne, John, & Manley, David. 2005. Review of Stephen Mumford's Dispositions. *Noûs*, **39**(1), 179–95.

Heil, John. 1999. Multiple Realizability. *American Philosophical Quarterly*, **36**(3), 189–208.

Heil, John. 2004. Natural Intentionality. *In:* Schantz, Richard (ed), *The Externalist Challenge*. De Gruyter.

Hitchcock, Christopher. 2015. *Lewis on Causation*. John Wiley & Sons, Ltd. Pages 295–311.

Hornsby, Jennifer. 1980. *Actions*. Routledge and Kegan Paul.

Hume, David. 1739. *A Treatise on Human Nature*. Oxford University Press.

Hume, David. 1965. *An Abstract of a Treatise of Human Nature, 1740*. Hamden, Conn., Archon Books.

Hurley, Susan L., & Noë, Alva. 2003. Neural Plasticity and Consciousness. *Biology and Philosophy*, **18**(1), 131–168.

Johnston, Mark. 1992. How to Speak of the Colors. *Philosophical Studies*, **68**(3), 221–263.

Kading, Daniel. 1959. Are There Really "No Duties to Oneself"? *Ethics*, **70**(2), 155–157.

Kant, Immanuel, & Muller, Max. 1896. *Immanuel Kant's Critique of Pure Reason, Tr. By F.M. Müller*.

Kelly, Sean D. 2001. The Non-Conceptual Content of Perceptual Experience: Situation Dependence and Fineness of Grain. *Philosophy and Phenomenological Research*, **62**(3), 601–608.

Kim, Jaegwon. 1989. The Myth of Non-Reductive Materialism. *Proceedings and Addresses of the American Philosophical Association*, **63**(3), 31–47.

Kim, Jaegwon. 1993. *Supervenience and Mind*. Cambridge University Press.

Kitcher, Philip. 1981. Explanatory Unification. *Philosophy of Science*, **48**(4), 507–531.

Kitcher, Philip, & Salmon, Wesley. 1989. *Scientific Explanation*. Univ of Minnesota Pr.

Kornblith, Hilary. 2001. Epistemic Obligation and the Possibility of Internalism. *Pages 231–248 of:* Fairweather, Abrol, & Zagzebski, Linda Trinkaus (eds), *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*. Oxford University Press.

Kripke, Saul A. 1971. Identity and Necessity. *Pages 135–164 of:* Munitz, Milton K. (ed), *Identity and Individuation*. New York University Press.

Kripke, Saul A. 1980. *Naming and Necessity*. Harvard University Press.

Lackey, Jennifer. 2007. Norms of Assertion. *Noûs*, **41**(4), 594–626.

LaPorte, Joseph. 2004. Rigid Designators for Properties. *Philosophical Studies*, **130**(2), 321–336.

Lean, Oliver M. 2014. Getting the Most Out of Shannon Information. *Biology and Philosophy*, **29**(3), 395–413.

Lewis, David. 1966. An Argument for the Identity Theory. *Journal of Philosophy*, **63**(1), 17–25.

Lewis, David. 1969. *Convention: A Philosophical Study*. Wiley-Blackwell.

Lewis, David. 1973. *Counterfactuals*. Blackwell.

Lewis, David. 1976. The Paradoxes of Time Travel. *American Philosophical Quarterly*, **13**(2), 145–152.

Lewis, David. 1986a. Causal Explanation. *Pages 214–240 of:* Lewis, David (ed), *Philosophical Papers Vol. Ii*. Oxford University Press.

Lewis, David. 1986b. *On the Plurality of Worlds*. Wiley-Blackwell.

Lewis, David. 1997. Finkish Dispositions. *Philosophical Quarterly*, **47**(187), 143–158.

Lipton, Peter. 2004. *Inference to the Best Explanation*. Routledge/Taylor and Francis Group.

Lowe, E. J. 2003. Individuation. *In:* Loux, Michael J., & Zimmerman, Dean W. (eds), *The Oxford Handbook of Metaphysics*. Oxford University Press.

Lycan, William G. 1986. Tacit Belief. *In:* Bogdan, R. (ed), *Belief: Form, Content, and Function*. Oxford University Press.

Mackie, J. L. 1974. *The Cement of the Universe*. Oxford, Clarendon Press.

Mackie, John L. 1972. *Truth, Probability and Paradox: Studies in Philosophical Logic*. Oxford, Clarendon Press.

Manley, David, & Wasserman, Ryan. 2008. On Linking Dispositions and Conditionals. *Mind*, **117**(465), 59–84.

Manley, David, & Wasserman, Ryan. 2011. Dispositions, Conditionals, and Counterexamples. *Mind*, **120**(480), 1191–1227.

Marmodoro, Anna. 2010. *The Metaphysics of Powers: Their Grounding and Their Manifestations*. Routledge.

Martin, Charlie B. 1994. Dispositions and Conditionals. *Philosophical Quarterly*, **44**(174), 1–8.

Martin, Charlie B., & Pfeifer, Karl. 1986. Intentionality and the Non-Psychological. *Philosophy and Phenomenological Research*, **46**(June), 531–54.

McDowell, John. 1994. *Mind and World*. Cambridge: Harvard University Press.

McDowell, John. 1998. *Mind, Value, and Reality*. Harvard University Press.

McGinn, Colin. 1976. On the Necessity of Origin. *Journal of Philosophy*, **73**(5), 127–135.

McGrew, Timothy. 2003. Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the Philosophy of Science*, **54**(4), 553–567.

Meckler, Lester. 1953. Normative and Descriptive Expressions. *The Journal of Philosophy*, **50**(19), 577–583.

Mellor, David H. 1974. In Defense of Dispositions. *Philosophical Review*, **83**(2), 157–181.

Millikan, Ruth G. 1984. *Language, Thought and Other Biological Categories*. MIT Press.

Millikan, Ruth G. 1993. *Spatial Representation*. Cambridge: Blackwell.

Millikan, Ruth G. 2004. *Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT Press.

Molnar, George. 2003. *Powers: A Study in Metaphysics*. Oxford University Press.

Moore, George E. 1922. The Nature of Moral Philosophy. *In: Philosophical papers*. Routledge and Kegan Paul.

Morreau, Michael. 1997. Fainthearted Conditionals. *Journal of Philosophy*, **94**(4), 187.

Mumford, Stephen. 1998. *Dispositions*. Clarendon Press.

Mumford, Stephen, & Anjum, Rani Lill. 2011. *Getting Causes From Powers*. Oxford University Press.

Myrvold, Wayne C. 2003. A Bayesian Account of the Virtue of Unification. *Philosophy of Science*, **70**(2), 399–423.

Nanay, Bence. 2010. A Modal Theory of Function. *Journal of Philosophy*, **107**(8), 412–431.

Neander, Karen. 1991a. Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of Science*, **58**(2), 168–184.

Neander, Karen. 1991b. The Teleological Notion of 'Function'. *Australasian Journal of Philosophy*, **69**(4), 454–468.

Noonan, Harold W. 1978. Sortal Concepts and Identity. *Mind*, **87**(346), 267–269.

Noonan, Harold W. 1979. Rigid Designation. *Analysis*, **39**(4), 174–182.

Okasha, Samir. 2002. Darwinian Metaphysics: Species and the Question of Essentialism. *Synthese*, **131**(2), 191–213.

Papineau, David. 1987. *Reality and Representation*. Blackwell.

Patrick, Kit. forthcoming. Unity As An Epistemic Virtue. *Erkenntnis*, 1–20.

Peacocke, Christopher. 1979. *Holistic Explanation*. Clarendon Press.

Peacocke, Christopher. 1983. *Sense and Content: Experience, Thought and Their Relations*. Oxford University Press.

Peacocke, Christopher. 1992. *A Study of Concepts*. MIT Press.

Peacocke, Christopher. 2001. Does Perception Have a Nonconceptual Content? *Journal of Philosophy*, **98**(5), 239–264.

Place, Ullin T. 1996. Intentionality as the Mark of the Dispositional. *Dialectica*, **50**(2), 91–120.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford University Press.

Price, Carolyn. 2001. *Functions in Mind: A Theory of Intentional Content*. Oxford University Press.

Prior, Elizabeth, Pargetter, Robert, & Jackson, Frank. 1982. Three Theses About Dispositions. *American Philosophical Quarterly*, **19**(3), 251–257.

Prior, Elizabeth W. 1985. *Dispositions*. Humanities Press.

Pritchard, Duncan. 2012. Anti-Luck Virtue Epistemology. *Journal of Philosophy*, **109**(3), 247–279.

Putnam, Hilary. 1963. Brains and Behavior. *In:* Butler, Ronald J. (ed), *Analytical Philosophy: Second Series*. Blackwell.

Putnam, Hilary. 1975. The Meaning of 'Meaning'. *Minnesota Studies in the Philosophy of Science*, **7**, 131–193.

Quine, Willard V. 1951. Two Dogmas of Empiricism. *Philosophical Review*, **60**(1), 20–43.

Quine, Willard V. 1956. Quantifiers and Propositional Attitudes. *Journal of Philosophy*, **53**(5), 177–187.

Quine, Willard V. 1960. *Word and Object*. The MIT Press.

Quine, Willard V. 1970. *The Web of Belief*. New York: Random House.

Robinson, William S. 1990. States and Beliefs. *Mind*, **99**(393), 33–51.

Rundle, Bede. 1997. *Mind in Action*. Oxford University Press.

Russell, Bertrand. 1919. On Propositions: What They Are and How They Mean. *Proceedings of the Aristotelian Society*, **2**, 1–43.

Ryle, Gilbert. 1949. *The Concept of Mind*. Hutchinson & Co.

Schwartz, Peter H. 1999. Proper Function and Recent Selection. *Philosophy of Science*, **66**(3), 210–222.

Segal, Gabriel. 2004. Reference, Causal Powers, Externalist Intuitions, and Unicorns. *Page 329 of:* Schantz, Richard (ed), *The Externalist Challenge*. De Gruyter.

Sellars, Wilfrid. 1962. Philosophy and the Scientific Image of Man. *Pages 35–78 of:* Colodny, Robert (ed), *Science, Perception, and Reality*. Humanities Press/Ridgeview.

Shoemaker, Sydney. 1980. Causality and Properties. *Pages 109–35 of:* van Inwagen, Peter (ed), *Time and Cause*. D. Reidel.

Shoemaker, Sydney. 1981. Some Varieties of Functionalism. *Philosophical Topics*, **12**(1), 93–119.

Shoemaker, Sydney. 2007. *Physical Realization*. Oxford University Press UK.

Silverstein, Matthew. 2016. Teleology and Normativity. *Oxford Studies in Metaethics*, **11**, 214–240.

Smith, A. D. 1977. Dispositional Properties. *Mind*, **86**(343), 439–445.

Smith, Michael. 1987. The Humean Theory of Motivation. *Mind*, **96**(381), 36–61.

Smith, Michael. 1994. *The Moral Problem*. Blackwell.

Soames, Scott. 1992. Truth, Meaning, and Understanding. *Philosophical Studies*, **65**(1-2), 17–35.

Squires, Roger. 1968. Are Dispositions Causes? *Analysis*, **29**(2), 45–47.

Squires, Roger. 1970. Are Dispositions Lost Causes? *Analysis*, **31**(1), 15–18.

Stalnaker, Robert. 1968. *Studies in Logical Theory*. Oxford: Blackwell.

Stevenson, Leslie. 1969. Are Dispositions Causes? *Analysis*, **29**(6), 197–199.

Steward, Helen. 1997. *The Ontology of Mind: Events, Processes, and States*. Oxford University Press.

Storer, Thomas. 1951. On Defining 'Soluble'. *Analysis*, **11**(6), 134–37.

Tappolet, Christine, & Rossi, Mauro. 2015. What is Value? Where Does It Come From? A Philosophical Perspective. *Pages 3–22 of:* Brosch, Tobias, & Sander, David (eds), *The Value Handbook: The Affective Sciences of Values and Valuation*.

Thomson, Judith Jarvis. 2008. *Normativity*. Open Court.

Thomson, Judith Jarvis. 2010. Normativity. *Analysis*, **70**(4), 713–715.

Tye, Michael. 2006. Nonconceptual Content, Richness, and Fineness of Grain. *Pages 504–30 of:* Gendler, Tamar S., & Hawthorne, John (eds), *Perceptual Experience*. Oxford University Press.

Vetter, Barbara. 2014. Dispositions Without Conditionals. *Mind*, **123**(489), 129–156.

von Wright, Georg H. 1963. *Norm and Action*. New York: Humanities.

Wedgwood, Ralph. 2006. The Internal and External Components of Cognition. *Pages 307–325 of:* Stainton, Robert J. (ed), *Contemporary Debates in Cognitive Science*. Malden MA: Blackwell.

Wedgwood, Ralph. 2007. *The Nature of Normativity*. Oxford University Press.

Willard, Mary B. 2014. Against Simplicity. *Philosophical Studies*, **167**(1), 165–181.

Williamson, Timothy. 1996. Knowing and Asserting. *The Philosophical Review*, **105**(4), 489–523.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.

Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Blackwell Pub.

Williamson, Timothy. 2009. Replies to Critics. *Pages 279–384 of:* Pritchard, Duncan, & Greenough, Patrick (eds), *Williamson on Knowledge*. Oxford: Oxford University Press.

Williamson, Timothy. 2013. How Deep is the Distinction Between A Priori and A Posteriori Knowledge? *Pages 291–312 of:* Casullo, Albert, & Thurow, Joshua C. (eds), *The A Priori in Philosophy*. Oxford University Press.

Williamson, Timothy. 2016. Modal Science. *Canadian Journal of Philosophy*, **46**(4-5), 453–492.

Williamson, Timothy. forthcoming. Gibbard on Meaning and Normativity. *Inquiry: An Interdisciplinary Journal of Philosophy*, 1–11.

Wright, Larry. 1973. Functions. *Philosophical Review*, **82**(2), 139–168.

Wright, Wayne. 2003. Projectivist Representationalism and Color. *Philosophical Psychology*, **16**(4), 515–529.

Zeman, J. Jay. 1973. *Modal Logic: The Lewis-Modal Systems*. London: Clarendon Press.