# A Defence of Dispositionalism

## Gregory Ashley Owens

Submitted in accordance with the requirements
for the degree of Doctor of Philosophy

University of Leeds

School of Philosophy, Religion and the History of Science

March 2014

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2014 The University of Leeds and Gregory Ashley Owens.

The right of Gregory Ashley Owens to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

# Acknowledgements

## Abstract

In this thesis, I develop and defend a distinctive version of a position I call (following Schwitzgebel (2002)) *phenomenal dispositionalism*. On this view, having such-and-such beliefs, desires, character traits etc. is just a matter of having such-and-such behavioural, cognitive and phenomenal dispositions; dispositions, roughly, to act, think and feel thus-and-so in such-and-such circumstances.

Phenomenal dispositionalism has its roots in Ryle (2000) (who, I argue, is no behaviourist). Just as Ryle frames his position as an alternative to the Cartesian 'Official Doctrine' of his day, I frame mine as an alternative to what Baker (1995) calls the 'Standard View' in contemporary philosophy of mind (roughly, the view that mental states are brain states). In Baker's view and in mine, Standard View theorists repeat the Cartesian error of construing the mind as a causal system.

I attack this error at what I take to be its root, arguing (*contra* Mumford (1998)) that disposition ascription does not and cannot work by picking out particular internal properties or states of the object of ascription, occupying particular causal roles. Nonetheless, I argue, disposition ascriptions (including mental state ascriptions) can explain - and (*pace* Ryle) explain *causally*.

The role of 'folk psychological' language, I argue, is not to pick out internal states occupying particular causal roles. Nor (*pace* Schwitzgebel) is it to assert subjects' conformity to 'dispositional stereotypes' for each individual mental state ascribed to them. Rather, it is holistically to describe subjects' *dispositional profiles* - their overall sets of behavioural, phenomenal and cognitive dispositions.

I argue that our rich, everyday mental-state taxonomy is fit for this purpose, and stands in no need of revision either by those who are inclined to boil it down to beliefs and desires, or those who posit 'aliefs' in order to fill the explanatory gaps this leaves us with (Gendler, 2008a).

# Contents

# 1 Introduction: Why Dispositionalism?

## 1.1 Introduction

Rylean dispositionalism - the thesis that for a person to be in a mental state such as *believing that P* is just for her to have certain dispositions - has long been conflated with the position variously known as logical, analytic, or philosophical behaviourism.[1] This is something Ryle saw coming, predicting that his theory would 'undoubtedly, and harmlessly, be stigmatized as "behaviourist"' (2000, p.308). In my view, there has been nothing 'harmless' about this stigmatization; it has meant that the failure of behaviourism has been seen as the failure of dispositionalism.

But of course, dispositionalism only amounts to behaviourism if one thinks that among all a subject's dispositions, only her *behavioural* dispositions are relevant to the question of what mental states may be ascribed to her. In this thesis, I will develop and defend a distinctive version of a more liberal position - a position I call (following Schwitzgebel (2002)[2]) *phenomenal dispositionalism*. On this view, it is not just a subject's *behavioural* dispositions that matter for the purposes of ascribing mental states to her, but also her *phenomenal* and *cognitive* dispositions - her dispositions, respectively, to have certain sorts of conscious experience and to reason in certain ways.[3]

---

[1]These terms have generally been employed in order to distinguish this strand of philosophical thought from a broadly contemporary, related, but distinct strand of thought in psychology, also known as 'behaviourism' (or, in its Skinnerian form, 'radical behaviourism') (Skinner, 1953; Watson, 1930; 1913). When I talk simply about 'behaviourism' in this thesis, it is *philosophical* behaviourism I have in mind.

[2]Although Schwitzgebel's project is in many ways the starting point for my own, it is not my intention simply to defend phenomenal dispositionalism as he formulates it. Indeed, I will argue in Chapter 6 that Schwitzgebel's formulation of the position is untenable, due to fundamental problems with the use he makes of the notion of *dispositional stereotypes* for the possession of particular mental states.

[3]For convenience, I will sometimes talk about behavioural, cognitive and phenomenal dispositions as dispositions to 'act, think and feel' a certain way. This loose terminology should be taken with a pinch of salt; no doubt there are phenomenal features of mental life that have at least as much to do with 'thinking' as with 'feeling' (e.g. consciously working out a maths problem in one's head).

# *1 Introduction: Why Dispositionalism?*

The position I am advocating is at odds with what Lynne Rudder Baker justifiably calls the 'Standard View' in contemporary philosophy of mind - the view that 'the attitudes, if there are any, are (or are constituted by, or are realized in) particular brain states' (1995, p.5). If I am right, it is a mistake to think - even if physicalism is true[4] - that the beliefs and desires (etc.) we routinely ascribe to subjects can only be said to exist if they can be identified with (or be said to be constituted by, or realized in) particular brain states of those subjects.

## 1.2 Dispositionalism and eliminativism

The sceptical reader may find himself wondering why we should trouble ourselves even to consider abandoning what has been a hard-won, albeit partial, consensus in the philosophy of mind in order to pursue a largely-abandoned line of thinking. By way of motivation, then, I want to suggest that dispositionalism may very well prove to be our last line of defence against a claim that really *would* shake up our consensus view of the mind: the eliminativist claim that there are no such things as beliefs, desires, and so on, with the consequence that ascriptions to subjects of such states are uniformly false.

Why should we think so? Consider an idea that is part and parcel of the Standard View, an idea sometime called the 'theory theory'. According to the theory theorist, beliefs, desires and the like are entities posited by a commonsense or 'folk' theory of the causal processes responsible for the production of behaviour. If we take this idea seriously, we must - in my view - also take seriously the possibility that the theory in question, so-called 'folk psychology', is fundamentally mistaken.

For the claims made by folk psychology, thus understood as a quasi- or proto-scientific theory, are far from modest. This theory does not restrict itself to claims about subjects' observable behaviour - claims, broadly, about how subjects who have in the past acted thus-and-so may be expected to act if exposed to such-and-such stimuli in such-and-such circumstances. Nor does it restrict itself to treating the mind as a 'black box' - asserting *that* the mind plays such-and-such a role in generating certain behavioural outputs in response to certain environmental inputs, but refraining from making assertions about *how* it plays that role. On the contrary, it makes bold assertions about the functional organization of the mind - about the mental processes causally mediating between stimulus and response. It provides us with a quasi- or proto-scientific

---

[4]I take physicalism to be the thesis that everything supervenes on the physical.

taxonomy of state types and tells us how particular token states are brought into being through exposure to certain stimuli, how they interact with one another, and how they ultimately cause subjects to act as they do.

It tells us, for instance, that a subject who has an ice cream placed in front of him will typically acquire a particular, individuable token state, a belief that there is an ice cream in front of him. It tells us that this token state will then interact causally with various other token states of the subject, so that if (for instance) the subject has tokens of the belief that ice creams melt quickly on hot days and of the belief that it is a hot day today, he will acquire a token of the belief that the ice cream in front of him will melt quickly. It tells us that if the subject has a token of the desire to eat a non-melted ice cream, he will acquire a token of the desire to eat the ice cream in front of him quickly, before it melts. It tells us that unless prevented from doing so by certain other token states - e.g. tokens of the belief that eating ice cream at the required rate would be likely to cause him to experience a sharp pain behind his eyes, and of the desire to avoid that experience - the aforementioned beliefs and desires will interact to produce ice-cream-eating behaviour by the subject.

It's true enough that the claims of folk psychology, as construed by the theory theorist, are modest in one respect: by characterizing mental states in purely functional terms, they allow for the multiple realizability of such states and so do not entail commitment to the truth of empirically defeasible assertions about (say) the physical composition of subjects possessing those states. This fact, however, does not strike me as being nearly so impressive as philosophers often appear to think. If I offer a speculative explanation of how an unfamiliar model of car works in terms of things like cylinder heads, fan belts and brake fluid reservoirs, I am allowing for the multiple realizability of these functionally-characterized components of four-stroke petrol engines and hydraulic braking systems. But I am not allowing for the multiple realizability of car engines and braking systems themselves. I might be quite wrong in asserting that the system mediating between inputs from the driver (e.g. presses on the accelerator and brake) and 'behavioural' outputs (e.g. changes in speed) has components playing such-and-such cylinder head roles, fan belt roles, brake fluid reservoir roles, and so on. Maybe the vehicle has some other sort of engine or braking system, performing the same *overall* functional role as a four-stroke petrol engine and a hydraulic braking system, but quite different in terms of its components and workings.

Similarly, the functional organization of the mind may - in principle - be quite different from what folk psychology (again, as construed by the theory theorist)

# *1 Introduction: Why Dispositionalism?*

supposes it to be. In arguing for an eliminativist position, Stich (1983) makes the following point:

> It is a fundamental tenet of folk psychology that *the very same* state which underlies the sincere assertion of 'p' also may lead to a variety of nonverbal behaviours. There is, however, nothing necessary or a priori about the claim that the states underlying assertions also underly nonverbal behaviour. There are other ways to organise a cognitive system. There might, for example, be a cognitive system which, so to speak, keeps two sets of books, or two subsystems of vaguely belief-like states. One of these subsystems would interact with those parts of the system responsible for verbal reporting, while the other interacted with those parts of the system responsible for nonverbal behaviour.
>
> (Stich, 1983, p.231)[5]

This being the case, it's perfectly plausible that a subject should possess a state responsible for the production of certain nonverbal behaviour, and yet lack a state responsible for the production of certain verbal behaviour we feel ought to go along with it. Indeed, Stich (1983, pp.231-233) adduces evidence from the psychological literature that subjects can be induced to go through certain mental processes, as evidenced by their non-verbal behaviour, which they sincerely and resolutely deny that they have gone through. This leads Nisbett and Wilson (1977, cited in Stich, 1983, pp.233-237) to hypothesize that subjects who are asked why they responded in a particular way to a particular stimulus do not report thought processes they actually went through, but rather make a judgement about what sort of process, according to some pre-existing theory, might plausibly have been responsible for their responding as they did. Introspective evidence about the processes actually responsible for their nonverbal behaviour might be quite unavailable to the systems responsible for producing their verbal behaviour.

To sketch one example of the sort of experiment Stich has in mind here: insomniac subjects given a placebo pill and told it would relieve their symptoms typically took *longer* to get to sleep as a result, while insomniac subjects given a placebo and told it would make their symptoms worse typically got to sleep *more quickly* as a result (Storms and Nisbett, 1970, cited in Stich, 1983, p.232). These results bore out the experimenters' hypothesis that subjects in the for-

---

[5]I should note that I see no principled reason why there should not be a system keeping *more* than two sets of books, with different subsystems involved in the production of different types of verbal and nonverbal behaviour. But focusing on the coarse-grained distinction between verbal and nonverbal behaviour suffices to make the point.

mer group would infer that their thoughts must be more troubling to them than usual (since they were still experiencing symptoms of insomnia in spite of having taking the pill), while subjects in the latter group would put their symptoms down to the pill they had taken and so infer that their thoughts must be less troubling to them than usual. When later asked to explain why they had found it easier or more difficult to get to sleep on the night they took the pills, subjects in the former group talked about their more-troubling-than-usual thoughts that night, and subjects in the latter group about their less-troubling-than-usual thoughts. But subjects did not believe that their coming to view their thoughts as more or less troubling than usual had anything to do with having taken the pill they took. The inferential process that had taken place had apparently done so without the knowledge of whatever cognitive system was responsible for producing the subjects' verbal reports. Those reports seem more like *ad hoc* rationalizations of subjects' behaviour than accounts of the processes actually responsible for that behaviour.

What this suggests is that there may be cognitive states which play the role in *some* cognitive processes associated by folk psychology with some particular belief, and yet which fail to play the role in *other* cognitive processes associated by folk psychology with that very same belief. And this is no less true in cases where verbal and nonverbal behaviour accord with one another; quite plausibly there are often two distinct cognitive states at work, playing two distinct causal roles, neither of which can be identified with any of the beliefs posited by folk psychology.

Much the same point could be made with reference to experiments involving 'split brain' patients who have had the lines of communication cut between the left and right hemispheres of their brains through the severing of the corpus callosum. Show these patients an image in only the left side of their visual field - that image being transmitted to the brain's right hemisphere - and they will be able to identify the object shown by drawing it, or picking out a related image with the hand under the control of that hemisphere (the left hand); but since the speech control centre is in the *left* hemisphere, which has no access to the image, they will be unable to identify the object verbally (even denying that they saw anything) (Gazzaniga, 2000, pp.1316-1318). No state of such a subject, surely, can be identified with or seen as realizing a folk psychological 'belief' about what the image shows, since the functional role associated (according to the theory theory) with such a folk psychological belief just is not being played by anything in such cases. That role, after all, includes giving rise to verbal as well as nonverbal behaviour.

# _1_ _Introduction: Why Dispositionalism?_

Split brain patients also engage in _ad hoc_ rationalizations of the nonverbal behaviour carried out by their left hands. One subject had an image of a snowy scene shown to his right hemisphere, and an image of a chicken claw shown to his left; when he then picked out an image of a shovel with his left hand, he explained that this was because shovels are used to clean out chicken coops (Gazzaniga, 2000, pp.1316-18). But the belief that shovels are used to clean out chicken coops can have played no role in producing the item of nonverbal behaviour he was trying to explain, since that behaviour was produced by the right hemisphere of his brain (which had no access to the image of the chicken foot, and was responding to the image of a snowy scene that it _did_ have access to).

Of course, in subjects with normal brains, the two hemispheres can communicate with one another. But that's no reason to think the distinct brain states and processes involved in the production of verbal and nonverbal behaviour in the brains of split-brain patients might not also be present in normal brains, at least some of the time. Our brains are quite plausibly highly modular, with different states of different systems responsible for doing different things - including things that folk psychology, according to the theory theory, takes to be done by one individuable state called, say, a 'belief that P'.

Perhaps it will be objected that the solution is simply to say that having a belief that P is a matter of one's left _and_ right hemispheres, and perhaps the tissue connecting them, collectively being in a state such that the right verbal _and_ nonverbal behaviour is produced in the right circumstances. This would allow us to insist that neurologically intact subjects, at least, really do have beliefs in virtue of which they act the way they do. But it seems to me that this would be to play into the hands of the eliminativist, who can point out that the 'belief' state here is explanatorily superfluous at best (since we could, in principle, explain and predict subjects' behaviour just as well by appealing to the individual verbal- and nonverbal-behaviour-generating states brought under belief's umbrella) and theoretically unsupportable at worst (since it actively obscures important distinctions between distinct states playing distinct roles in producing distinct types of behaviour).

It must be emphasized that the empirical questions here, about the actual functional organization of the brain, are incidental. The point is just that if we take our everyday mental-state language to induce commitment to a particular theory about what cognitive state types there are and what roles they play, we leave ourselves wide open to the empirical falsification of all sorts of everyday assertions about the mind. A better theory really might come along.

A certain line of eliminativist argument (Churchland, 1981) presses the point that this cannot be dismissed as a merely theoretical, remote possibility. Why, after all, should we expect the posits of *this* folk theory to turn out really to exist, when the posits of *other* folk theories generally turn out not to exist? The gods of folk meteorology, held responsible for sending the rain and so on, don't exist. Nor do the forces folk medicine takes to be at work when people fall ill or recover - disruption to or restoration of the flow of *qi* around the body's meridians, say. The posits of folk physics and folk chemistry have fared no better. So isn't it rather unlikely that our distant ancestors got lucky when they posited the existence of beliefs, desires and so on to explain one another's behaviour?

To this it is tempting to reply that whereas we never had direct experience of the gods we once took to be behind the rain and winds (say), we do have direct experience of our own mental states. The existence of beliefs, desires and so on is therefore not open to doubt in the way that the existence of the gods is open to doubt. But however seriously we take introspective evidence, this will not do. From the fact that I can, for instance, make internal utterances such as 'it's sunny' or even 'I believe that it's sunny', and perhaps feel confidence in the truth of those statements, it does not follow that I am either directly conscious of or entitled to infer the existence of an internal state playing just the role reserved by folk psychology, according to the theory theory, for the belief that it's sunny. Perhaps there just is not any single, individuable state responsible *both* for the production of those internal utterances (and their accompanying feelings), *and* for the production of certain items of behaviour folk psychology would - again, according to the theory theory - take to be produced by the belief that it's sunny; raising my hand to shade my eyes, say.

All that being said, however, I am not in the least inclined to doubt that the claims of eliminativism are false. *Of course* subjects have beliefs and desires. And this is so, I will argue, because beliefs and desires are not posits of a folk psychological theory about the functional organization of the mind at all. They are not internal or intrinsic states of subjects at all. So long as a subject has the right behavioural, phenomenal and cognitive dispositions, nothing science tells us about *why* he has those dispositions - in terms of what cognitive or neurological states and processes play what roles in mediating between sensory inputs and behavioural outputs - has any bearing on the truth of ascriptions to that subject of such-and-such beliefs, desires, and so on. Thus dispositionalism can insulate folk psychology - properly understood - from the eliminativist's attack.

# *1* *Introduction: Why Dispositionalism?*

Before I set out how I propose to go about developing and defending the phenomenal dispositionalist position, I should perhaps say something about how I think folk psychology *should* be understood. It should be clear that I think it is a mistake to think of it as a theory about the functional organization of the mind, about the states and processes mediating between sensory inputs and behavioural outputs. However, it would be too strong, I think, to say that it isn't a theory in any sense. As practitioners of folk psychology, we are in the business of theorizing about what sorts of behaviour, inferential processes and conscious experiences typically 'go together' in what sorts of subject. We are adept at predicting subjects' future behaviour from their past behaviour (verbal and nonverbal). We readily infer, on the basis of subjects' behaviour, that those subjects are having (or have had) conscious experiences of certain kinds. We confidently assume that subjects with access to certain facts will have made certain inferences. But this has, in my view, far more to do with our ability to recognize patterns in the way subjects act, think and feel in different circumstances than with any presumed knowledge about the functional organization of the mind. We need no more possess a theory about the functional organization of the mind in order successfully to practice folk psychology than we need possess a theory about the functional organization of four-stroke petrol engines and hydraulic braking systems in order successfully to drive a car.

## 1.3 The case for phenomenal dispositionalism

The foregoing was intended to show why I think there is at least a *prima facie* case for giving the phenomenal dispositionalist case serious consideration. Let me now set out how I propose to make that case.

The starting point for any dispositionalist position in the philosophy of mind, and my starting point in Chapter 2, is *logical behaviourism*. As formulated by Carnap (2002), in response to verificationist concerns about the meaningfulness of mentalistic language, logical behaviourism held that all sentences of psychology (including folk psychology) were logically equivalent to (= reducible to, = intertranslatable with) sentences about observable behaviour - specifically, about the observable behaviour subjects are disposed to exhibit in a range of external circumstances.

Logical behaviourism, I concede, fails because for many mental states, there is no set of behavioural dispositions the possession of which is necessary and sufficient for the possession of that state. A respectable dispositionalism must

admit the relevance of *non*-behavioural dispositions to the question of what mental states a subject has, as well as offering some account of how it is that mental states interact in such a way that, for instance, subjects with a given belief in common come to have different dispositions when they have different desires.

Though usually characterized as a behaviourist, Ryle (2000) did in fact admit the relevance of non-behavioural dispositions to the question of what mental states a subject has. The roots of the position I advocate, phenomenal dispositionalism, are therefore - I suggest - present in Ryle. However, Ryle did not offer a satisfactory account of how mental states interact. Nor did he respect the intuition that explanations of subjects' behaviour that cite their mental states are *causal* explanations.

My discussion of Ryle is followed by a discussion of Stout's (2006) defence of behaviourism. I argue that this defence ultimately fails, in part because of lingering concerns about the relevance of non-behavioural dispositions to questions about mental states, and in part because it rests on a problematic account of mental causation.

In closing Chapter 2, I note that an influential 'functionalist' line of thinking has sought to resolve certain key problems facing the Rylean dispositionalist by identifying mental states with categorical states of subjects playing certain causal roles (specifically, brain states). This allows for mental states to interact in a straightforward, causal sense, and respects the intuition that mental-state explanations are causal explanations. I further note that Stephen Mumford has offered a functionalist account of dispositions in general, according to which all dispositional properties are identical with their categorical bases. To adopt this view of dispositional properties would be to offer phenomenal dispositionalism as a variant of functionalism; if it is to be a distinctive position, therefore, this view of dispositions must be refuted and an alternative account of the interaction of mental states and their role in causal explanations must be offered.

Chapter 3 consists of my attempt to refute Mumford's functionalist account of dispositions. Central to this account is his 'argument from identity of causal role'. Mumford claims that the dispositional properties we prescientifically ascribe to objects occupy the same causal role *conceptually* that is revealed by science to be occupied *in fact* by certain intrinsic, categorical properties of those objects, and that those dispositional properties must therefore be identical with those categorical properties. (This move is crucial to the bringing together of functionalism and physicalism by thinkers such as Lewis (2002); more on this in Chapter 4). I argue that Mumford's argument fails, because

# *1 Introduction: Why Dispositionalism?*

the concept of 'occupancy' is such that there are cases in which two entities may occupy the same thing and yet fail to stand to one another in a relation of identity. I further argue that if we grant that dispositional properties conceptually occupy causal roles, these roles include the roles not only of categorical properties of the object of ascription, but also those of categorical properties of other objects with which they interact - and, indeed, of other features of the environment in which those interactions take place. For instance, I argue that, conceptually, sugar's disposition to dissolve in water would have to be taken to occupy the causal roles not just of certain categorical properties of sugar, but also of certain categorical properties of water; and water's disposition to dissolve sugar would have to be taken to occupy those very same causal roles. Finally, I argue that there is no reason to think of dispositional properties as intrinsic properties of the objects of ascription at all; better to think of them as *relational* properties of those objects.

In Chapter 4, I consider some of the implications for the philosophy of mind of taking such a relational view of dispositions. If the conclusions of the previous chapter are correct, Lewis's (2002) functionalist argument for the identity of mental states with neural states is deeply misguided. To be in such-and-such a mental state is not to be in an intrinsic state playing such-and-such a causal role, but to stand in a certain relation to certain objects and to certain actual or possible events. Just as the water-solubility of a sugarcube has its categorical basis in physical properties of the sugarcube, of water, and of the environment in which sugar and water interact, so a subject's fear of spiders (say) has its categorical basis in physical properties of that subject, of spiders, and of the environment in which they interact. These properties all have distinct causal roles to play, none of which is individually identifiable with the causal role conceptually played (according to the theory theory of folk psychology) by the subject's fear of spiders.

Chapter 4 concludes with a discussion of Lynne Rudder Baker's 'Practical Realism', a position according to which *'S* believes that *p* if and only if there are certain counterfactuals true of *S*, where the content of the counterfactuals may be intentionally characterized' (1995, p.21) and which can, I think, reasonably be characterized as dispositionalist. Two aspects of Baker's work are especially relevant to my project. Firstly, she undermines the idea that for a state of something to be causally relevant, it must be an internal, physical state of the object of ascription. The 'state of financial health' of a Savings & Loan, she points out, is causally relevant to its performance, but is not identical with any internal state of the S & L; and the same goes for subjects' mental states.

Secondly, she makes clear that we do not have to deny, with Ryle, that explanations of subjects' actions in terms of their mental states are causal explanations - even if we deny that mental states are identical with causally efficacious internal states of subjects. I take Baker successfully to have demonstrated that the dispositionalist can oppose functionalism and the Standard View while respecting the intuition that mental-state explanations of subjects' behaviour are causal explanations.

In Chapter 5, I set out Schwitzgebel's version of phenomenal dispositionalism, this being the contemporary starting point for my own. Schwitzgebel argues (2002, p.253) that 'To believe that P... is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P' - this being 'the cluster of dispositions that we are *apt* to associate with the belief that P'. These include not just behavioural dispositions, but also phenomenal and cognitive dispositions - dispositions, respectively, to have certain sorts of conscious experience and to reason in certain ways. There is no determinate set of dispositions someone must have in order to qualify as believing that P; whether or not the belief that P should be ascribed to someone will depend on the context of ascription and on what we know about the subject's other beliefs, desires, character traits etc. In some contexts (e.g. when a subject has atypical desires, or there is something preventing him from acting as he normally would), we will recognize that there are 'excusing conditions' that explain the non-manifestation of a disposition belonging to the stereotype. And sometimes a subject will be 'in between' having and lacking a certain belief, meaning we have to decide on pragmatic grounds whether it makes sense in a given context to ascribe that belief to him, deny that he has it, or simply spell out the respects in which he does and does not conform to the stereotype for possessing that belief.

Phenomenal dispositionalism, at least in the form I advocate, is an externalist position in the sense that it does not take the content of subjects' mental states to be determined solely by the way things are 'in her head', but also by the way things are in her environment. However, Schwitzgebel (2002) notes an apparent tension between externalism and dispositionalism: dispositionalism seems to require that we ascribe the same dispositions, and so the same mental states, to physically identical subjects in different environments (say, Earth and Twin Earth) - because, assuming physicalism, those subjects would have identical responses to being placed in identical circumstances - but externalism holds that these subjects have different mental states. However, Schwitzgebel points out that we can respect externalism by specifying some of the disposi-

# 1 Introduction: Why Dispositionalism?

tions subjects have in external terms - allowing for physically identical subjects in Earth and Twin Earth environments to have different mental states. And it should be borne in mind that phenomenal dispositionalism is not in the business of providing binary yes-or-no answers to the question of whether subjects have or lack certain mental states. On the phenomenal dispositionalist view, we ascribe such states to subjects for pragmatic reasons when it's useful to do so; thus it makes perfect sense to deny that my Twin Earth counterpart has any beliefs about water while he's going about his business on Twin Earth, but to ascribe such beliefs to him the moment he is teleported to Earth and we become concerned to predict and explain his interactions with water.

In Chapter 6, I argue that Schwitzgebel's reliance on dispositional stereotypes containing 'dispositions that we are *apt* to associate with the belief that P' is problematic, since this invites two quite different readings. Do these stereotypes include *all* the dispositions we're apt to associate with a given belief in any circumstances whatever - e.g. when we know a subject to have very unusual desires or habits of thought? In that case, pretty much any disposition may qualify as belonging to the stereotype for any belief. Or do they include only those dispositions we're apt to associate with a given belief 'by default' - on the assumption that subjects' desires etc. are more or less typical (e.g. the disposition to take an umbrella when going outside)? I argue that neither reading can be integrated coherently into Schwitzgebel's overall position, and that in fact the idea that there is any such thing as a dispositional stereotype for the possession of any given belief (or other mental state) is unsustainable. Rather, there is a plausibly infinite number of dispositional clusters associated with each belief: one for having that belief plus *these* desires, *these* character traits, etc., one for having that belief plus *those* desires, *those* character traits, etc., and so on. I therefore argue that the phenomenal dispositionalist should reject the idea that we can account for mental states one-by-one by referring to a 'stereotype' for the possession of each individual belief, desire etc. A more holistic account is needed, according to which a subject's having the overall set of mental states he has is a matter of his having a certain *dispositional profile* - a certain overall set of behavioural, phenomenal and cognitive dispositions.

In Chapter 7, I suggest that in seeking to accommodate the holism of the mental, the phenomenal dispositionalist can draw on Dennett's (2002) version of interpretationism for inspiration. The interpretationist does not attempt to account for subjects' mental states on a one-by-one basis; rather, he acknowledges that interpretation involves ascribing complex sets of complementary beliefs, desires etc. to subjects all at once. I suggest, however, that interpre-

tationism tends to place too much weight on rationality as the arbiter of correct interpretation; in fact we need to make more room for non-rational features of mental life, including the phenomenal qualities of conscious experience. In cases where rival schemes of interpretation are available - where different sets of beliefs and desires fit a subject's behaviour equally well - it may be that a subject's phenomenal dispositions are what settle, in principle, the question of what mental states she has.

By the end of Chapter 7, my version of a Ryle/Schwitzgebel-type phenomenal dispositionalist position will be on the table. To see what that position amounts to in practice, however, I think it will be very helpful to apply it to a specific problem.

Chapter 8, therefore, takes the form of a 'case study'. In it, I consider Alston's (1989) attempt to show that a functionalist account of mental-state terms allows for them to be applied univocally to human beings and to God, while respecting both God's personhood and his 'radical otherness'. I argue that while this account fails to accommodate a number of important theistic convictions concerning the nature of God, a phenomenal dispositionalist account of the meaning of these terms fares rather better. A broader point emerging from this discussion is that phenomenal dispositionalism - unlike functionalism - is in tune with common-sense intuitions when it comes to questions about the mental states that may be attributed to atypical agents such as God, computers, and the Martians (Lewis, 1980), 'super-Spartans' (Putnam, 1963) and 'Blockhead'-type systems (Block, 1981) that populate philosophical thought experiments. On a phenomenal dispositionalist view, (full?) mentality may not be ascribed to systems that have the right behavioural dispositions but the wrong (or no) phenomenal dispositions, while (limited?) mentality may be ascribed to subjects that have phenomenal dispositions but the wrong (or no) behavioural dispositions.

In Chapter 9, I make a last-ditch attempt to illustrate the dangers of sticking with a Standard View approach to mental states in the face of arguments that our folk psychological taxonomy of such states is mistaken. Gendler (2008a) attempts to maintain her footing on the eliminativist's slippery slope by proposing the inclusion of a new state type - alief - in that taxonomy, to fill a causal gap that is supposedly left behind once one recognizes that the concept of 'belief' is too bound up with rationality for us to accept beliefs as being responsible for a whole range of arational behaviour. I argue, firstly, that the concept of mental-state content used by Gendler in defining alief is obscure at best; and, secondly, that once we embark on this project of 'discovering' new mental

state types to account for the full range of human behaviour, there is no saying where it will end. Why not distinguish, for instance, between alief-like states in which the role of *habit* is crucial, and alief-type states in which *instinct* is key? Or between affect-laden and non-affect-laden aliefs? As philosophers, I argue, we should not be in the scientific business of taxonomizing the mental state types that 'really' drive our behaviour at all. Rather, we should focus on the way everyday mentalistic language enables us to make sense of one another - which is, in my view, by helping us to build up pictures of subjects' dispositional profiles.

## 1.4 Scope and approach

Let me conclude this introductory chapter by clarifying a couple of points about the scope of this thesis and the general approach I propose to take.

In the sense in which 'dispositional' is contrasted with 'occurrent', it is uncontroversial that some mental states are dispositional states. To say that a subject believes that apples are green and desires that he should eat an apple a day, for instance, is not to report that something is happening to that subject right now - e.g. that an image of an apple is hovering before his mind's eye with two labels attached to it reading 'this is green' and 'I should eat one of these a day'. Nor are we reporting any such mental occurrences when we say that a subject is confident, vain, wise, quick-witted etc., or that a subject loves his dog but hates next door's cat. We use this sort of language - we ascribe *dispositional mental states* of these types (propositional attitudes, character traits etc.) - when we want to say something about what a subject is like in general. The reporting of mental occurrences, meanwhile, has its own language - *he decided to catch the bus, the right word suddenly popped into her head, he felt a twinge of regret, she realized that it was lunchtime already*.

I offer phenomenal dispositionalism as an account of the sort of dispositional mental states just mentioned. Things like realizations and twinges might count as manifestation events for certain dispositions, but they are not themselves subject to a dispositional analysis. This makes the whole enterprise sound rather trivial; just how controversial can it be to argue that dispositional mental states should be understood in dispositional terms? But I hope that the manner of the dispositional understanding I propose is sufficiently distinctive - when contrasted, especially, with a Standard View understanding of these dispositional mental states as being identifiable with (or realized by, or constituted by) particular categorical brain states - to be interesting.

Wherever possible, I will try to avoid treating mental states as a special sort of mystery. My hope is that if we can get clear on how disposition ascription works in general - what its purpose is and what the relationship is between something's dispositional states or properties and its categorical states or properties - certain problems in the philosophy of mind can be dispelled, and others reframed in a more constructive way.

# 2 Beyond Behaviourism

## 2.1 Introduction

Any discussion of dispositionalism necessarily takes place in the shadow of logical behaviourism. Assessing the claims of phenomenal dispositionalism means understanding how they differ from the claims of behaviourism, and especially how it is that the phenomenal dispositionalist thinks he can answer certain objections to behaviourism that have generally been found compelling.

In this chapter, therefore, I want to begin by sketching (in section 2.2) a position that deserves the name 'logical behaviourism' if anything does - the position expressed by Carnap in his (2002) - and spelling out the objections that have seen this position consigned to the dustbin of history. In section 2.3, I will highlight some crucial differences between Carnap's views and the views of Ryle, as expressed in his (2000). My aim here will be to remove the stigma of behaviourism from Ryle's work, by drawing attention to Ryle's frank acknowledgment of the relevance of non-behavioural dispositions to the question of what mental states may be ascribed to subjects. I will also seek to clarify the sense in which I take Ryle to be opposed to a 'Cartesian' view of the mind: not, that is, in challenging the idea that we can meaningfully talk about unobservable 'inner' experiences and processes, but in challenging the 'category mistake' made by anyone who thinks that mental-state talk is talk about the causal workings of a mechanical or para-mechanical system called a 'mind'. This section should make clear that to be a dispositionalist in the Rylean tradition is not to embrace behaviourism, and so is not to be subject to certain compelling objections levelled against that position.

Before moving on, in future chapters, to consider what a liberal, non-behaviourist version of dispositionalism ought to look like, I will briefly consider (in section 2.4) Stout's (2006) defence of behaviourism itself. I will argue that Stout's central claim - that being in such-and-such a mental state is just a matter of having such-and-such behavioural dispositions - remains untenable. However, I will suggest that his attempt to show how behaviourism must and can accommo-

date the holism of the mental has implications for the development of a more liberal dispositionalist position.

## 2.2 Behaviourism

### 2.2.1 Carnap's logical behaviourism

In his (2002), Rudolf Carnap makes the claim that '*all sentences of psychology describe physical occurrences, namely, the physical behaviour of humans and other animals*' (p.39, Carnap's italics). This claim is open to misinterpretation, and so it is worth pinning down just what Carnap seems to have in mind.

Carnap does not mean that every time we say something about a subject's mind, what we are really doing is reporting that that subject is now behaving thus-and-so, or behaved thus-and-so at such-and-such a time. Rather, he means that every time we say something about a subject's mind, we are asserting 'the existence of a physical structure characterized by the disposition to react in a specific manner to specific physical stimuli' (Carnap, 2002, p.43). To say that a subject, Mr. A, is 'excited', for instance, is to assert 'the existence of some physical structure (micro-structure) of Mr. A's body (especially of his central nervous system) that is characterized by a high pulse and rate of breathing, which, on the application of certain stimuli, may even be made higher, by vehement and factually unsatisfactory answers to questions, by the occurrence of agitated movements on the application of certain stimuli, etc.' (Carnap, 2002, p.43). These testable claims about subjects' behavioural dispositions are, says Carnap, analogous to testable claims about the behavioural dispositions of physical objects; to claim that a wooden support is 'firm', for instance, is to assert the existence of some physical structure of that support 'that is characterized by the fact that, under a slight load, the support undergoes no noticeable distortion, and, under heavier loads, is bent in such and such a manner, but does not break' (2002, p.43).

Carnap's claim needs to be understood in the context of his commitment to *verificationism*. As a verificationist, Carnap believes that claims are meaningful only if there are, in principle, observations that one could make that would tend to confirm or disconfirm the truth of those claims. If no observable feature of the world could ever serve either to give weight to the claim that P or to undermine that claim, then the claim that P is (according to the verificationist) literally meaningless; it doesn't say anything about the world at all.

If claims about a subject's mind were claims about private, unobservable goings-on in a sort of hidden inner realm, therefore, those claims would be meaningless on a verificationist view. In order to maintain the meaningfulness of such claims, Carnap proposes that sentences expressing such claims are inter-translatable with (or have the same meaning as) sentences expressing testable claims about something observable. This 'something' is behaviour. A sentence like 'Bob wants an apple' does not, on this view, express the untestable claim that (say) Bob is holding an image of an apple before his mind's eye while experiencing a sense of longing, but rather the testable claim that Bob will accept an apple if offered one, assent to the proposition that he wants an apple, and so on.[1]

Carnap denies, therefore, that sentences of psychology make assertions about 'a consciousness, a certain power or entity' (2002, p.43) existing alongside subjects' physical bodies. Insofar as they were interpreted as making such assertions *in addition to* assertions about subjects' behavioural dispositions, sentences of psychology would not be testable and so would be meaningless. The existence of such a parallel entity is just not, in principle, something we could confirm empirically.

In short, then: Carnap holds that claims about subjects' mental states are logically equivalent to, or reducible to, claims about their dispositions to exhibit certain overt, observable behaviour in certain circumstances. This is a view that seems to me to be deserving of the name 'logical behaviourism'. As I will argue in section 2.3, however, it would be a mistake to attribute this view to a philosopher often regarded as its most prominent advocate, Gilbert Ryle.[2]

Before I move on to talk about Ryle, though, I think is is worth pausing to note

---

[1]There are some thorny issues to do with the validity of inductive reasoning in the vicinity here. Claims such as 'Fs always A in circumstances C' and 'if this F had (counterfactually) found itself circumstances C at $t_2$, it would have A-ed' should plausibly be regarded as meaningless on a verificationist view, since there is seemingly no observable difference between a world in which Fs always A in circumstances C and a world in which Fs just happen to A in circumstances C when we happen to be looking; or between a world in which F would have A-ed if it had found itself in circumstances C at $t_2$ and a world in which F would not have A-ed if it had found itself in circumstances C at $t_2$ (even if F has been observed to A in circumstances C at $t_1$, $t_3$ etc.). I don't propose to get into these issues; suffice it to note that Carnap thinks observations of particular events *can* serve to confirm or disconfirm general claims of this sort (concerning laws, dispositions etc.); we can test the truth of such general claims by testing the truth of specific claims that are deducible from them (e.g. we can test the truth of the claim that Fs always A in circumstances C by putting some Fs in circumstances C at different times and seeing whether they consistently A or not) (Carnap, 2002, pp.40-41).

[2]Julia Tanney notes that while the question of whether Ryle is a behaviourist is still the subject of some debate, the 'standard interpretation [of] Ryle's view is that statements containing mental terms can be translated, without loss of meaning, into subjunctive conditionals about what the individual will do in various circumstances' (Tanney, 2009a, xxv).

some of the important objections faced by the behaviourist. The success of any alternative version of dispositionalism will, after all, depend on its being able to answer these objections.

## 2.2.2  Objections to behaviourism

As arguments put forward by Chisholm (1957, ch.11) and Geach (1957, p.8) make clear, behaviourism demands an impossibly tight connection between the possession of particular mental states and the possession of particular behavioural dispositions. To take a stock example, consider the belief that it is raining. According to the behaviourist, the claim that a subect has that belief is reducible to the claim that he is disposed (among other things) to use an umbrella while outdoors if possible. But it is easy to imagine cases in which a subject lacks that behavioural disposition, and yet (intuitively) possesses that belief: for instance, he might simply enjoy getting wet. And the obvious response - tweaking the specification of the circumstances in which a subject with the belief that it is raining would use an umbrella, so as to exclude circumstances in which he has the desire to get wet - is not available to the logical behaviourist. The logical behaviourist, after all, is trying to 'analyze away' all references to mental states, and so cannot refer explicitly to a subject's 'desires' in her analysis of that subject's beliefs; nor, though, can she simply translate that reference to a subject's desire to get wet into the language of behavioural dispositions. For consider what this would involve. The behaviourist would want to reduce the claim that a subject desires to get wet to the claim that that subject is disposed (among other things) to venture outside without an umbrella when it's raining; and so we end up with a reduction of the claim that Bob believes it's raining to the claim that Bob is disposed (among other things) to use an umbrella outdoors if possible unless he is disposed (among other things) to venture outside without an umbrella when it rains. Plainly this won't do; rather than telling us anything meaningful about the way people's actions, beliefs and desires relate to one another, the behaviourist appears to be prescribing the reduction of all mental-state ascriptions to vacuous assertions that subjects are disposed to act thus-and-so, unless they're not.

This objection rests on the claim that the possession of a particular set of behavioural dispositions is not *necessary* for the possession of some particular mental states; one can believe that it's raining, for instance, without being disposed to take an umbrella when one ventures outdoors. A second objection rests on the claim that the possession of a particular set of behavioural dispo-

sitions is not *sufficient* for the possession of particular mental states of certain sorts. *Prima facie*, after all, certain mental states are the states they are partly in virtue of their phenomenal character. A subject could not be 'in pain', for instance, unless his conscious experiences had a particular character - even if he was disposed to exhibit 'pain behaviour' in appropriate circumstances. He might be, as Block puts it (1981, p.12), a 'perfect pain-pretender'. (This objection is not going to cut any ice with the committed verificationist, of course, since he will think that assertions about 'conscious experiences' are nonsensical insofar as they are irreducible to assertions about behavioural dispositions; but it strikes me as overwhelmingly plausible that at least some mental states must be characterized at least in part with reference to their phenomenal features. It also strikes me as overwhelmingly *im*plausible that no meaning can be attached to a claim such as 'in addition to being disposed to limp (etc.) in circumstances C, this animal/robot/Martian experiences feelings of pain in its leg in circumstances C'.)

The phenomenal dispositionalist response to these objections is simply to concede that the possession of a particular set of behavioural dispositions may indeed fail to be either necessary or sufficient for the possession of a particular mental state. (More than this, as we shall see in Chapters 5 and 6, the phenomenal dispositionalist concedes that there is not even any determinate set of behavioural *and non-behavioural* dispositions the possession of which is necessary and/or sufficient for the possession of a particular mental state; given the holism of the mental, this sort of one-at-a-time reduction of mental states to sets of dispositions is not possible.)

A further objection, often raised by philosophers in the functionalist tradition (e.g. Armstrong, 1981), cannot be so straightforwardly dodged by the phenomenal dispositionalist. Behaviourism, so the objection goes, is incompatible with the view that mental states such as beliefs and desires are the causes of subjects' actions. If 'believes that P' just meant 'tends to act thus-and-so in such-and-such circumstances,' then *prima facie* it wouldn't tell us anything about the *cause* of a subject's acting thus-and-so in those circumstances. That can't be right, so the objection goes, because conceptual analysis reveals that we take beliefs and desires to be the causes of subjects' behaviour. Furthermore, behaviourism is unable to account for the causal interaction of mental states like beliefs and desires. (The thought here is this: when someone puts up his umbrella, his belief that it is raining and his desire to remain dry are somehow combining to cause that behaviour. So a successful account of mental states will be one that takes those states to interact causally with one

another as well as with sensory stimuli and behavioural responses.)

This objection is a powerful one, and Chapter 3 is devoted to the consideration and ultimate rejection of one possible way of answering it: the adoption of what we might call a 'compatibilist' view, according to which being in such-and-such a mental state is indeed just a matter of having such-and-such dispositions, but those dispositions are themselves the causes of a subject's actions. In Chapter 4, I will defend the view that we don't in fact need to think of mental states as playing a *causal* role in the production of subjects' behaviour in order to account for their *explanatory* role in causal explanations of such behaviour.

# 2.3 Ryle's dispositionalism

## 2.3.1 Ryle vs. Cartesianism

There are, of course, similarities between Carnap's views and Ryle's. Ryle, like Carnap, is hostile to the notion that in addition to their physical bodies, the histories of which are histories of public, observable events occurring in the physical world, people also possess nonphysical minds, the histories of which are histories of private, hidden events occurring in a mental world (Ryle, 2000, 13ff.). However, his objection to such Cartesian views is not, first and foremost, that they invite us to conceive of the mind as something nonphysical, private and hidden; it is, rather, that they invite us to conceive of the mind as a system of causes and effects:

> My destructive purpose is to show that a family of radical category-mistakes is the source of the double-life theory. The representation of a person as a ghost mysteriously ensconced in a machine derives from this argument. Because, as is true, a person's thinking, feeling and purposive doing cannot be described solely in the idioms of physics, chemistry and physiology, therefore they must be described in counterpart idioms. As the human body is a complex organized unit, so the human mind must be another complex organized unit, though one made of a different sort of stuff and with a different sort of structure. Or, again, as the human body, like any other parcel of matter, is a field of causes and effects, so the mind must be another field of causes and effects, though not (Heaven be praised) mechanical causes and effects.
>
> (Ryle, 2000, p.20)

Ryle is equally opposed, therefore, both to Cartesian dualism and to the position Lynne Rudder Baker calls 'Cartesian materialism' (1995, p.6) - the view that the mind is a field of physical causes and effects. Both positions, for Ryle, involve the very same category mistake: the mistake of thinking that terms like 'belief' refer to states of a causal system generating behaviour. As Julia Tanney points out, 'to suppose the explanatory power of mental conduct terms depends on their designating an event or state that is causally related to the performance [of the action to be explained] is to accept another version of the "paramechanical" hypothesis, even though it is now couched in ontic-neutral or physicalist terms' (2009b, section 6.1). For Ryle, mental states are not causes, and explanations of subjects' behaviour which cite their mental states are not causal explanations.

## 2.3.2 Ryle and non-behavioural dispositions

Like Carnap, then, Ryle thinks that psychological claims are claims not about the hidden causes of subjects' responses to stimuli, but about their *dispositions* to have certain responses. However, while he certainly emphasizes the fact that much of our mental-state talk has to do with subjects' dispositions to exhibit overt behaviour, he does not share Carnap's commitment to the view that assertions about other features of subjects' mental lives are meaningless. Indeed, he often makes such assertions himself. Schwitzgebel (2002) draws attention to the following passage as evidence that Ryle countenances the relevance of phenomenal and cognitive dispositions, as well as behavioural dispositions, to the question of what mental states may be ascribed to a subject:

> 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 peoples' assertions to that effect, in objecting to statements to the contrary, in drawing consequences from the original proposition, and so forth. But it is also to be prone to skate warily, to shudder, to dwell in imagination on possible disasters and to warn other skaters. It is a propensity not only to make certain theoretical moves but also to make certain executive and imaginative moves as well as to have certain feelings.

> (Ryle, 2000, p.129)

Elsewhere, Ryle could hardly be more explicit in claiming that to ascribe certain motives[3] to subjects in explaining their actions is to ascribe to them an assortment of behavioural, phenomenal and cognitive dispositions:

> To say, then, that a certain motive is a trait in someone's character is to say that he is inclined to do certain sorts of things, make certain sorts of plans, indulge in certain sorts of daydreams, and also, of course, in certain situations to feel certain sorts of feelings.
>
> (Ryle, 2000, p.90)

These passages are noteworthy for the explicitness with which they appeal to non-behavioural dispositions - 'pronenesses', 'propensities' and 'inclinations' privately to think and feel a certain way - but in many ways they are quite unexceptional; Ryle often makes assertions about 'inner' experiences without any suggestion that they can be reduced to assertions about behaviour. For instance, he happily asserts (2000, p.28) that 'Much of our ordinary thinking is conducted in internal monologue or silent soliloquy, usually accompanied by an internal cinematograph-show of visual imagery' - a strange way for a behaviourist to characterize 'thought', but not something a more liberal dispositionalist needs to apologize for. Ryle does not set out to deny the existence of 'inner', private mental goings-on, but to debunk the idea that explanations of behaviour in terms of beliefs, desires and so on are reports of causal transactions between inner entities.

Thus we find in Ryle no doomed attempt to translate sentences making psychological claims into sentences making claims only about observable behaviour - to reduce the assertion that a subject is in a particular mental state to the assertion that that subject would behave like *this* in circumstances $C_1$, like *this* in circumstances $C_2$, like *this* in circumstances $C_3$, and so on. Instead, a disposition ascription like 'Jones believes the earth is round' must be understood, as Tanney puts it, in terms of what would satisfy it: in terms, that is, of 'an open-ended (infinite) list of inferrings, imaginings, saying [sic] and doings (etc.) on the part of Jones' (2009b, section 9) that we would take to reflect Jones's possession of that belief.

There is, no doubt, a price to pay for abandoning the behaviourist's insistence that our criteria for mental-state ascription must boil down to something publicly observable - i.e. the overt behaviour of the subject of ascription. To cite

---

[3]According to Ryle, 'Roughly, "believe" is of the same family as motive words' (2000, p.129). So too is 'desire' (2000, p.85), along with a whole range of other words used to explain why a subject acts in a certain way - 'vanity, kindliness, avarice, patriotism and laziness,' for instance (2000, p.82.)

phenomenal and cognitive dispositions is to invite a series of hard questions, generally outside the scope of this thesis, about things like phenomenal consciousness and our knowledge of other minds. This is not, in my view, a weakness of phenomenal dispositionalism, since these hard questions are real and interesting, and so we should be more wary of positions in the philosophy of mind that see them analyzed away than of positions that remind us of them. However, rather than make haste towards the development of a more liberal dispositionalist position, I think it's worth pausing at this point to consider a recent defence of behaviourism offered by Rowland Stout (2006). While I will argue that Stout's position fails convincingly to answer certain long-standing objections to behaviourism, and invites certain important objections of its own, I will suggest that there are important lessons for any dispositionalist in Stout's handling of the problem of mental holism.

# 2.4 Stout's defence of behaviourism

## 2.4.1 Teleological dispositions

In his (2006), Stout offers a novel version of behaviourism built around the following account of behavioural dispositions:

> [B]ehavioural dispositions are teleological dispositions. . . [that] must be characterised in normative terms. They are dispositions to do the right thing to achieve the right goal, according to some conception of what constitutes rightness. . . This does not mean that behavioural dispositions are dispositions to respond to a person's internal representations of what is right. . . Instead, we should understand the idea of being disposed to behave in a norm-governed way as being disposed to do what the rules of practical rationality dictate. . . So, mental states, the framework causes of human behaviour, are dispositions to behave in ways that are sensitive to practical rationality.
>
> (Stout, 2006, p.97)

This stands in need of some clarification.

Stout insists that mental states must be conceived of as what he calls *framework causes* of behaviour, as distinct from *input causes* of behaviour. That is to say, they must be thought of as consisting in the existence of mechanisms

through which certain behaviour is produced, and not as events or facts feeding into such mechanisms (2006, 72ff.). So it is a mistake to think of things like beliefs and desires as inputs to a psychological mechanism for producing behaviour. To be in the dispositional state that is believing that P is, rather, to embody a mechanism for producing such-and-such behaviour in such-and-such circumstances.

So Stout takes what we might call a *non-psychological* view of practical rationality, in the following sense: the facts about what behaviour a certain version of practical rationality recommends are not facts about what recommendations for behaviour happen to be generated by certain psychological processes - say, a process in my head that takes as inputs my belief that it will rain and my desire to stay dry, and generates the recommendation that I should take an umbrella if going outside. Versions of practical rationality should, according to Stout, be thought of as abstract systems, external to subjects, the inputs to which are *facts*. For instance, there is at least one version of practical rationality which, taking as inputs the facts that it will rain and that I need to go out, would yield as an output the recommendation that I should take an umbrella (Stout, 2006, pp.99-100). According to this version of practical rationality, the right goal to have when it will rain is the goal of staying dry, and the right way to achieve that goal is to take an umbrella.

If my way of behaving can be recognized as being in accordance with the rules characterizing this version of practical rationality - e.g. 'if it will rain and you need to go out, take an umbrella' - I can thereby be said to have the belief that it will rain. What this means is that I embody a mechanism for producing behaviour that is in accordance with the rules generated by this version of practical rationality. This mechanism, or dispositional state, is a framework cause of my taking an umbrella when I go out. An input cause of my taking an umbrella, meanwhile, is a fact - the fact that according to this version of practical rationality, I should take an umbrella with me if it will rain and I need to go out. (Note that it is this fact itself that is the input cause of my behaviour, according to Stout - not, for instance, a brain state representing this fact, or an event of this fact's coming to be held true by me.)

## 2.4.2 Stout on causal explanation

The foregoing is a rather crude sketch of a rather complex position, but it does bring out roughly the shape of Stout's answer to one of the key objections to behaviourism discussed above: the objection that explanations of behaviour in

terms of behavioural dispositions are not causal explanations. On Stout's view, they *are* causal explanations. If I explain that Bob took his umbrella because he believed that it would rain, I am asserting that an input cause of Bob's behaviour was the fact that according to some version of practical rationality, taking an umbrella with you is the thing to do if it will rain; and that a framework cause of Bob's behaviour was his embodiment of a mechanism for behaving as that version of practical rationality recommends one should behave.

In one important respect, I am very sympathetic to Stout's project. He is right, I think, to resist the functionalist idea that when we talk about a subject's beliefs, desires and so on, we are asserting the existence of discrete states of those subjects playing defined roles in mediating between sensory inputs and behavioural outputs (and so, in my view, leaving ourselves open to the risk that advances in empirical science will prove us wrong and the eliminativist right). It seems much more plausible to me to think that claims about mental states are rather more modest assertions - about the way subjects behave in certain circumstances, rather than about the causal mechanisms underlying that behaviour. For reasons that will become apparent in Chapter 3, I would not agree with Stout that the mechanism through which S's behaviour is produced is a mechanism *embodied by S* if, as seems plausible, this implies that S's dispositional properties are internal or intrinsic properties of S (which, I shall argue, they are not). Still, it seems broadly right to say that assertions about S's mental states are (at least in part) assertions that as a result of some mechanism or other, the functional details of which we may not know even in outline, S can be expected to behave thus-and-so in such-and-such circumstances.

However, I confess that I struggle to make sense of the idea that facts about versions of practical rationality might be the input causes of behaviour. If facts are to be regarded as causes, surely they must be grounded in the actual world in something like the way Mellor (1995) grounds certain facts in the actual world by identifying them with spatiotemporally located states of affairs. But a fact about an abstraction like a version of practical rationality is not thus grounded in the actual world. How might the fact that (say) according to some version of practical rationality, taking an umbrella with you is the thing to do if it will rain, go about inserting itself in a chain of spatiotemporally located causes and effects leading from the beating down of rain on Bob's roof to Bob's taking of an umbrella when he heads outside? I understand Stout's reasons for not wanting to regard Bob's belief in the truth of that fact as an input cause of Bob's behaviour; but why not take the input causes of his behaviour to be such events as *rain's beating down upon Bob's roof*, or *sound waves caused by*

*the rain beating down upon his roof reaching Bob's ears*? Or, if a fact-based ontology of causation is preferred, such spatiotemporally located facts as *that rain beats down upon Bob's roof* or *that the sound waves caused by the rain beating down upon his roof reach Bob's ears*? Nothing about this picture, so far as I can see, precludes us from characterizing Bob's dispositions as Stout wants to characterize them - that is, in normative terms, as dispositions to act according to the rules generated by a certain version of practical rationality. The facts about what those rules are, are plainly relevant to the question of what specific instances of behaviour we can expect to see from Bob in what circumstances; they are relevant, that is, to the question of what dispositional state he is in. But *causes* of his behaviour? Surely not.

### 2.4.3 Stout on the phenomenal character of mental states

As we have seen, a further key objection to behaviourism is that the possession of at least some mental states is a matter of having certain types of conscious experience, as well as of exhibiting certain behaviour. In answering this objection, Stout again pursues what we might call an 'externalizing' strategy. Just as he conceives of the input causes of behaviour as being external to subjects - as being facts about what should be done (according to some version of practical rationality), rather than beliefs about what should be done - so he conceives of phenomenal qualities as being external to subjects:

> This response shifts what we might think of as qualitative nature from being an aspect of inner experience to being an aspect of the world - an aspect that is only perceivable from a certain perspective. It is the world that has phenomenal qualities, not the mind. If there is any difficulty in knowing the qualitative nature of things it is not a difficulty concerning knowledge of mental things but a difficulty concerning knowledge of the world. So having knowledge of the way the world is includes having what Block describes as 'phenomenal' consciousness, since it includes awareness of the qualitative nature of things.
>
> (Stout, 2006, pp.209-210)

Elsewhere, Stout says:

> Being conscious of the dusty redness and ineffable earthy smell of a tomato is to know the dusty redness and ineffable earthy smell of the tomato through their being present to one's visual and olfactory

capacities. Sensory qualities are *out there*, and when their presence to us enables us to know them they enter into our conscious states.

(Stout, 2006, p.19, Stout's italics)

The strategy here is to reduce *phenomenal* consciousness to *epistemic* consciousness, so that my being conscious of a tomato's redness is really just my knowing that it's red. And my knowing that it's red, of course, is (on a behaviourist view) really just my being disposed to act as if it's red - e.g. to describe it as being red, or to pick it out when asked to pick out a red object. There is nothing fundamentally more mysterious going on here than there is in the case of a thermostat being 'conscious' of the temperature in a room (= 'knowing' what the temperature is, = being disposed to act as if that's what the temperature is - by turning down the heating, say).

So Stout takes what is, I think, best termed a 'naive realist' view of secondary qualities.[4] I think we have good reason to reject this view, however. While I perceive a tomato as a dusty red, my dog (so I understand) probably perceives it as a browny grey. But the tomato is not 'really' one colour or the other; what's 'out there' is just an object that reflects light at a certain wavelength, and that therefore produces one sort of visual experience in me and another in my dog (given the differences in our sensory apparatus). No doubt there is a sense in which my dog and I are perceiving the same quality or property of the tomato - its capacity to reflect light at such-and-such a wavelength - but this is surely not a *phenomenal* quality of the tomato named both by 'dusty red' and by 'browny grey'. To return to the thermostat: both it and I have the capacity to detect a certain quality or property of the air in a very hot room, its high temperature. But this quality of the air is not identical with the phenomenal quality of my experience when I sit, sweltering, in that room.

So Stout's attempted reduction of phenomenal consciousness to epistemic consciousness fails, in my view. Hence it remains far from obvious that phenomenal dispositions can be reduced to behavioural dispositions.

---

[4] I am aware that I am skimming over some deep epistemological waters in this brief discussion; deeper than I propose to dive into here. To be clear, though: I do not mean to disparage or reject out of hand the direct realist position *per se*, understood as the thesis that in perception we are directly aware of external objects (rather than of intermediaries such as sense data). Rather, I am challenging the supplementary thesis that these external objects of which we are directly aware possess such phenomenal properties as *redness* independently of perceivers. The conjunction of the two theses is what I mean by 'naive realism'.

### 2.4.4 Stout on holism

A third key objection to logical behaviourism was that it demands an impossibly tight connection between the possession of particular mental states and the possession of particular behavioural dispositions. To believe that P is not *in itself* to have any particular behavioural dispositions; it all depends what desires one has, what other beliefs one has, what physical and mental capacities one has, and so on.

Stout acknowledges that a successful version of behaviourism must be holistic and non-reductive; it cannot seek to reduce claims about particular mental states to claims about particular behavioural dispositions. However, he argues that the behaviourist need not be wedded to the idea that each mental state we can ascribe is associated with just one set of behavioural dispositions:

> [Ryle's] view is that when we apply a mental predicate we are describing the way someone is disposed to behave. But this does not commit him to the view that for each mental predicate there is just one corresponding way of behaving. Describing something as concave is to describe its shape; but it does not follow that there is only one shape corresponding to the predicate 'concave'.
>
> (Stout, 2006, p.110)

Stout argues that the claims of behaviourism should be expressed explicitly in terms of the links between subjects' overall states of mind and their overall sets of behavioural dispositions, rather than in terms of links between particular mental states and particular behavioural dispositions:

> According to a holistic behaviourist, knowledge of someone's overall state of mind is knowledge of how that person behaves. On this conception there is not just one way of behaving per person. Nor is there one way of behaving per psychological attitude. There are as many ways of behaving as there are possible overall states of mind. There is one way of behaving which is believing that it will rain at the same time as wanting to stay dry and intending to go out, and so on. There is another way of behaving that is believing that it will rain at the same time as wanting to get wet and believing that going out in the rain is the best way of getting wet, and so on.
>
> (Stout, 2006, p.113)

Although I reject Stout's behaviourism, I endorse his holistic approach to dispositionalism. The reductive behaviourist approach of trying to say, in dispositional terms, what it is to have one particular belief (say) will not do; nor, as we

shall see in Chapter 6, does Schwitzgebel's attempt (2002) to deal with these issues in terms of conformity to 'dispositional stereotypes' stand up to scrutiny. My claim will be that for a subject to have a particular set of (dispositional) mental states[5] is just for them to have a particular set of behavioural, cognitive and phenomenal dispositions; but to have any particular belief (say) is neither to have any particular dispositions, nor to conform closely to any particular dispositional stereotype.

## 2.4.5 A note on rationality

Although Stout allows that 'versions of practical rationality' might sometimes be rather irrational - they might involve some rather eccentric inferences from facts about what situation someone is in to recommendations about what they ought to do - his position still strikes me as being, in a sense, too rationalistic.

Perhaps the best way to illustrate the problem as I see it is with an example. Suppose, then, that we observe two subjects - one an arachnophobe, the other not - acting in just the same way in just the same circumstances: each sees a venomous false widow spider on the floor in front of him and backs away from it. Each can therefore be described, in Stout's terms, as acting according to the rules of a version of practical rationality according to which the goal to have when there is false widow spider on the floor in front of you is to avoid that spider, and the way to avoid it is to back away from it. Does this suffice for the ascription to both subjects of the (*de re*) belief that false widow spiders are dangerous? We had better say 'no', since the arachnophobe might in other respects not be disposed to act in such a way as to justify the ascription of that belief. For instance: he might be disposed not to assent to the proposition that the spider in front of him is dangerous, but rather to say something like: 'I'm sure it's harmless, but I can't help being frightened of it anyway'. The ascription of the belief that false widow spiders are dangerous had better, then, be conditional on a subject's acting according to the rules of a version of practical rationality that recommends, in the circumstances described, not only that one should avoid false widow spiders by backing away from them, but also that one should assent to the proposition that the spider in front of one is dangerous, etc.

All this is fine so far as the non-arachnophobe is concerned - he *does* act

---

[5]As I noted in the Introduction, the mental states I am concerned with in this thesis are those states that are dispositional in the sense in which 'dispositional' is contrasted with 'occurrent' - states such as beliefs, desires, fears, prejudices, character traits, and so on, as opposed to states such as the state of experiencing a certain sensation.

according to such a version of practical rationality, and we can (very plausibly) express this fact by ascribing to him the belief that false widow spiders are dangerous (together with other beliefs and desires that similarly make rational sense of his behaviour). But what of the arachnophobe? It would be futile, surely, to try to come up with a set of beliefs and desires that make rational sense of his behaviour in the same way. We can make sense of it all right, but only by citing his *irrational* fear of spiders. And this is surely not a matter of attempting to characterize a 'version of practical rationality' which generates rules according to which the arachnophobe acts. Rather - as I shall argue in subsequent chapters - it is a matter of attempting to characterize the arachnophobe's *dispositional profile*; to express something about his overall set of behavioural, phenomenal and cognitive dispositions.

## 2.5 Conclusion

I have suggested that we find in Ryle the roots of a form of dispositionalism that is more liberal than Carnap's logical behaviourism in two important senses. Firstly, it is less restrictive about the classes of dispositions that are relevant to the question of whether a subject does or does not possess a particular mental state; a subject's phenomenal and cognitive dispositions matter, as well as his behavioural dispositions. Secondly, it denies that there is any determinate 'checklist' of particular dispositions the possession of which is necessary and sufficient for the possession of any particular mental state; as we have seen noted by both Tanney (2009b) and Stout (2006), rather than there being just one set of behaviours that a subject must exhibit in order to be said to believe that P, there may be a whole range of behaviours (and - on Tanney's reading of Ryle - a whole range of imaginings, inferrings, etc.) that would satisfy the ascription to a subject of that belief.

This sort of phenomenal dispositionalism therefore escapes two of the most important objections levelled at behaviourism: it does not demand the aforementioned tight connection between the possession of particular mental states and the possession of particular behavioural dispositions, and it does not require that we ascribe states such as 'pain' to systems which have the right sort of behavioural dispositions, but the wrong sort of conscious experiences (or no conscious experiences at all). However, it does not so easily escape the objection that a dispositional account of mental states is at odds with our folk-psychological conception of mental states as things that occupy particular causal roles in mediating between sensory inputs and behavioural outputs,

and that, in doing so, interact causally with one another. And this, in my view, is where we need to go beyond Ryle and show how the phenomenal dispositionalist can accommodate the powerful intuition that to offer an explanation of subjects' behaviour by citing their mental states is to offer a *causal* explanation of their behaviour.

There is more at stake here, however, than simply accommodating that intuition. The dispositionalist must also accommodate the holism of the mental; he must be able to say how the range of behaviours (etc.) that would satisfy the ascription of a particular belief (say) to a subject changes depending on what other mental states that subject has. The functionalist deals with this issue by allowing the causal roles played by particular mental states to include interacting with other mental states in particular ways; but how should the dispositionalist deal with the same issue?

One reply to this objection might be this: dispositional accounts of mental states *are* compatible with our folk-psychological conception of mental states as occupants of particular causal roles, because we conceive of dispositions *themselves* as occupants of particular causal roles. On this view, for a subject to have such-and-such a belief, say, would just be for her to be in some state (or: to possess some property) occupying the right causal role in mediating between sensory inputs and behavioural outputs. And part of that causal role includes interacting causally with other such states, so that the range of behaviours caused by any state considered individually changes depending on what other states are around to interact with.

To offer this answer would be to propose phenomenal dispositionalism as a variant of functionalism - one that takes mental states to be the states they are solely in virtue of the 'forward-looking' causal roles they occupy in the processes giving rise to certain sorts of behaviour, experiences and patterns of reasoning, and not also in virtue of the 'backward-looking' causal roles they occupy in the processes through which they themselves came to be possessed by a subject.[6] In Chapter 3, I will consider and reject an avowedly functionalist account of dispositions of the sort that would lend itself to such a view, inviting us to identify dispositional mental states with categorical brain states. In doing so, I will begin to undermine the idea that the usefulness of disposition

---

[6] On this view, a subject would count as having the belief that snow is white, for instance, if she were in some state that played a particular causal role in producing behaviour such as verbally assenting to the proposition that snow is white, experiences such as mentally picturing a snow-covered field as white, and inferences such as the inference that snow is a different colour from grass. How the subject came to be in that state would not matter; she would count as having that belief whether she acquired it (say) by looking out of the window at a snow-covered field, or as a result of some freak event in her brain.

ascriptions depends on their entailing the ascription to subjects of causally efficacious, intrinsic properties or states.

# 3 Mumford on Dispositional and Categorical Properties

## 3.1 Introduction

Like Schwitzgebel (2002) - and Ryle (2000), for that matter - I am primarily concerned to offer a useful account of the role of disposition ascriptions in mentalistic discourse rather than to puzzle over the question of where curious things called 'dispositions' fit into the ultimate structure of reality. However, in proposing a dispositional analysis of mental-state terms, it would seem remiss wholly to neglect the question of what dispositions *are*. Something must be said about how the disposition ascriptions we make relate to what's 'really', fundamentally going on in the world. How, especially, do objects' *dispositional* properties relate to their intrinsic *categorical* properties[1] - the properties of shape, structure and so on in virtue of which we typically take objects to have the dispositions they do? Addressing this question is essential if we are to be clear on just what the claims of phenomenal dispositionalism amount to.

One appealing suggestion is that an object's dispositional properties are simply identical with their *categorical bases* - the categorical properties which are their causal grounds. On this view, a vase's fragility, say, *just is* the microstructural property of the vase which would cause it to break if subjected to certain stimuli under certain conditions (if knocked off a table onto a tiled floor, and so on). In general terms, the appeal of this view is that it accounts for the apparent value of disposition ascriptions in causal explanation in the most straightforward way possible - the vase's fragility causally explains its breaking simply because it *causes* its breaking - while denying that an object's causal powers are grounded in anything other than the sort of categorical properties science

---

[1]I do not mean to imply that all categorical properties are intrinsic properties: 'being ten per cent larger than the Moon' and 'being a mile from Leeds' look like categorical properties to me, but plainly they are not intrinsic properties. However, it is specifically *intrinsic* categorical properties that I am concerned with in this chapter. I explain my reasons for this in section 3.3.

concerns itself with. An object does not, on this view, require an ethereal second class of properties, dispositional properties, in addition to its observable categorical properties, in order to enter into causal transactions with other objects.

As we saw in Chapter 2, this view also has a particular appeal to someone who wants to analyze psychological language in dispositional terms: by allowing for the identification of dispositional mental properties such as *believing that P* with categorical physical properties such as *being in brain state S*, it allows for things like beliefs and desires causally to interact with one another and causally to give rise to manifestation events such as instances of behaviour. This would both account for the role of dispositional mental state ascriptions in predicting and explaining behaviour, and allow for the holism of the mental by allowing individual mental states to send causal ripples through the whole network of states possessed by a subject.

This is the view of dispositions defended by Stephen Mumford in his (1998), and it's the view I want to consider in this chapter. Spelling out just what I take to be wrong with it will, I hope, bring us closer to an understanding of what it is for something or someone to have such-and-such dispositions. By challenging a functionalist account of dispositions in general, rather than a functionalist account of mental states in particular, I hope to show that the dispositionalist about mental states need not rely on special pleading to make his case, and that the functionalist about mental states - and anyone else who takes the mental states of folk psychology to be the occupants of causal roles - is resting his case on some highly questionable assumptions.

Section 3.2 of this chapter will consist of a broad-strokes sketch of Mumford's account of dispositions. In section 3.3, I'll address Mumford's pivotal 'argument from identity of causal role'. After setting aside a couple of initial objections, I will proceed by provisionally conceding to Mumford three central points: that dispositions are causes of the events that are their manifestations; that categorical property instances are causes of those events; and that dispositions and categorical property instances neither individually underdetermine nor jointly overdetermine those events. I will then draw attention to a certain slipperiness in the concepts both of *occupancy* and of *causal role*, and suggest that Mumford's argument fails to go through even when the aforementioned concessions are made, because the premises are only plausible if the expression 'the occupant of causal role R' is given a different reading in each. In section 3.4, I will suggest that, in light of the foregoing discussion, Mum-

ford's widely-shared[2] view that dispositions are intrinsic properties of objects is not sustainable; dispositional properties are in fact *extrinsic*, relational properties. Section 3.5 will be something of a mopping-up exercise, taking - again - a more broad-strokes view of Mumford's position and drawing attention to what I take to be its fundamental flaws. My conclusion will be that the relationship between dispositional and categorical properties is best understood in terms of supervenience rather than identity.

Note that because my concern in this chapter is just to block the route to a functionalist account of dispositions and so to a functionalist understanding of phenomenal dispositionalism, I do not propose to get into the details of what an alternative account of dispositions should look like. That account would, I think, have to be based on a *conditional* analysis, but as to how simple (Choi, 2008; 2006; Gundersen, 2002) or sophisticated (Lewis, 1997; Mellor, 2000; Fara, 2005; Manley and Wasserman, 2008) that analysis should be - that is a question I am happy to leave open.

## 3.2 Mumford's functionalist account of dispositions

In his (1998), Stephen Mumford argues for the view that every disposition - excepting abstract and ungrounded dispositions[3] - is identical with its categorical base. The kind of identity alleged is token-token, not type-type, allowing for the *variable realization* of dispositions: the fragility of vase V might be identical with, say, V's having molecular structure $m_1$ (the microstructural feature of V which would be causally responsible for the manifestation of fragility in the relevant conditions), while the fragility of glass G might be identical with G's having molecular structure $m_2$ (the microstructural feature of G which would be causally responsible for the manifestation of fragility in the relevant conditions).

Mumford describes his account of dispositions as a *functionalist* one, on the basis that dispositions are to be characterized in terms of their causal or functional roles. It is to be distinguished, however, from the functionalist account of Elizabeth Prior (1985), according to which a disposition D is itself a causally

---

[2]Choi and Fara (2012, section 5) note that the intrinsicness of dispositional properties has been defended or take for granted by philosophers including Harré (1970), Armstrong (1973), Mellor (1974), Mackie (1977), Lewis (1997), Bird (1998), Molnar (2003), Choi (2005b), and Menzies (2009).

[3]Abstract dispositions include such things as the divisibility by two of the number eight; ungrounded dispositions include such things as the fundamental properties of subatomic particles (e.g. an electron's charge). See section 3.4.

impotent, second-order property - the property of having some causally potent, categorical first-order property occupying such-and-such a causal role. On Mumford's view, the fragility of V, say, *just is* V's causally potent, categorical first-order property of having molecular structure $m_1$; on Prior's view, the fragility of V is a second-order property possessed by V in virtue of the fact that V has some causally potent, categorical first-order property occupying the relevant causal role - a property which happens to be 'having molecular structure $m_1$'.[4]

Mumford is opposed to *property dualism* - the view that there are, fundamentally, two different types of property out there in the world, dispositional and categorical. He is also opposed, however, to four possible varieties of *property monism*: reductive and eliminativist varieties of *categorical monism* (according to which there is, fundamentally, just one type of property out there in the world - categorical - and dispositional properties are either reducible to such properties or wholly eliminable), and reductivist and eliminativist varieties of *dispositional monism* (according to which there is, fundamentally, just one type of property out there in the world - dispositional - and categorical properties are either reducible to such properties or wholly eliminable). He argues instead for a position he calls *neutral monism*, according to which there is, fundamentally, just one type of property out there in the world, but it's equally legitimate to characterize those properties as either dispositional or categorical; there's just no need to insist that they're 'really' one or the other.

While Mumford denies that there is an *ontological* distinction to be drawn between dispositional properties and categorical properties, he does accept that there is a *conceptual* distinction between the two. The distinction, he suggests, is that dispositional properties entail conditionals by conceptual necessity, while it's an *a posteriori* matter that certain conditionals are entailed by certain categorical properties. So, for instance, to know that V is fragile is already to know that if it were dropped (in the right conditions) it would break; but to know that V has molecular structure $m_1$ is not yet to know any such thing. That the possession of molecular structure $m_1$ occupies such-and-such a causal role in relation to breaking events is something we can come to know only on empirical grounds. This characterization of the conceptual categorical-

---

[4]The discussion in subsections 3.3.2 and 3.3.3 brings out the main reasons one might have for preferring one of these accounts over the other. Roughly, though, Mumford's account might be preferred by someone who was concerned primarily to respect the intuition that dispositions are the causes of their manifestations, while Prior's account might be preferred by someone who was concerned primarily to respect the intuition that two 'fragile' objects (say) instantiate the very same dispositional property even if each instantiates a different categorical property which is the causal ground of its fragility.

dispositional distinction is one I am happy to accept, for the sake of argument at least.

## 3.3 The argument from identity of causal role

### 3.3.1 The argument

With this background in place, we can consider Mumford's argument for the identity of dispositions and their categorical bases - the 'argument from identity of causal role'. I present the argument below, along with Mumford's statement of the identity conditions for a disposition token *d* and a categorical base token *c*:

**The argument from identity of causal role:**

1. disposition $d_1$ = the occupant of causal role R [by conceptual necessity]

2. categorical base $c_1$ = the occupant of causal role R [an empirical fact]

*Therefore*: disposition $d_1$ = categorical base $c_1$

(Mumford, 1998, p.146)[5]

**Identity conditions for a disposition token d and a categorical base token c:**

$\forall d \forall c$ ((*d* = *c*) $\longleftrightarrow$ $\exists x$ (*d* causes or is caused by *x* & *c* causes or is caused by *x*)

&

$\neg \exists y$ ((*d* causes or is caused by *y* & $\neg$(*c* causes or is caused by *y*))

---

[5]As we might expect from a functionalist theory of dispositions, Mumford's argument here closely parallels functionalist arguments for the identity of mental states and brain states - notably this one, presented by David Lewis:

Mental state *M* = the occupant of causal role *R* (by definition of *M*).
Neural state *N* = the occupant of causal role *R* (by the physiological theory).
[Therefore:] Mental state *M* = neural state *N* (by transitivity of =).
(Lewis, 2002, p.88)

I will return to Lewis's argument in Chapter 4.

∨(¬(*d* causes or is caused by *y*) & *c* causes or is caused by *y*)))

(Mumford, 1998, p.162)

(In the argument from identity of causal role, 'the occupant of causal role R' is intended as a definite description; in the statement of identity conditions for *d* and *c*, *x* and *y* are variables ranging over actual and physically possible events.)

## 3.3.2 Two initial objections

A *prima facie* objection to this argument, which Mumford attributes to Cummins (1974), is that properties - whether categorical or dispositional - simply can't be causes because the relata of the causal relation are events (Mumford, 1998, pp.126-128). Mumford replies that while dispositions cannot, indeed, be 'initiating' or 'efficient' causes, they nonetheless occupy *some* causal role and are thus, in a perfectly acceptable sense, 'causes'.

Note that Mumford certainly does not think dispositions are merely causally *relevant*, in just the way we might say that (e.g.) facts about physical laws or about the truth of conditional statements are causally relevant. The picture he favours is one on which dispositional properties *combine* with events to produce effects; dispositions are causally potent (or efficacious) properties of objects that wait around, as it were, for events to provide the extra bit of 'oomph' required for their manifestations. As we shall see, this insistence on the causal potency, rather than the mere causal relevance, of an object's dispositions, is an important feature of Mumford's position: 'The explanatory value of appeal to dispositions,' he says, 'typically resides in them being *causally efficacious* and being *properties*' (Mumford, 1998, p.14, Mumford's italics).

This talk of causally efficacious properties has the potential to mislead, so let me be clear. Although Mumford talks loosely about properties being causally efficacious, he does not mean that properties themselves, *qua* universals, are causally efficacious; rather, he means that particular instantiations of properties - or *property instances* - are causally efficacious:

> When we say that the weight of the apple caused the pointer on the scales to move, for example, we do not mean that a property of weight in general, construed as a universal, caused the moving of the pointer. Rather it was this particular weight of this particular apple that caused the pointer to move... Unless we accept some notion of properties being instantiated in particulars, then it

seems difficult to sustain the evident link between a thing's properties and the causal transactions into which it enters. It is difficult, in short, to see how, unless we allow that there are particular property instances possessed wholly by objects and substances, a thing's properties can have causal effects. What causes a square peg to fit a square hole? It is not a timeless universal that exists nowhere, rather it is something about this hole and this peg

(Mumford, 1998, pp.160-161)

A full consideration of the issues raised by the Cummins objection, which would involve a lengthy excursion into the metaphysics of causation, would be outside the scope of the present discussion. I propose, therefore, simply to concede that Mumford may be right to think that property instances can, in principle, be causes;[6] it is on the question of whether dispositions may be identified with any such causally efficacious property instances (specifically *intrinsic, categorical* property instances) that I want to challenge him.

As I mentioned in footnote 1 above, it is only *intrinsic* categorical property instances with which I am concerned in this chapter. This is simply because, in the context of Mumford's argument, intrinsic categorical property instances are the only categorical property instances with which dispositions could plausibly be identified. There are two reasons why this is so. Firstly, Mumford insists that dispositions are themselves intrinsic properties - 'instantiated properties which inhere completely within the object of ascription' (1998, p.74). And secondly, Mumford thinks that non-intrinsic, relational properties, 'if they are properties at all, [arguably] need bestow no causal powers on the particular in which they are instantiated' (1998, p.122); dispositional properties, on the other hand - those that are not abstract, anyway - *necessarily* bestow causal powers on the particulars in which they are instantiated, in Mumford's view. The property instances with which Mumford thinks dispositions may be identified must therefore be intrinsic as well as categorical.

A second, equally fundamental objection to the argument from identity of causal

---

[6]Mumford doesn't really pin down just what sort of thing property instances are supposed to be, so it's hard to say how plausible this is. My hunch is that Mumford has got the relation between physical objects and their properties essentially back to front. It is not a curious object called 'this apple's weighing a hundred grams' that provides the causal 'oomph' to move the pointer on the scales; it is the apple itself. And it is because the apple itself does so - not because some other object does so on its behalf - that we may truthfully say of the apple that it weighs a hundred grams (or, if we like, that there exists a property instance called 'this apple's weighing a hundred grams'). But I won't press this objection - partly because I don't think Mumford's account of property instances provides a clear enough target, and partly because to do so would mean tackling metaphysical questions that are outside the scope of the present discussion.

role has been raised (in personal communication) by Helen Steward (2010), who suggests that Mumford is guilty of equivocation in his use of the expression 'the occupant of causal role R'. The thought is this: 'disposition $d_1$' stands for the name of a property, and if Mumford's argument is to go through, this must be the very same property picked out by the definite description 'the occupant of causal role R' in both premises. But in the first premise - grounded as it is in conceptual necessity - 'the occupant of causal role R' must pick out the property 'having a property that is the occupant of causal role R';[7] it couldn't be a matter of conceptual necessity that 'disposition $d_1$' names (and so rigidly designates) the very property that is in fact the occupant of causal role R (the property of having such-and-such a molecular structure, say).

So in the case of a particular dispositional property of a particular object - let's say its fragility - that first premise ought really to look like this:

1. Fragility = having a property that is the occupant of causal role R
[by conceptual necessity]

In the second premise, though, 'categorical property $c_1$' clearly *is* the name of the very property that is the occupant of causal role R, and not of the property 'having a property that is the occupant of causal role R'. So in the case of a particular categorical property of a particular object - let's say its having molecular structure m - that second premise ought really to look like this:

2. Having molecular structure m = the property that is the occupant of causal role R [an empirical fact]

And because 'having a property that is the occupant of causal role R' and 'the property that is the occupant of causal role R' are different properties, these premises simply do not license the move to the concusion:

*Therefore*: Fragility = having molecular structure m

In order to assess this objection, I want to introduce two bits of jargon: by 'A-worlds', I mean worlds with the same natural laws as the actual world (including the actual world itself), and by 'D-worlds' I mean worlds with natural laws which are at least somewhat different from those of the actual world. Now, if dispositional terms name properties that objects tending to exhibit behaviour F in A-worlds have in common with objects tending to exhibit behaviour F in D-worlds, I think Steward's objection is fatal to the argument. If 'fragility', say, names a property possessed both by an A-world vase that would break if dropped in virtue of having molecular structure m, and by a D-world counterpart that would break if dropped in virtue of having molecular structure n,

---

[7]This is very much in line with Prior's view of dispositions as second-order properties (see section 3.2).

then it can't be the case (and so can't be a matter of conceptual necessity) that 'fragility' names the property 'having molecular structure m'. The D-world vase, after all, does not have that property at all - but, *ex hypothesi*, it does have the property named by 'fragility'. More plausibly, 'fragility' names the property Steward suggests it names: the property 'having a property that is the occupant of causal role R', which we might think is a property possessed both by the A-world vase and by its D-world counterpart.

However, Mumford explicitly denies that the property named by 'fragility' would be possessed by the counterpart vase in the scenario just outlined - and with some plausibility, I think. After all, as he points out (Mumford, 1998, p.156), when we ascribe fragility to an actual-world vase, we are saying something about the way it is disposed to respond to certain stimuli in A-worlds; nothing at all is being said about the way it is disposed to respond to similar stimuli in D-worlds. Disposition ascriptions must therefore be regarded as world-relative (or, more generally, as relative to a set of conditions fixed by the context of ascription - conditions which include the relevant natural laws) (Mumford, 1998, pp.155-156).[8]

It might be suggested that this is no objection at all to the view suggested by Steward - the view on which an easily-broken A-world vase and an easily-broken D-world vase have the property named by 'fragility' in common. For one might think that, on this view, disposition ascriptions are world-relative in the perfectly good sense that fragility, say, is being thought of as a property which the former vase has 'relative to' A-worlds but which the latter has 'relative to' (some) D-worlds. That being the case, this view seems perfectly capable of accommodating the fact that in ascribing fragility to an actual-world vase, we are saying something about how it is disposed to respond to certain stimuli in A-worlds and not in D-worlds. Why, then, should we think that the world-relativity of disposition ascriptions rules out the possibility that 'fragility' names a property possessed both by easily-broken A-world vases and by easily-broken D-world vases?

I think what a defender of Mumford would say here is something like this: 'The concept of a dispositional property like fragility is the concept of a *causally*

---

[8] In a similar vein, Galen Strawson argues that an object's 'fundamental dispositions include the disposition to behave in way F in nomic environment 1, the disposition to behave in way G in nomic environment 2, and so on' (2008, p.277) (so that a vase that would break easily in A-worlds but not in (some) D-worlds, and a vase that would break easily in (some) D-worlds but not in A-worlds, have different fundamental dispositions). Mumford is making the further claim that when we, as A-world occupants, use dispositional terms like 'fragility', we are naming the properties responsible for objects behaving a certain way just in our own, A-world, nomic environment.

*efficacious* property of an object, a property that is itself the occupant of a particular causal role. But the property Steward suggests is named by 'fragility' - the property 'having a property which is the occupant of causal role R' - just doesn't fit the bill, conceptually speaking; it's not a property that could itself *be* the occupant of causal role R. And Steward is wrong to think that it couldn't be a matter of conceptual necessity that a term like 'fragility' names the very property that is the occupant of the relevant causal role in the actual world. Why? Because the relevant causal role only *exists* - it only *has* an occupant - in the actual world; and so 'the occupant of causal role R' really does succeed in picking out the very property that is, uniquely, the occupant of that role. Only in A-worlds, after all, does anything mediate causally between A-world droppings and A-world breakings - between dropping-events and breaking-events that occur under A-world conditions. *This* is what it means to say that disposition ascriptions are 'world-relative'. A vase that would break if dropped in some D-world, but would not break if dropped in an A-world, plainly does not have any property that occupies the role of causing it to break if dropped in an A-world; therefore it does not have the property we're ascribing to an A-world object when we say it's fragile.'[9]

What all this boils down to, I think, is a largely intuition-driven disagreement over the correct conceptual analysis of dispositional terms like 'fragility'. Mumford thinks they name causally efficacious properties of objects; to say that an object is fragile is to pick out the property in virtue of which it would break if dropped. If he's right about *that*, I think he can defend the argument from identity of causal role against both the objections just considered.

### 3.3.3 Why think dispositions are causally efficacious properties of objects?

Still, *prima facie* at least, it looks quite possible that such terms don't name causally efficacious properties of objects at all; rather they name second-order, causally impotent properties of objects (like the property of having some first-order property that occupies such-and-such a causal role - a Prior-type functionalist view of the sort Steward's objection seems to presuppose). So what can Mumford say to support his analysis?

Well, according to Mumford:

---

[9]I don't think this is to rule out the possibility that we could use the term 'fragility' to name the property of some D-world object in virtue of which it would break if dropped; it's just that, given the world-relativity of disposition ascriptions, we'd be naming a *different* property from the one we're naming when we ascribe fragility to an A-world object.

> The justification for regarding dispositions as causally efficacious is as follows. If type-identical stimuli are applied to two objects and one reacts differently from the other, then the difference in reactions must be accounted for in terms of some difference between the objects and this is a difference that has a causal effect on the reaction... For there to be explanatory value in... [a] disposition ascription... [it] would have to be taken as an ascription of something that is causally efficacious of such behaviour in such conditions, namely, a *property* of the object.
>
> (Mumford, 1998, pp.14-15)[10]

A related point made by Mumford has to do with our realist intuitions when ascribing dispositions to objects: '[t]he concept of a disposition,' he says, 'is a concept of something that lies behind what occurs' (1998, p.63). So disposition-talk can't just be reduced to talk about the actual and possible events that would constitute the evidence for the possession of dispositions; it has to be construed as talk about 'real instantiations of properties which afford possibilities' (1998, p.63). The fact that we take precautions to prevent the breakage of fragile objects, for instance, shows that we take them to possess some relevant property during times when the conditions for the manifestation of that property are not satisfied.

A third reason for taking dispositions to be causally efficacious properties of the objects to which they are ascribed is suggested by Mumford's consideration of the '*virtus dormitiva*' objection (1998, 136ff.) - the objection that causal explanations given in terms of dispositions are necessarily trivial if dispositions are just defined as properties of an object occupying certain causal roles. ('Why does opium make one sleep?' 'It has a property such that it makes one sleep.') Mumford argues that while such explanations are indeed trivial, this is only so because we begin by supposing that opium really does make one sleep. And to say that opium really does make one sleep, by virtue of one of its properties, is to say something informative: it is to say that sleep follows the ingestion of opium not because of divine intervention, say, or by pure coincidence, but because of *something about the opium* (Mumford, 1998, p.138).

These are points to which I will return in section 3.4. They clearly have some

---

[10]Note that elsewhere, Mumford (1998, pp.119-120) argues in the same way for the causal *relevance* rather than the causal efficacy of a property of the object - seemingly taking 'being causally relevant', 'being a cause' and 'being causally efficacious' to be equivalent. That they *are* equivalent is by no means obvious (see for instance Steward (1997, especially chs.5-6); Jackson and Pettit (1990; 1988)). I will argue in Chapter 4 that there is a good sense in which things can be causally relevant without being causally efficacious. Such things might include facts, laws, and - in my view - dispositional properties.

intuitive force, however, and so for now, I will simply let them stand.

### 3.3.4 On the concept of 'occupancy'

Having set aside a couple of objections to the argument from identity of causal role, and considered some of the reasons Mumford offers for supposing that dispositions are causally efficacious, I propose to concede the following to Mumford for the sake of argument:

1. Dispositions are causes of the events which are their manifestations.

2. Instances of categorical properties are causes of those same events.

3. In the context of the argument from identity of causal role, there couldn't be two wholly distinct property instances occupying the very same causal role. (It could not be the case, that is, either that two wholly distinct property instances are both necessary to produce the relevant effects, or that two wholly distinct property instances are each sufficient and so jointly overdetermine their effects).

Now, Mumford is clearly right to think that *one* way of reconciling these three presumed facts is just to identify dispositions with the instances of categorical properties that are the causes of their manifestation. But, I suggest, it is not the *only* way to do so.

To see why, I want to set aside questions about dispositions, property instances and causal roles for a moment and say something about the concept of *occupancy* in general*.*

Lots of things occupy (are the occupants of, take up, fill) other things. People occupy jobs; intellectual projects occupy minds; objects occupy regions of space. Note, however, that none of these 'arguments from identity of thing occupied' is at all persuasive:

1. Mr Smith = the occupant of job 'account keeper for ACME Electricals'

2. Smith & Jones Accountancy = the occupant of job 'account keeper for ACME Electricals'

*Therefore*: Mr Smith = Smith & Jones Accountancy

This is clearly wrongheaded. If Mr Smith keeps the accounts for ACME Electricals in his capacity as a partner at Smith and Jones Accountancy, then both Mr Smith and Smith and Jones Accountancy can correctly be described as the

holder, occupant or 'doer' of that job. Now, clearly Mr Smith and Smith & Jones Accountancy aren't two distinct things doing the same work twice over, or else sharing the task between them. Neither, though, is Mr Smith just identical with Smith & Jones Accountancy; rather, he in some way belongs to, or acts on behalf of Smith & Jones Accountancy.

> 1. The designing of the conservatory = the thing that occupied Mr Brown's mind on January 5
>
> 2. The designing of the house = the thing that occupied Mr Brown's mind on January 5
>
> *Therefore*: The designing of the conservatory = the designing of the house

This is equally misguided. The designing of the conservatory is just *part* of the designing of the house, and so either one can correctly be said to have occupied Mr Brown's mind on January 5.

> 1. The base of the filing cabinet = the thing that occupies that area of the floor
>
> 2. The filing cabinet = the thing that occupies that area of the floor
>
> *Therefore*: The base of the filing cabinet = the filing cabinet

Once again: if part of the filing cabinet occupies some region of space, we can equally well say that the filing cabinet occupies that region of space.[11]

Obviously there are important differences between regions of space, jobs, minds and causal roles; so no doubt there are important disanalogies between these arguments. But it seems clear enough that the concept of occupancy is, in general, such that if x and y can both be legitimately described as being 'the occupant of Z', and if there aren't two wholly distinct things which are

---

[11]Note that these faulty arguments do not involve straightforward cases of failure to refer to a unique object, as we see in the following:

> 1. John Lennon = The Beatles' lead singer
> 2. Paul McCartney = The Beatles' lead singer
> *Therefore*: John Lennon = Paul McCartney

In *that* case, it's clear enough what something would have to do in order to qualify as being 'The Beatles' lead singer' (that is, a lead singer of The Beatles); the definite description itself is not ambiguous. It's just that at least two things meet the criteria (satisfy the description), and so the description fails to refer to a unique object.

By contrast, when we consider what something would have to do in order to qualify as being 'the occupant of job "account keeper for ACME Electricals"', 'the thing that occupied Mr Brown's mind on January 5', or 'the thing that occupies that area of the floor', it's apparent that the definite descriptions themselves are distinctly slippery; they can be given different readings under which they succeed in referring to different unique objects which, while closely related, are not simply identical.

the occupant of Z, then it may nonetheless be that x is not identical with y. Rather, it may be that some other relation holds between x and y. Perhaps, most obviously, it may be that one of them is part of, or partly constitutive of the other.[12]

Perhaps I have been labouring an obvious and uncontroversial point. What the foregoing discussion makes clear, though, is that we have good reason to be on the lookout for a certain kind of equivocation when we see an expression like 'the occupant of...' popping up in an argument of the sort Mumford presents. It could be that we are being invited to identify two things that are not identical at all, but which stand in some quite different relation.

### 3.3.5 Implications for the relation between dispositional and categorical properties

Is this plausible in the present context, however? Obviously a person might be an employee of a company, a filing cabinet base might be part of a filing cabinet, and an instance of conservatory-designing might be partly constitutive of an instance of house-designing; but could a categorical property instance and a disposition stand in some similar sort of relation?

Could a categorical property instance just be 'part of' a dispositional property instance, say? Well, maybe not - I'm not sure property instances are the sort of things that have parts (although it seems fair enough to say that 'part of' what it is, say, for my kitchen to instantiate the property of containing a table and chairs is for it to instantiate the property of containing a table).

Let's be clear, though. We have been given reasons to think, and have provisionally conceded (in section 3.3.4), two things: that dispositions are the causes of the events that are their manifestations, and that categorical property instances are the causes of those events. What the foregoing discussion shows, however, is that it just doesn't follow from this that dispositions *just are* property instances at all. Still less does it follow that they are identical with particular categorical property instances.

---

[12]This may not exhaust the possibilities, though; it's not entirely obvious to me, for instance, that persons (like Mr Smith) are actually *constituents* of the companies they work for, the charities they represent, and so on. One might prefer to say that it's not the whole person Mr Smith, but just the skills, knowledge and labour he contributes to Smith and Jones Accountancy, that serve to make the firm what it is; and that this is why, if Mr Smith were to be replaced by another individual contributing the same skills, knowledge and labour, the firm would remain essentially the same.

One alternative possibility is that categorical property instances are somehow constitutive of dispositions - that what it is for an object O to have disposition d is for such-and-such categorical properties to be instantiated. Or perhaps the disposition in question could be identified with some *set* of categorical property instances.[13] In either case, it would seem legitimate to say both that a categorical property instance c was the occupant of some causal role R, and that d was the occupant of that role. Mumford would be guilty of equivocation, however, since the sense in which c is the occupant of R is different from the sense in which d is the occupant of R - just as the sense in which Mr Smith has the job of keeping ACME Electricals' accounts is different from the sense in which Smith & Jones Accountancy has the job of keeping ACME Electricals' accounts.

So much for 'occupancy'. Time to say something more about 'causal roles'.

### 3.3.6 The causal roles of dispositional and categorical properties

If some disposition d did all and only the causal work done by some categorical property c - if, that is to say, Mumford's identity conditions for c and d were met - then the discussion in sections 3.3.4 - 3.3.5 would get us precisely nowhere. Such a case would be one where some categorical property c, all by itself, made true the counterfactuals entailed by some disposition ascription; where c was able to provide all the causal 'oomph' required (in addition to the 'oomph' provided by some stimulus event) to bring about the manifestation of the relevant disposition. In a case like *that*, Mumford would be home and dry; there

---

[13]Note that in such cases, it need not (in principle) be the object to which the disposition is ascribed which instantiates all the relevant categorical properties. To deny that a disposition d of an object O must be identical with some intrinsic categorical property of O - and Mumford does insist that it's *intrinsic* properties of O we're dealing with when we talk about dispositions (see section 3.4) - is to open up the possibility that d has to do, somehow, with properties of objects other than O. So these possible accounts of the relation between dispositional and categorical property instances are more radically different from Mumford's identity-based account than they might at first appear. (More will be said about the way the dispositions of one object are related to the categorical properties of other objects in subsection 3.3.6 and in section 3.4.)

 Considerations of parsimony might be thought to provide a *prima facie* reason to prefer Mumford's account to the sort of accounts just suggested (if, for instance, one found it ontologically extravagant to bring an additional entity called a 'set' of property instances into the picture). Equally, one might be inclined to prefer Mumford's account on the grounds of simplicity - identity being the simplest possible relation that could hold between one thing and another. However, in subsection 3.3.6 I will argue that there are compelling reasons to deny that a disposition of an object O could simply be identical with some categorical property of O.

just wouldn't be any more to being d than there was to being c. We could, if we liked, talk about c being a 'part' of d, but only in the sense that any object considered as a whole is part of itself; we'd still be entitled just to identify c with d.

The trouble is, there just aren't any such cases. To see why not, consider how everyday causal explanations typically work. Here are three alternative answers to the question 'Why does sugar dissolve in water?', offered by members of the public on a 'Yahoo! Anwers' web page (Anonymous, no date):

**1. Sugar can be dissolved in water because it is soluble in water.**

This is surely a paradigm of causal explanation in terms of an object's dispositions. In Mumford's view, the value of such an explanation is that it ascribes a causally efficacious property, water-solubility, to the sugar. (That very same property can also be characterized in categorical terms as a feature of sugar's molecular composition, though.)

**2. The molecules of sugar bond with the water molecules. This is because water is a solvent, which are substances that can bond to another kind of subtances [sic] easily.**

This answer tells us what sugar dissolving in water *is* - the bonding of sugar molecules with water molecules - then moves on to offer a causal explanation in terms of a dispositional property of the water - its propensity (as a solvent) to bond with other substances. No properties are explicitly ascribed to the sugar.

**3. Table sugar dissolves in water because when a sucrose molecule breaks from the sugar crystal, it is immediately surrounded by water molecules. The sucrose has hydroxyl groups that have a slight negative charge. The positive charge of the oxygen found in the water molecule binds with the sugar. As the hydration shell forms around the sucrose molecule, the molecule is shielded from other sugar molecules so the sugar crystal does not reform.**

Here we're told quite an involved story about the physical process through which water and sugar molecules bond. The causally relevant property of the sugar is said to be its possession of molecules having hydroxyl groups with a negative charge (let's call this property $m_3$), and the causally relevant property

of the water is said to be its possession of molecules having oxygen with a positive charge (which we'll call $m_4$).

So we have three alternative explanations: one ascribing a dispositional property to sugar, one ascribing a dispositional property to water, and one ascribing a categorical property to sugar and a further categorical property to water. On Mumford's view, the causal role *conceptually* occupied by sugar's water-solubility - the role that, prescientifically, we think and speak of as being occupied by sugar's water-solubility - should be the same as the causal role science reveals *in fact* to be occupied by $m_3$. (In the same way, the causal role occupied by water's sugar-solvency[14] should be the same as the role occupied by $m_4$.) So - is it?

No, it's not. For if sugar's water-solubility conceptually occupies a causal role at all, as we're supposing, it occupies a role which is something like: 'together with the event of sugar being placed in water (in the relevant conditions), causes sugar to dissolve in water'. (It provides all the causal 'oomph' required, in addition to that provided by the relevant stimulus event itself, to bring about its manifestation.) It does *not* occupy a role something like: 'together with the event of sugar being placed in water (in the relevant conditions) and water's sugar-solvency, causes sugar to dissolve in water', or 'together with the event of sugar being placed in water (in the relevant conditions) and $m_4$, causes sugar to dissolve in water'. Conceptually, *no* causal factor is required *in addition to* sugar's water-solubility to bring about a dissolving-event when sugar is placed in water (in the relevant conditions).

Similarly, *no* causal factor *in addition to* water's sugar-solvency is required to bring about the same sort of event in the same conditions. Conceptually, in fact, water's sugar-solvency would surely have to be taken to occupy the *very same* causal role occupied by sugar's water-solubility - the role 'together with the event of sugar being placed in water (in the relevant conditions), causes sugar to dissolve in water'. Conceptually, for sugar to be disposed to dissolve in water *just is* for water to be disposed to dissolve sugar; for sugar to be water-soluble *just is* for water to be sugar-solvent. But clearly $m_3$ and $m_4$ occupy distinct causal roles here, each of which must be occupied if a dissolving-event is to occur.[15]

---

[14]This term has a slightly awkward ring to it, but the neatness of the analogy with 'water-solubility' serves to remind us that the properties named by these terms, *being disposed to dissolve sugar* and *being disposed to dissolve in water*, mirror one another in a particular way when ascribed to water and to sugar respectively.

[15]Heather Logue (2010) has suggested (in personal communication) that distinguishing between sugar's water-solubility and water's sugar-solvency is a way of keeping track of the distinct causal contributions made to dissolving events by sugar and by water - or, in Mum-

What a conceptual analysis of paradigmatic causal explanations seems to reveal, then - at least if we take Mumford's view that dispositions are causes of their manifestations - is that sugar's water-solubility occupies (at least) two distinct causal roles: the one occupied by $m_3$, and the one occupied by $m_4$. And water's sugar-solvency occupies (at least) those same two roles.

So Mumford's claim that in causal explanation, dispositions 'stand in' for categorical properties - so that we might start out, prescientifically, by saying dissolving events are caused by sugar's water-solubility, then work out that they're caused by $m_3$, and so ultimately be in a position to say that the sugar's water-solubility *just is* $m_3$ - just isn't borne out by a consideration of paradigmatic causal explanations. Water-solubility just can't straightforwardly 'stand in' for $m_3$ in causal explanations of dissolving events.

The only way to make things appear otherwise is to perform a sleight of hand when it comes to specifying what are to count as causes and what are to count as background conditions. One can indeed say either of these two things:

> (1) in the relevant conditions, an event of sugar being placed in water, plus the sugar's water-solubility, cause an event of the sugar dissolving in water

> (2) in the relevant conditions, an event of sugar being placed in water, plus sugar's possessing $m_3$, cause an event of the sugar dissolving in water

- and it certainly looks as if, in (2), a categorical property of the sugar has just been straightforwardly slotted into the place occupied in an otherwise identical explanation, (1), by a dispositional property of sugar. But where does water's sugar-solvency - or the categorical property Mumford would identify with it, $m_4$ - fit into this picture? Mumford certainly wouldn't want to deny that that property is causally efficacious, or that it plays a role in the production of the specified dissolving event.[16] But it's not mentioned explicitly in either of these putative

---

ford's terms, of distinguishing between the causal role occupied by some property of the sugar and the causal role occupied by some property of the water. However, I simply don't see how one could specify distinct causal roles occupied by sugar's water-solubility and by water's sugar-solvency in a way that plausibly reflected our ordinary usage of those dispositional concepts. And I think any attempt to do so would necessarily involve a substantial deviation from Mumford's analysis of such concepts, since it appears inescapable on his view (1998, p.89) that either of these dispositions on its own would suffice, in the right conditions and in conjunction with the right sort of stimulus event, for the production of a dissolving event. There is just no leftover causal role for the other disposition to fill.

Of course, Mumford wouldn't want to deny that properties of sugar and of water have distinct causal roles to play in the production of dissolving events; my point is just that it only appears to be possible to specify what those distinct roles are if one characterizes those properties in categorical rather than dispositional terms.

[16]Perhaps a defender of Mumford could argue that the event of the sugar's dissolving in water

causal explanations of that event. Where's it hiding?

Well, in (2), I think water's possession of $m_4$ must be counted among the 'relevant conditions'. The causal picture being painted must look something like this:

(2')

*Relevant conditions:*

Obtaining background (environmental) conditions & water's possessing $m_4$

*Stimulus event:*

Sugar placed in water

*Causally efficicacious property instance(s):*

The sugar's possessing $m_3$

*Effect:*

Event of sugar dissolving in water

But according to the prescientific explanation offered in (1), surely, the picture looks like this:

(1')

*Relevant conditions:*

Obtaining background (environmental) conditions

*Stimulus event:*

Sugar placed in water

*Causally efficicacious property instance(s):*

The sugar's water-solubility

*Effect:*

Event of sugar dissolving in water

In (1), then, sugar's solubility is effectively 'standing in' not just for $m_3$, but for $m_4$ too.

Suppose a defender of Mumford wanted to insist that water's possession of $m_4$ must be counted among the relevant conditions in (1) as well as (2), and so the picture in fact looks like this:

---

is not identical with the event of the water's dissolving the sugar, and that neither dispositional nor categorical properties of the water are strictly causally relevant to the production of the former event. I don't believe that would be Mumford's own view, and I won't attempt to construct such an argument on behalf of an imagined opponent who would be raising metaphysical questions I could not pretend to find meaningful.

(1")

*Relevant conditions:*

Obtaining background (environmental) conditions & water's possessing $m_4$

*Stimulus event:*

Sugar placed in water

*Causally efficicacious property instance(s):*

The sugar's water-solubility

*Effect:*

Event of sugar dissolving in water

This would be to insist that for a dissolving event to occur, it is necessary not only that sugar possesses water-solubility, that the relevant background conditions obtain, and that there occurs an event of sugar being placed in water, but also that water possesses $m_4$. But that just seems wrong: surely it's partly in virtue of water's possessing $m_4$ that sugar possesses water-solubility, and so absurd to talk as if, once sugar's water-solubility has been established, there's a lingering question as to whether water is going to hold up its side of the bargain in bringing about some dissolving event.

Note that on Mumford's view, to insist that $m_4$ must be counted among the 'relevant conditions' in (1) is equally to insist that water's sugar-solvency must be counted among the relevant conditions. On Mumford's view, therefore, (1") is equivalent to (1'''):

(1''')

*Relevant conditions:*

Obtaining background (environmental) conditions & water's sugar-solvency

*Stimulus event:*

Sugar placed in water

*Causally efficacious property instance(s):*

The sugar's water-solubility

*Effect:*

Event of sugar dissolving in water

Again, this is intuitively at odds with the way we use dispositional concepts. If sugar is water-soluble, then necessarily water is sugar-solvent; so there is just no sense in insisting that if sugar is water-soluble and is placed in water, the occurrence of a dissolving event is conditional upon its *also* being the case that water is sugar-solvent. One might as well insist that in order for Bob to obtain a Green Card, it is necessary not only that he should marry a US citizen, but also that a US citizen should marry him.

But there's another problem here: once we slot in that second dispositional property - water's sugar-solvency - we can't help but start running in circles. Both sugar-solvency and water-solubility, on Mumford's account, are ascribed to their possessors relative to some set of relevant conditions; and (1''') expresses the view that sugar is water-soluble relative to a set of conditions which includes water's being sugar-solvent. But if that's right, it must surely also be the case that water is sugar-solvent relative to a set of conditions which includes sugar's being water-soluble. So we end up with a picture on which sugar is water-soluble relative to conditions in which water is sugar-solvent relative to conditions in which sugar is water-soluble relative to... and so on.

And what are we to say about the following disposition ascription?

> Sugar and water are disposed to form a solution when mixed together.

Such uses of disposition ascriptions are by no means unusual (we can and do say things like 'piano and cello tend to sound good together', or 'Josh and Amy are apt to make an awful mess if I leave them in the kitchen together'[17]). They look distinctly awkward for Mumford, though. Conceptually, sugar and water here seem to be sharing a single disposition between them. That would mean, I think, that Mumford would have to say they jointly instantiate some categorical property - a sort of 'umbrella' property, presumably, that includes the causally relevant categorical properties of both sugar and water. And so we seem to be back with the idea that some dispositions, at least, are somehow made up of more than one distinct property instance (see subsections 3.3.4 - 3.3.5).

Alternatively, I suppose, Mumford could just insist that such locutions need to be analyzed into ascriptions of two distinct dispositions: so 'Sugar and water are disposed to form a solution when mixed together' becomes something like 'sugar is disposed to dissolve in water with which it is mixed and water is disposed to dissolve sugar with which it is mixed'. But those two ostensibly different disposition ascriptions just don't seem to be telling us two different

---

[17]I'm assuming here that the pairing of Josh with Amy is what gives rise to their tendency to make a mess; it's not just that Amy has that tendency and so does Josh.

things. They seem to express the very same fact about the way the world is, in certain respects, with regard to sugar and water - just as surely as 'Bob is married to Bella' and 'Bella is married to Bob' tell us the very same thing about how the world is in certain respects with regard to Bob and Bella.

Am I, then, suggesting that dispositions are not *intrinsic* properties of objects at all, as is usually supposed? Yes, I am. In fact dispositions seem to me to be best characterized as extrinsic, *relational* properties[18]: to have such-and-such a disposition is precisely to stand in such-and-such a relation to certain actual and possible objects and events.

In section 3.4, I will argue that in spite of certain metaphysical misgivings, this suggestion is not as controversial as it might appear.

## 3.4 Dispositional properties as relational properties

Suppose that an objector, bridling at the suggestion that dispositional properties are not intrinsic properties of the objects of ascription, were to argue as follows:

'Intrinsic properties are properties an object has wholly in virtue of the way it is in itself; extrinsic properties are properties an object has partly in virtue of the way things are outside of itself. An object's intrinsic properties can't be changed or got rid of by adjusting the rest of the world, and are possessed by all duplicates of that object in different possible worlds; not so for its extrinsic properties.

'With that in mind, consider a duplicate of an actual-world sugarcube sitting alone in an otherwise empty possible world. Unlike its actual-world counterpart, it doesn't have the relational properties of being the biggest sugarcube in the bowl, of being the sugarcube nearest the teapot, or of being the the thing that will sweeten my next cup of tea; there are just no other sugarcubes, or teapots, or future cups of tea around in its environment for it to stand in such relations to. But it does have the same intrinsic properties as its actual-world counterpart: most obviously, it's the same shape, has the same structure and is made of the same stuff.

---

[18]Plausibly there are some properties that are relational in a sense, but that are nonetheless intrinsic: 'being wider at the top than in the middle', for instance. Hence the need to make explicit the claim that dispositional properties are extrinsic as well as being relational in character.

'And of course it also has the same dispositional properties as its actual-world counterpart. It's still true even of this lonely sugarcube, for instance, that if it were to be placed in water in actual-world conditions, it would dissolve. The fact that actual-world conditions don't obtain in its environment, and the fact that there isn't any water around in its environment for it to stand in any relation to whatever, just don't have any bearing on the truth of that conditional (which, of course, entails its possession of the dispositional property 'being water-soluble').

'Furthermore, all duplicates of that sugarcube have just the same dispositional property even in watery D-worlds with different natural laws from the actual world. Even in worlds where those sugarcubes wouldn't dissolve if placed in water, they have the property named by 'being water-soluble' because, again, it's still true of them that if they were to be placed in water in actual-world conditions, they would dissolve.

'So there you have it: the dispositional property 'being water-soluble' *must* be an intrinsic property, because it's possessed by duplicates of actual-world sugarcubes regardless of the way things are outside of themselves in the rest of their environment.'

In light of the discussion in section 3.3.6, however, I would have to insist that this lonely sugarcube and its D-world duplicates do not possess the property 'being water-soluble' wholly in virtue of the way they are in themselves, but rather partly in virtue of the way water is (intrinsically, in any world) and partly in virtue of the way the actual world is in general. There are, therefore, properties which objects possess partly in virtue of the way things are in possible worlds other than those in which those objects are situated; and these properties are best characterized as extrinsic. While my imagined objector is broadly correct in his statement of the instrinsic/extrinsic distinction, he is therefore mistaken to think that *none* of the extrinsic properties possessed by an object in one possible world are shared by its duplicates in other possible worlds; its dispositional properties are important exceptions.

However strong one's intuitions that dispositions just *must* be causally potent, then - and however successful Mumford's arguments to that effect - the claim that dispositions are 'instantiated properties which inhere completely within the object of ascription' (Mumford, 1998, p.74) just does not appear to be supported by an analysis of dispositional concepts as they feature in causal explanation. Even 'addicts of the superstition that all true indicative sentences either describe existents or report occurrences' (Ryle, 2000, p.119) are not, it seems, obliged to conclude that the existents described in this case inhere completely

within the object of ascription (because it could be that disposition-ascribing true indicative sentences describe existents which do *not* inhere within that object). To reach that conclusion, they must also make 'the preposterous assumption that every true or false statement either asserts or denies that a *mentioned* object or set of objects possesses a specified attribute' (Ryle, 2000, p.115; my italics).

If all this seems rather controversial, it shouldn't. We know perfectly well that a scientist who wanted to know what 'makes it true' that sugar is disposed to dissolve if placed in water - who wanted, that is, to discover the causal ground of that disposition - would not restrict his investigations to the sugar. He would also be looking for causally relevant features of the water, and indeed of the environment in which sugar and water interact (such as temperature, gravity or atmospheric pressure). If the causal ground of sugar's water-solubility were to be identified with a categorical property of anything, in fact, it seems to me more plausible that it should be with a categorical property of the world as a whole rather than with a categorical property of sugar itself.[19] (Though note that in possible worlds where sugar wouldn't dissolve if placed in water, there would be no possibility of identifying its water-solubility with any categorical property of that world as a whole. The causal ground of that disposition, after all, would be partly located not in that world, but in the actual world.)

So far, I have been describing cases in which it is reasonable to suppose that an intrinsic, categorical property of the object of ascription plays *some* causal role in bringing about a disposition manifestation (even if it does not play the very causal role conceptually played by the disposition itself). But I think the connection between an object's intrinsic categorical properties and its dispositional properties can be loosened still further; for it seems clear to me that we can easily imagine cases in which it would be perfectly legitimate to ascribe a disposition to an object even if none of its intrinsic properties played any causal role in bringing about a manifestation of that disposition.

Suppose, for instance, that God got bored one day and decided that she was going to turn any sugarcube placed in water into a diamond. All the sugarcubes in the world, surely, would thereby gain the disposition to turn into a diamond if placed in water.[20] Yet none of their intrinsic properties would have changed at

---

[19]I am, therefore, not arguing (like Simon Blackburn (1991)) merely that dispositional 'role states' of an object O may not be identified with categorical 'realizing states' of O; I am arguing that there *are* no states of O (at least if by 'state of O' we mean something about the way O is intrinsically) that stand in the relation of 'role state' to 'realizing state'. It is the way the world is quite generally, and not just the way sugar is (say), that 'realizes' sugar's water-solubility.

[20]Would they also *lose* the disposition to dissolve in water? One could argue that they

all; they would simply stand in a somewhat different relation to things like God, diamonds, water, and events of being placed in water.

This is a silly example, of course, but there's surely no doubt that most of the people who have successfully wielded dispositional concepts over the millennia have believed in the reality of things like divine fiats, blessings, curses, spells, fate, karma and so on. They have believed, that is, that sometimes objects do the things they do not even partly in virtue of the way they are intrinsically, but wholly in virtue of the causal influence of supernatural forces. (Depending on his understanding of or commitment to the claim that everything is caused by God, a theist might even take this to be the rule rather than the exception.) In light of that simple fact, however, and of the discussion in sections 3.3.4 - 3.3.5, Mumford's reasons for thinking that the concept of a disposition is the concept of a causally efficacious property of the object of ascription (see section 3.3.3) just don't support the idea that these properties must be *intrinsic* properties. Taking these reasons one by one:

> 1. 'For there to be explanatory value in... [a] disposition ascription... [it] would have to be taken as an ascription of something that is causally efficacious of such behaviour in such conditions, namely, a *property* of the object.' (Mumford, 1998, pp.14-15)

What the discussion in sections 3.3.4. - 3.3.6 shows is that even if Mumford is right that disposition ascriptions tell us about the causally efficacious properties responsible for an object's behaviour, it does not follow (and is not plausible) that they tell us only about intrinsic properties of *the object of ascription*. Intrin-

---

wouldn't, strictly speaking - it would still be as true as it ever was that they are disposed to dissolve in water *in the relevant conditions*. It's just that those conditions don't include a whimsical interventionist God. Similarly, I suppose one could insist that sugarcubes have *always* been disposed to turn into diamonds following such a divine command, and so wouldn't, strictly speaking, gain any dispositions in the circumstances described. Still, it surely isn't the case that sugarcubes have always had that disposition even partly in virtue of any of their intrinsic categorical properties; if they have always had it, they have always had it solely in virtue of God's power.

Picking up on that point, a defender of Mumford could just insist that in the circumstances described, sugarcubes do not have (and never had) the disposition to turn into diamonds, since any sugarcube-to-diamond transformation events are wholly caused by God and not at all by intrinsic properties of diamonds. (Note that such events plausibly *are* (partly?) caused by *relational* properties of diamonds: the property of being the subject of a relevant divine decree, for instance.) This, of course, is just to insist that dispositions are intrinsic properties of objects, which is the very point at issue.

In my view, any dispute as to whether or not sugarcubes had the disposition to turn into diamonds when placed in water in such-and-such conditions could be settled conclusively, without philosophical exertion, simply by placing some sugarcubes in water in those conditions and seeing what happened. Someone who observed the sugarcubes turn into diamonds and yet insisted that it might nonetheless be the case that they have no disposition (or tendency, or ability) to do so could not, I think, plausibly claim to be using such terms in their ordinary senses.

sic properties of water, for instance, must (on Mumford's view) be taken to play a causal role in the production of dissolving events, and so help make it true that sugar is water-soluble. And the precise relation between certain properties of water and of sugar is crucial: it matters causally that sugar has hydroxyl groups with a negative charge, while water molecules contain oxygen with a positive charge.

> 2. We take precautions to prevent the breakage of fragile objects, which shows that we take them to possess some relevant property during times when the conditions for the manifestation of that property are not satisfied.

Following on from the previous point, it's worth noting that this argument is rather less forceful if rather than thinking about cases in which we straightforwardly 'mask' (Johnston, 1992) a relevant property of the object of ascription - as when we wrap a vase in bubble-wrap, say - we think about cases in which we take steps to modify the object's environment. We could put cushions on the floor around the table on which the vase stands, for instance, or (more fancifully) create a low-gravity environment within which to display the vase - which plausibly says more about our beliefs concerning the properties of hard floors and normal-gravity environments than about the properties of the vase itself.

More interesting is a case like the following. Suppose someone believes that a previously robust vase - made out of titanium, perhaps - has been cursed by a powerful warlock and, as a result, would break if dropped. So he takes extra precautions to ensure that the vase is not dropped - putting it in a locked cabinet, say. Now, this misguided soul does indeed think the vase possesses some relevant property during times when the conditions for the manifestation of that property are not satisfied. But the property in question is not an intrinsic property of the vase, but its relational property of having been cursed by a powerful warlock. So while this argument might do something to undermine a simple conditional analysis of dispositions, it does nothing at all to support the view that the concept of a disposition is the concept of a causally efficacious intrinsic property of the object of ascription.

> 3. (The *virtus dormativa* point): to say that opium has the disposition to make one sleep is to say that it makes one sleep by virtue of one of its properties; and this is to say something informative - that sleep follows the ingestion of opium not because of divine intervention, say, or by pure coincidence, but because of *something about the opium* (Mumford, 1998, p.138).

The crucial point here is just that we can perfectly well ascribe to opium the disposition to make one sleep whether or not we think it's some *intrinsic* property of opium that's doing the causal work. If a scientist with expertise in the area of narcotics were to bump into someone who believed that opium poppies had been blessed by the god of dreams, the two parties surely wouldn't disagree on the question of whether opium had the disposition to make one sleep (although see footnote 20, above). Their disagreement would be over the *causal ground* of that disposition: the scientist would point to the fact that the drug (and the brain) have certain intrinsic, categorical properties, while his opponent would point to the fact that the flower from which the drug is extracted has the relational property of having been blessed by the god of dreams.

These examples are all rather fanciful, of course; things like curses and blessings don't really do any of the causal work that gets done in the actual world. But I think they decisively undermine the claim that the *very concept* of a disposition is the concept of an intrinsic property of an object. For all that conceptual analysis can tell us, dispositions might very well be relational properties *even if* they're causally efficacious properties of the object of ascription.

## 3.5  Some further concerns

Mumford urges 'modest realism' about subject-independent reality (Mumford, 1998, p.192); we should assume, he says, that there is such a reality, but not be overconfident about our ability accurately to describe that reality. In particular, we should refrain from projecting our conceptual categorical/dispositional distinction on to the world, and assume instead that there is really just one type of property out there.

But what is modest about the claim that our pretheoretical disposition ascriptions work by neatly ascribing, to precisely the right objects, precisely the properties that a completed science would appeal to in offering causal explanations of everything that happens (albeit under merely functional characterizations)? However 'realist' we are about causation, surely this can't be right. Surely the supposition that some mentioned object has some causally efficacious intrinsic property is not the only basis on which we could ever hand someone a valid 'inference-ticket' (Ryle, 2000, p.117 *et passim*) - a licence to reason from cause to effect and from effect to cause. Quite apart from the sort of fanciful cases discussed in section 3.4, don't we say, for instance, that the rate of inflation tends to rise when interest rates fall; that public confidence in the police

is apt to drop when there are fewer bobbies on the beat; that the Conservative Party is vulnerable to smear campaigns; that democracy in Iraq is still fragile? These inference-tickets are perfectly valid; they license perfectly sound inferences. But it's not at all clear that we're here talking about causally efficacious intrinsic properties of 'the rate of inflation', 'public confidence in the police', 'the Conservative Party' and 'democracy in Iraq'. (Certainly such things don't have the sort of categorical properties - physical, microstructural properties - that Mumford thinks are the causal grounds of dispositions.)

Perhaps Mumford would dismiss such cases as 'atypical' and hence beside the point of a general theory of dispositions. This is something he does rather a lot. 'Abstract' and 'ungrounded' dispositions, such as (respectively) the number eight's divisibility by two or the charge of a subatomic particle, also escape his account; so too (or so it sometimes appears) do objects' dispositions to affect sentient beings in certain ways. Mumford suggests, for instance, that where an object's function is essentially dependent on our responses - like the function of a road sign to instruct or of a flag to add grandeur - 'it would seem a mistake to ascribe a disposition to such an object' (Mumford, 1998, p.203). Similarly, Mumford warns against generalizing from the case of *provocativeness* - the disposition a red cape has to anger a bull - as this 'should be classed among judgment- or response-dependent dispositions'; the 'extra element of response-dependence' in such cases apparently makes it fundamentally different from cases like that of solubility (Mumford, 1998, p.205).

Mumford is, of course, using the term 'response-dependence' to mean a dependence on the responses of sentient beings. But it is worth noting that *all* dispositions, on Mumford's account, are response-dependent in a perfectly good sense: they are dependent on the 'responses' of objects to events and to the instantiation of certain properties. If water molecules did not respond to the presence of sugar in a certain way, sugar would not be soluble; if vases did not respond to being dropped in a certain way, vases would not be fragile; and yes, if bulls did not respond to the presence of a red cape in a certain way, red capes would not be provocative. Mumford presumably thinks there's something especially contingent about the responses of sentient things like human beings and bulls, as opposed to things like vases and water molecules, but so far as I can see he does not do anything to back up this intuition. And since he endorses the Humean principle that the laws of nature are contingent - it is always a contingent matter what follows what - I fail to see how he could do so.

Even if Mumford has an answer to this point, however, note that he apparently excludes a huge range of dispositions from his account; an account which, as

we have just noted, already seems unable to accommodate many everyday uses of disposition ascriptions. The points he makes about provocativeness and the functions of flags and road signs seem to apply equally to the dispositions things might have to look, taste, feel, smell or sound a certain way to human or non-human subjects, to produce aesthetic, emotional, intellectual or sexual responses in human or non-human subjects, and so on. In fact it looks very much as if psychologists, social scientists, anthropologists, zoologists, aestheticians, perfume designers, economists, animal trainers, chefs and anyone else who is not concerned strictly and solely with the interaction of unconscious lumps of matter is going to have to look elsewhere for an account of the way the disposition ascriptions they use every day actually work.

Mumford's treatment of 'ungrounded' dispositions is also of concern. Such cases - which Mumford concedes may very well lie at the end of any chain of explanation, once we get down to the level of subatomic particles with no structural properties - are indeed atypical; but they are also absolutely fundamental. In saying that most dispositions have categorical bases and so our theory of dispositions should be shaped accordingly, Mumford seems to me rather like the man who insists that the earth is supported 'all the way down' by a tower of turtles; the fact that such a 'support from below' theory has to accommodate only one, atypical, unsupported turtle at the bottom of the tower does not make it any more appealing. If we recognize that there can be objects that are unsupported from below, plainly we may as well skip the turtles entirely and theorize that the earth is itself unsupported from below. Similarly, if we recognize that we are perfectly able usefully and correctly to ascribe dispositions to objects even if we do not take those dispositions to be causally grounded in categorical properties of those objects, there is just no compulsion to identify dispositions with their categorical bases in order to explain the evident value of disposition ascription in causal explanation.

## 3.6 Conclusion

In this chapter I have left a number of threads hanging. Having granted to Mumford, for the sake of argument, that there may be property instances which are causally efficacious of events in the physical world, I have been obliged to leave open the possibility that dispositions may be identical with things like sets of property instances (subsection 3.3.5), or categorical properties of the whole world (section 3.4), as well as the possibility that relational properties of objects may be causally efficacious (section 3.4). I hope, however, that I have

succeeded in closing off the possibility that dispositions are simply identical with - or, I think, constituted by or realized in - intrinsic, categorical properties of objects, 'instantiated properties which inhere completely within the object of ascription' (Mumford, 1998, p.74); that view, I have argued, is one that we have good reasons to reject even if we judge Mumford's property instance ontology to be acceptable.[21]

This is emphatically *not* to reject the view that dispositions are properties of the objects to which they are ascribed; they are. It's just that they are *relational*, and not intrinsic properties of those objects. To have such-and-such a disposition, as I suggested in section 3.4, is just to stand in such-and-such a relation to certain actual and possible objects and events. The relationship between the dispositional and categorical properties of objects, I suggest, is perhaps best understood in terms of what we might call *transglobal supervenience* rather than identity; once the Gods of all possible worlds have dished out to all the objects in those worlds all their categorical properties, fixed the laws of nature in those worlds, and promised not to interfere in the day-to-day running of things, the question of what dispositions objects have has been settled. In Chapter 4, I will consider some of the implications for the philosophy of mind of taking this sort of relational view of the mental states we ascribe to one another.

---

[21]My own inclination would be to refrain from attributing causal *efficacy* to properties and their instances (whether dispositional or categorical), while asserting their causal *relevance*. On this view, it certainly *matters* causally to certain dissolving-events both that this sugarcube instantiates the property $m_3$, and that this sugarcube instantiates the property of being water-soluble; but neither property instance should be credited with *causing* those dissolving-events. The question is not, therefore, what roles are played by dispositional and categorical property instances in generating events, but rather what roles are played by ascriptions of dispositional and categorical properties in explaining and predicting events. These themes are explored in more detail in Chapter 4.

# 4 Mental States and the Causal Explanation of Behaviour

## 4.1 Introduction

In the previous chapter, I argued that dispositional properties are not, as is often supposed, intrinsic properties of the objects to which they are ascribed. Rather, they are relational properties of those objects: to have such-and-such a disposition is to stand in such-and-such a relation to certain other actual or possible objects and events. It is therefore a mistake to think that the categorical features of the world which are the causal ground of a given dispositional property must all be intrinsic features of the object of ascription. Rather, they may be widely distributed. An object's dispositional properties supervene not 'locally' on its own categorical properties, but 'transglobally' on the categorical properties of objects in its own world as well as in other possible worlds.

In this chapter, I want to consider the implications all this has for the philosophy of mind. If mental properties such as *believing that P* are dispositional properties, and if these dispositional properties are relational properties, and if these relational properties supervene partly on the categorical properties of objects other than the object of ascription - in this case, the subject who is said to have such-and-such beliefs, etc. - well, what then? In what sense, if any, are things like beliefs and desires *states* of the subject to which they are ascribed? Are explanations of subjects' behaviour which cite their mental states *causal* explanations? Are a subject's beliefs and desires the causes of her behaviour?

In section 4.2, I will try to get clear on what the dispositionalist should say about the relationship between subjects' mental states and their brain states, in light of my rejection in Chapter 3 of Mumford's functionalist position identifying dispositional properties with categorical properties of the object of ascription. Then, in Section 4.3, I will draw on Lynne Rudder Baker's 'Practical Realism' to show how I think the dispositionalist can reconcile a rejection of functionalist arguments with a defence of the reality and causal relevance of mental states

- without invoking any 'ghosts in the machine'.

## 4.2 Occupants of causal roles again

In Chapter 2, I suggested that if one takes dispositional properties to be the occupants of causal roles, phenomenal dispositionalism is essentially a variant of functionalism. Having considered Mumford's 'functionalist' account of dispositions in Chapter 2, it should now be easy to see why this is so.

Consider this influential functionalist argument for the identity of mental states and brain states:

> Mental state *M* = the occupant of causal role *R* (by definition of *M*).
>
> Neural state *N* = the occupant of causal role *R* (by the physiological theory).
>
> [Therefore:] Mental state *M* = neural state *N* (by transitivity of =).
>
> (Lewis, 2002, p.88)

As noted in Chapter 3, this argument is simply what one gets if one applies Mumford's general 'argument from identity of causal role' to the particular case of dispositional *mental* states and categorical *brain* states. Someone who subscribed to Mumford's view of the relationship between dispositional and categorical properties in general would, I think, have no difficulty in subscribing to Lewis's view of the relationship between mental states and brain states in particular. (I'm not sure much turns here on whether one runs the argument in terms of properties, like Mumford, or in terms of states, like Lewis.)

In light of the discussion in Chapter 3, it should be clear that phenomenal dispositionalism, as I understand it, is *not* a functionalist position. In putting forward his particular argument from identity of causal role, Lewis makes, in my view, precisely the mistake Mumford makes in his more general argument. The problem, again, is that - as I put it in Chapter 3 - 'the concept of occupancy is, in general, such that if x and y can both be legitimately described as being "the occupant of Z", and if there aren't two wholly distinct things which are the occupant of Z, then it may nonetheless be that x is not identical with y'. So the conclusion simply does not follow from the premises. It is perfectly consistent with those premises that M is (for instance) *part* of N, or partly constitutive of N; or conversely, that N is part of M, or partly constitutive of M.

If we are to treat dispositional mental states just as we treat dispositional states in general - which seems a reasonable starting point, at least - I think what

we should say is going on here is this: just as the water-solubility of sugar is causally grounded in features of sugar, of water, and of the environment in which sugar and water interact (or *would* interact in some circumstances), so dispositional mental states are causally grounded in features of the subject to which those states are ascribed, of objects including, but not limited to, the objects those states are 'about', and of the environment in which the subject interacts (or *would* interact in some circumstances) with those objects and others.

Consider a paradigmatic dispositional mental state: Bob's lifelong fear of spiders. Just as we can prescientifically predict and explain sugar's interactions with water by appealing to its water-solubility, so we can prescientifically predict and explain Bob's interactions with spiders by appealing to his fear of spiders. Functionalist arguments of the Lewis/Mumford sort invite us to think of Bob's fear of spiders as being identical with some state of Bob's brain. But if - in light of the discussion in Chapter 3 - we ask ourselves what is the categorical basis of Bob's fear of spiders, a different picture suggests itself. For of course it is flatly false that certain categorical properties of Bob's brain are the only properties of anything that are causally relevant to the production of such manifestation events as Bob's running away when he sees a spider. Certain categorical properties of the spider are relevant, for starters - the spindliness of its legs, say. So too are some quite general properties of the environment - e.g. the presence of light, which enables Bob to see the spider.

An objector might be inclined here to argue that it is only insofar as features of Bob's environment, including the spindliness of the legs of any spiders in that environment, are represented in Bob's brain, that those features play a causal role in producing his behaviour. To this I would say two things. Firstly: the most this argument could hope to establish is that the proximal or immediate causes of Bob's behaviour are his brain states rather than features of his environment; and this by no means demonstrates that features of, or events occurring in, his enviroment play *no* causal role in the production of his behaviour. If the presence of a spider causes Bob to acquire a mental representation of a spider in front of him, and if that mental representation plays some causal role in producing spider-avoidance behaviour in Bob, then the presence of the spider plays a causal role in producing spider-avoidance behaviour in Bob. And it is in virtue of certain categorical properties of the spider, and certain features of the environment, that the presence of the spider plays that role. (The evidence for this is just that Bob responds differently to the presence of spiders in different environments, e.g. pitch black rooms, and to the presence of objects with

categorical properties different from those of spiders, e.g. kittens or cream cakes.) And secondly: since Bob's behaviour is embedded in his environment, the manifestation events for Bob's fear of spiders will be the events they are in virtue of features of that enviroment and not simply in virtue of what's going on inside Bob. If Bob squashes the spider that scares him, it is in virtue of the weakness of the spider's body, as well as the force with which Bob strikes, that that event occurs; whether or not the spider's body is represented in Bob's brain as being weak is irrelevant. So it would not be right to say even that features of Bob's environment can only be relevant to the occurrence of fear-of-spiders manifestation events if they are represented in Bob's brain.

Of course, the causal role played by intrinsic properties of Bob's brain in the production of such manifestation events is distinct from the causal role played by intrinsic properties of spiders. But for the reasons set out in Chapter 3, it would be a mistake to think that we can pick out the occupants of these distinct causal roles simply by ascribing distinct dispositions to Bob and to spiders. In fact, just as for sugar to be water-soluble *just is* for water to be sugar-solvent, so for Bob to be afraid of spiders *just is* for spiders to inspire fear in Bob (or: to be capable of inspiring fear in Bob, or: to be disposed to inspire fear in Bob). These two dispositions do not play distinct causal roles at all, and each of them is causally grounded in the very same categorical properties of Bob, of spiders, and of the world in general.

This has some interesting implications for philosophical questions concerning intentionality. Consider, first of all, the question: how is it that a state of someone's brain - a purely physical arrangement of stuff - can be *about* anything? This is a question that confronts anyone who thinks that a mental state such as *Bob's fear of spiders* is a brain state. But to someone who thinks that Bob's fear of spiders is not an intrinsic brain state at all, but rather a relational, dispositional state that is (in part) causally grounded in categorical properties of spiders, it ought to appear quite unmysterious that that state of Bob's is *about* spiders. It is *about* spiders just because it *has to do with* spiders in the right way - spiders play a particular role in the production of relevant manifestation events.

It might be felt that the anti-functionalist thrust of this discussion, and of the discussion in Chapter 3, seriously threatens the important principle that subjects' behaviour can be causally explained by citing their mental states. In section 4.3, I will draw on Lynne Rudder Baker's (1995) work in developing the position she calls 'Practical Realism' in an attempt to demonstrate that this is not the case.

## 4.3 Baker's Practical Realism

Lynne Rudder Baker's (1995) critique of what she calls the 'Standard View' has clear parallels with Ryle's critique of the 'Official Doctrine' (2000) (see Chapter 2). Indeed, Baker goes so far as to say that the Standard View - roughly, the view 'that the attitudes, if there are any, are (or are constituted by, or are realized in) particular brain states' (1995, p.5) - 'may be thought of as "Descartes without dualism"' (1995, p.6). Philosophers who subscribe to the Standard View make - in Baker's view, and in mine - the Cartesian error of construing the mind as a causal system of internal states of subjects.[1]

Baker's 'Practical Realist' alternative to the Standard View is also distinctly Rylean:

> If Practical Realism is correct, beliefs are not theoretical entities, like electrons; they are not spatiotemporal entities or internal states at all. Since the term 'belief' is just a nominalization of 'believes that', *S* has a belief if and only if there is some proposition *p* such that *S* believes that *p*. Whether *S* believes that *p* depends solely on what *S* would do, say, and think in various circumstances. Although *S* may not always manifest beliefs in behaviour, there must be some circumstances in which *S*'s belief makes a difference to what *S* would do, say, or think... The Practical Realist view of belief is this: *S* believes that *p* if and only if there are certain counterfactuals true of *S*, where the content of the counterfactuals may be intentionally characterized.
>
> (Baker, 1995, p.21)

Note that although she sticks to the language of counterfactuals in her (1995), Baker is clear that 'one may speak instead of dispositions, provided that dispositions are construed relationally and intentionally' (1995, p.21n.41). In her (2001a), Baker seems rather happier to express her claims in explicitly dispositional language - suggesting, for instance, that there is 'a set of dispositions that *constitute* [a subject's] believing that there is beer in the fridge' (2001a, p.186, my italics). It seems clear enough to me, then, that Baker's Practical Realism is a dispositionalist position. However, Baker doesn't see herself as

---

[1] I don't propose to get into the scholarly question of whether Descartes did or did not in fact endorse an Official Doctrine view of the mind as characterized by Ryle; if the 'Cartesian' views Ryle, Baker and I have in our sights are rooted in familiar caricatures or misrepresentations of Descartes' thought, then the label 'Cartesian' should simply be taken with a pinch of salt.

offering a Rylean account of belief and other attitudes. This is because she asserts something Ryle denies: that belief explanations are causal explanations.

'Causal explanation' is a slippery phrase, though, and Ryle has a quite specific sort of explanation in mind when he denies that belief explanations are causal explanations. Taking 'causal explanations' to mean 'explanations citing causes', and 'causes' to be happenings or events, what Ryle in fact denies is just that belief explanations (and motive explanations, etc.) cite particular events that stand to other events in the relation of cause to effect (2000, p.86 *et passim*). Baker seems to be aware of this - she points (1995, p.27) to Ryle's assertion that 'Motives are not happenings and are not therefore of the right type to be causes' (2000, p.109) - yet also thinks Ryle would assent to the proposition that 'Unless beliefs were brain states, they could not causally explain behaviour' (Baker, 1995, p.17; pp.27-28). However, states are not happenings any more than motives are happenings. Hence I think it's clear that Ryle would *deny* that brain state explanations are causal explanations, for just the same reason he denies that motive explanations are causal explanations.

In the sense in which Ryle denies that belief explanations, motive explanations and so on are 'causal explanations', I think Baker *also* denies that they are causal explanations. After all, Baker agrees with Ryle that such explanations do not explain by citing internal mental events like volitions or pangs of desire that precede and are the causes of people's actions; and Ryle would surely agree with Baker that equally, such explanations do not explain by citing events internal to the brain that precede and are the causes of people's actions.

The real question, then, is whether Ryle is being too restrictive in his use of the term 'causal explanation'. Do some explanations citing things other than events - things like dispositions and background conditions, perhaps - nonetheless deserve to be called 'causal'? Nested within this question is a further question: do all explanations deserving to be called 'causal' cite things that deserve to be called 'causes' (where a 'cause' is a bearer of causal efficacy)?

To the first of these questions, I am inclined - with Baker - to answer 'yes'. If I explain that the vase broke because the hammer hit it, I have cited two events that stand in the relation of cause to effect; if I explain that the vase broke because it was fragile, I have cited a disposition of the vase to break in response to stimuli such as being struck with a hammer. These are certainly different sorts of explanation, but they are nonetheless two sides of the same coin, part and parcel of the same successful explanatory practice. The latter explanation makes sense of the former; it is precisely because we know that the vase was fragile that we can recognize the events of its being struck and of

its breaking as standing in a relation of cause to effect.

If this seems less than obvious, suppose we witness the following events occur simultaneously in Bob's living room: the dog barks, the light flickers, and Bob presses the power button on his TV remote. A moment later, a picture appears on the TV screen. What enables us to identify Bob's pressing of the power button, rather than the barking of the dog or the flickering of the light, as having caused the picture to appear on the TV screen? The answer, surely, is our knowledge that the TV is disposed to display a picture when the power button is pressed (but not when the dog barks or when the light flickers). Without that knowledge of the TV's dispositional properties, we just have one event following a bundle of other events.

So to insist that explanations are causal only insofar as they cite particular events standing in the relation of cause to effect is to insist that insofar as they are causal, explanations are always incomplete - not just in the harmless sense that they do not cite each and every link in a causal chain of events, but in the pernicious sense that they remove any element of generality from such explanations, any attempt to fit particular pairs of events into intelligible, predictable patterns. I think it's reasonable, therefore, to regard both types of explanation as belonging to the same family - causal explanations. Both types of explanation draw our attention to what is 'causally relevant' to the occurrence of actual or possible events.

When it comes to the second question - do all explanations deserving to be called 'causal' cite things that deserve to be called 'causes'? - I am more hesitant. So long as one is not one of Ryle's 'addicts of the superstition that all true indicative sentences either describe existents or report occurrences' (2000, p.119), there can be no objection in principle to the everyday use of a locution such as 'a desire for money is the cause of many betrayals'; that is a good enough way to assert the causal relevance of certain people's desire for money to certain events of betrayal, putting us in a position to predict, explain, prevent, or, if we like, promote such events.[2] But addiction to that superstition remains widespread, and so I think that to talk of beliefs and desires as 'causes' - as bearers of 'causal efficacy' - invites the sort of mistaken thinking both Ryle and Baker set out to undermine. It invites us to reify attitudes, to think of them as

---

[2]Baker argues, rightly I think, that 'we know that we have an adequate causal explanation when it affords control over phenomena of the type explained' (1995, pp.121-122). For instance: if we know that we can change people's behaviour in a certain way by changing their beliefs, we know we have an adequate causal explanation of that sort of behaviour in terms of that sort of belief. Beliefs are 'real', in Baker's view and in mine, just because 'what is causally explanatory is real' (1995, p.217). (See section 4.4, however, for some concerns about the detail of Baker's account of causal explanation and 'control over phenomena'.)

entities occupying the same field of causes and effects as spatiotemporally located, physical events such as a vase's breaking or the taking of a £20 note from someone's wallet. Without wanting to refute Baker's claims for the reality of (unreified) attitudes, therefore, I will avoid such locutions myself and take the line that mental states such as the belief that P are causally relevant to (and so causally explanatory of) the occurrence of certain events, but are not themselves the causes of any events.[3]

I would like explicitly to endorse Baker's account of the sense in which things like beliefs are *states* of a person. This is another locution that has, in my view, invited much confusion by leading philosophers to think of beliefs in terms of *internal* states of their bearers - particular arrangements of the stuff of which their bearers are made.

According to Baker:

> a belief is a global state of a whole person, not of any proper part of the person, such as the brain... An attitude is a state in the attenuated sense in which a state of financial health... is a state... there are genuine facts about financial health even though 'state of financial health' does not refer to anything inside a person or to anything inside an institution like a Savings and Loan (S & L)... The state of an S & L's financial health cannot be identified with any particular internal state; financial health is a state of the S & L as a whole... So, if a state of belief is like a state of financial health, then we should not expect to identify it with any particular internal state of the believer.

> (Baker, 1995, p.154)

The typical rationale for identifying a belief with an internal state of a person (say, *brain state S*) is, I take it, the same as the rationale for identifying a belief with an intrinsic, categorical property of a person (say, *being in brain state S*). In each case, it is the causal role occupied by the relevant state or property that is thought to make it a token belief with some given content. So it is interesting to

---

[3]The distinction made here between causally efficacious states or properties on the one hand, and causally relevant or explanatory states or properties on the other, has been made even in a functionalist context - notably by Jackson and Pettit (1988, pp.395-396 *et passim*), who regard dispositional explanations as functional state explanations and functional state explanations as 'programme explanations' that work by citing a causally explanatory but inefficacious state of some object (e.g. fragility) the presence of which 'programmes for' the presence of some causally efficacious categorical state or other of that object occupying the appropriate causal role. In light of the discussion in Chapter 3, it should be clear that the use I am making of the distinction is very different from the use made of it by Jackson and Pettit.

consider cases like these in the light of my critique of Mumford's 'argument from identity of causal role' in Chapter 3: if we ask ourselves what is the 'categorical basis' of an S & L's good state of financial health, the answer is not going to be 'such-and-such an internal arrangement of S & L stuff'; we are going to have to cast our net much more widely and recognize that the financial health of the S & L supervenes on a great many categorical properties of a great many things (e.g. the structural properties of its mortgage holders' homes).

It must be emphasized that none of this is to deny the reality, the causal efficacy or the causal relevance of internal states of believers, including brain states. Of course states of subjects' brains make a difference to the way they act, think and feel, and of course these states can be the subject of legitimate scientific inquiry (just as the internal states of an S & L's mortage holders' homes can be the subject of legitimate scrutiny by property surveyors, say). The point is just that the ordinary language we use to ascribe mental states, or states of financial health, does not work by picking out and naming particular internal states of the object of ascription (or anything else) playing particular causal roles.

## 4.4 Elugardo's challenge to Baker on causal explanation

Elugardo (2001, pp.114-115) offers a counterexample to Baker's 'Control Thesis' (which spells out a sufficient condition for the the occurrence of an event F causally to explain the occurrence of an event G). I don't think Baker's attempted refutation of this counterexample (2001a, p.189) works - rather, I think it opens her conception of causal explanation up to serious objections - and so I want to offer my own thoughts on where it leaves the Control Thesis, and how the dispositionalist should respond to it.

Here is Baker's Control Thesis:

> An occurrence of F in context C causally explains an occurrence of G if: (i) if an F had not occurred in C, then a G would not have occurred in C; and (ii) given that F did occur in C, an occurrence of G was inevitable.

> (Baker, 1995, p.122)

Elugardo suggests the following as a counterexample:

> Suppose that my brother and I live in the same apartment. He is blind but can hear; I am deaf but am sighted. To help us know when someone is at the door, the apartment is electrically wired in such a way that our doorbell will ring when and only when the light bulb in our doorway, which is always on, glows dim and bright in an alternating pattern. Because of the internal circuitry, the correlation between the doorbell ring and the light bulb glowing holds as a matter of law. Consequently, you cannot cause the light bulb to alternate between dim light and bright light by pressing the doorbell button without also bringing about a situation in which my brother's eardrums vibrate (when he is home, within earshot, etc.). For, as a matter of law, doing the first is nomically correlated in this context with causing the doorbell to ring, which in turn normally causes my brother's eardrums to vibrate. We may suppose that the doorbell ringing is the only cause of the vibrations in this context. Therefore, given the facts of the case, if the lightbulb had not dimmed and then glowed brightly, my brother's eardrums would not have vibrated. And, given that the lightbulb did glow in this pattern, the vibration of my brother's eardrums was inevitable.

> (Elugardo, 2001, pp.114-115)

Elugardo argues (2001, p.115) that, contrary to the Control Thesis, the light-bulb's glowing dim and bright in an alternating pattern does not causally explain the vibrating of his brother's eardrums when the doorbell is pressed. This, he says, is because 'there is no causal path that traces the second kind of event back to the first without bypassing the causal connection between the sound of the doorbell ringing and my brother's eardrums vibrating' (2001, p.115).

Let us be clear on the imagined scenario here. No doubt there are various ways in which one could rig up a system in which the pressing of a button caused both a characteristic glowing of a light bulb and the ringing of a bell (so that there was a nomic correlation between these two events). One could set things up so that the pressing of the button closed an electrical circuit, causing the glowing of a lightbulb, while a separate device detected that glow and triggered the ringing of the doorbell. Or one could set things up so that the pressing of the button closed an electrical circuit, causing the ringing of the doorbell, while a separate device detected that ringing and triggered the glowing of a lightbulb. Or one could set things up so that the pressing of the button closed two circuits A and B simultaneously, with the closing of circuit A causing the lightbulb to glow and the closing of circuit B causing the bell to

ring.

In the *first* of these possible scenarios, plainly one *can* trace a causal path back from the vibrating of eardrums to the glowing of a lightbulb without bypassing the causal connection between doorbell ringing and eardrums vibrating (a path that, traced forwards, looks like this: lightbulb glows > light detected > doorbell rings > eardrums vibrate). So clearly Elugardo does not have a scenario like this in mind.

In the *second* scenario, there's a causal path that looks like this: doorbell rings > sound detected > lightbulb glows. But if one wanted to add "> eardrums vibrate" to the end of that path, thereby asserting a causal connection between lightbulb glowing and eardrums vibrating, the (shorter) path back from the latter event to the former would not include the ringing of the doorbell. The causal connection between doorbell ringing and eardrum vibrating is therefore missing from that shorter path, and in that sense 'bypassed'.

In the *third* scenario, two distinct causal paths can be traced forwards from the pressing of the button: button pressed > circuit A closed > lightbulb glows; and button pressed > circuit B closed > doorbell rings. To trace a (supposed) causal path back from the vibrating of eardrums to the glowing of the lightbulb, one would have to add "> eardrums vibrate" to the end of the former path. And that, again, would be to bypass the causal connection between doorbell ringing and eardrums vibrating.

Having spelled that out, I want to consider Baker's response:

> The eardrums' vibrating is caused by the doorbell's ringing, which is caused by ([nomically][4] connected to) the alternating dimming and brightening of the light, which is caused by the pressing of the doorbell.

> Since we are taking the peculiarities of the wiring to be part of the context (and thus fixed), any event along this path is a causal explanation of the eardrums' vibrating in this context. We have here a predictive/explanatory pattern (i.e., a counterfactual-supporting pattern). As Dennett has said, when one finds a certain kind of predictive pattern, 'one has ipso facto discovered a causal power - a difference in the world that makes a subsequent difference testable by standard empirical methods of variable manipulation.'

---

[4]In the text as published, this word is rendered as 'nominally'. It is clear from the context, however, that 'nomically' is correct; and indeed, this is how the word is rendered in what I take to be Baker's final draft of the same chapter (2001b, p.9), made available by Baker on her website.

(Baker, 2001, p.189; Dennett, 1998, p.112.)

To my mind, there is a distinct tension in what Baker says here. On the one hand, we have what looks like a bullet-biting assertion that the doorbell's ringing's being *nomically connected to* the lightbulb's dimming-and-brightening suffices for its being *caused by* the lightbulb's dimming-and-brightening. (After all, Baker surely isn't assuming that the context here is something like the first scenario just discussed, with the dimming-and-brightening somehow triggering the doorbell's ringing. And her Control Thesis can certainly be read, as it is by Elugardo, as entailing just such an assertion.) Yet on the other hand, we have an endorsement of 'standard empirical methods of variable manipulation' by which we test what is causing what to happen.

Here is how I think the tension creeps in. It's true enough that in Elugardo's counterexample - let's assume scenario 3 is what he has in mind - we have a predictive and counterfactual-supporting pattern such that if the lightbulb glows in a certain way, we can be sure that his brother's eardrums will vibrate. And the inference-ticket also covers the return journey; if the brother's eardrums vibrate (in the relevant way), we can be sure that the lightbulb has glowed in a certain way. Still, this does not amount to an *explanatory* pattern. This ought to be evident if one considers that there is a perfectly analogous predictive and counterfactual-supporting pattern relating the ringing of the doorbell, rather than the glowing of the lightbulb, to the vibrating of the brother's eardrums: if the doorbell rings, we can be sure that the brother's eardrums will vibrate, and if the brother's eardrums vibrate (in the relevant way), we can be sure that the doorbell has rung. For according to the assumptions underpinning 'standard empirical methods of variable manipulation,' the claim that the vibrating of the brother's eardrums is causally explained by the glowing of the lightbulb is in competition, in the imagined scenario, with the claim that the vibrating of the brother's eardrums is causally explained by the ringing of the doorbell. Both claims are based on the recognition of a predictive, counterfactual-supporting pattern of nomic connections between the glowing of the lightbulb, the ringing of the doorbell and the vibrating of the brother's eardrums. Yet one of those claims might be right and the other wrong. By manipulating variables - e.g. removing the lightbulb or muffling the doorbell - we can work out whether the glowing of the lightbulb really does cause the vibrating of the brother's eardrums, or whether it is just nomically correlated, for some reason, with ringings of the doorbell or vibratings of the brother's eardrums.

So Elugardo's criticism of Baker's Control Thesis hits its target, I think: there is a difference, not captured by that thesis, between nomic correlation and causal

explanatoriness. The former does not suffice for the latter, although it does license the making of predictions that such-and-such an event will happen or inferences that such-and-such an event has happened.

It seems to me, however, that Practical Realism - and dispositionalism more broadly - can survive this attack. In fact, I think the failure of the Control Thesis reflects a failure to give dispositions their due by focusing too narrowly on the particular context of a causal explanation. If the occurrence of F in context C is causally to explain the occurrence of G in that context, it is not sufficient that some nomic correlation between Fs and Gs exists *in just that context* (as it does between glowings of the lightbulb and vibrating of the brother's eardrums in the scenario considered above); rather it must exist in some suitable range of contexts of the sort we might create if we were manipulating variables in order to settle on a *correct* causal explanation. That the brother's eardrums are disposed to vibrate when and only when the lightbulb glows *in just that context* tells us nothing about the role of the lightbulb glowing in causal explanations of his eardrums vibrating; only by learning what his eardrums are disposed to do in *other* contexts can we find our way to the right explanation. Once we understand that these particular eardrum-vibration events fit into a pattern of similar eardrum-vibration events accompanying doorbell-ringing events, and not a pattern of similar eardrum-vibration events accompanying lightbulb-glowing events, *then* we have a causal explanation of the eardrum-vibrating events that passes Baker's own test of 'afford[ing] control over phenomena of the type explained' (1995, pp.121-122) in a meaningful way. We now know, for instance, that we can bring about similar eardrum-vibrating events by sounding a similar bell, but not by creating a similar lightshow.

According to the version of phenomenal dispositionalism I am advocating, the causal explanation of a subject's actions, thoughts and feelings in terms of her dispositional mental states is similarly all about fitting her particular actions, thoughts and feelings into patterns - about building up intelligible pictures of how she typically responds to situations like *this*, and *this*, and *this*. To get the pattern wrong - to be wrong about the relevant features of the subject's *dispositional profile* (see Chapter 7) - is to get the explanation wrong.

## 4.5 Conclusion

In this chapter, I have argued that dispositional mental properties, just like other dispositional properties, can and should be conceived of as relational

properties - and that conceiving of them as such does not threaten their causal relevance to subjects' actions (and thoughts and feelings). Things like beliefs and desires *do not* conceptually occupy causal roles that are in fact occupied by certain states of subjects' brains; rather, to have the mental states one has is just to be disposed to act, think and feel certain ways in certain circumstances. The causal grounds or categorical bases of these dispositions, far from being wholly located within the subject to which they are ascribed, are widely distributed. Nonetheless, dispositional mental states are real states of subjects that can serve causally to explain *why* those subjects act, think and feel as they do.

This chapter marks the end of what has been a groundwork-laying phase of this thesis. In the next chapter, I will turn to a detailed consideration of the phenomenal dispositionalist position itself - as formulated, in the first instance, by Eric Schwitzgebel (2002).

# 5 Schwitzgebel's Phenomenal Dispositionalism

## 5.1 Introduction

I suggested in Chapter 2 that the dispositionalism of Ryle (2000) is considerably more liberal than has generally been recognized; far from trying to reduce ascriptions of mental states to ascriptions of dispositions to exhibit certain observable behaviour in certain circumstances, Ryle in fact acknowledges that being in such-and-such a mental state might involve being disposed not just to act a certain way, but also 'to make certain theoretical... and imaginative moves [and] to have certain feelings' (Ryle, 2000, p.129).

Schwitzgebel (2002) offers an account of belief in this liberal tradition. On his phenomenal dispositionalist view, believing that P is a matter of having not just certain behavioural dispositions, but also certain cognitive dispositions and certain phenomenal dispositions - dispositions, respectively, to reason in certain ways and to have certain sorts of conscious experiences.

In sections 5.2 and 5.3, I want to set out Schwitzgebel's position in some detail and with a minimum of criticism and commentary (though I will clarify or expand on certain points as I feel appropriate - notably in section 5.3, in which I will consider Schwitzgebel's response to an apparent tension between dispositionalism and externalism about mental content. I think this response invites several objections that are worth seeing off in order to defend phenomenal dispositionalism against the charge that it cannot be reconciled with externalism and so must be false). There is, in my view, much of value in Schwitzgebel's account, and it is well worth following his line of thinking in an unbroken way. Once we have a clear picture of how that account is supposed to hang together, however, I will go on to argue (in Chapter 6) that one of the pillars of that account - the 'dispositional stereotype' - will not bear the weight Schwitzgebel places upon it.

I should note that Schwitzgebel has recently (2013) generalized his phenome-

nal dispositionalist account of belief to cover attitudes in general - and not only propositional attitudes such as believing and desiring, but also a whole range of other attitudes such as resenting, appreciating, loving and valuing.[1]However, I think it will be beneficial in this chapter to take a relatively narrow but deep view of his position by looking at his account of belief in particular, rather than a relatively broad but shallow view by looking at his account of attitudes in general. We will not, I think, misunderstand Schwitzgebel's current position if we treat his (2002) as a case study of belief *qua* paradigmatic attitude, rather than as an account of belief *qua* belief.

## 5.2 Schwitzgebel's account of belief

In his (2002), Schwitzgebel offers a phenomenal dispositionalist account of belief as a pragmatic alternative to prevailing representationalist accounts - accounts, that is, according to which 'to believe something is to have a representation of some sort in one's mind' (Schwitzgebel, 2002, p.249). I say 'pragmatic alternative' because Schwitzgebel does not set out to tell us what beliefs ultimately *are*, metaphysically speaking; rather, he sets out to tell us something about the way belief ascription works (or ought to work) in practical terms when we ask ourselves whether some particular subject does or does not have some particular belief. Nor does Schwitzgebel put forward his account as being preferable to representationalist accounts simply by virtue of being right where they are wrong; rather, he suggests only that a phenomenal dispositionalist approach might be more useful to adopt in certain contexts:

> Representational approaches to belief have played an important role in cognitive science, and this account is not meant to displace them, but to supplement or complement them. In some contexts, I believe, the present account will prove more useful; in others, a representational approach will work as well or better.

> (Schwitzgebel, 2002, p.270.)

So Schwitzgebel makes fairly modest claims for phenomenal dispositionalism; it's supposed to be a way of thinking about belief ascription that might sometimes be useful, and that's all. The claims I am making in this thesis are less modest. I am claiming that the identification of states like belief with internal

---

[1]Schwitzgebel also assumes that character traits can similarly be analyzed in dispositional terms. So his account is intended, I think, to cover the same ground as mine, although I have chosen to lump attitudes and character traits together under the heading of '(dispositional) mental states'.

states of the subject of ascription - whether those states are characterized as representations having a certain content, in terms of their causal role, or in any other way - is a simple mistake. It must be emphasized that this is *not* to claim that there are no internal states of subjects that might be usefully and correctly characterized by cognitive scientists in terms of their causal role or representational content; that is for those scientists themselves to decide. It is to claim only that our ordinary, folk psychological language of mental states does not work by picking out these very states, providing us with a ready-made taxonomy that we can be assured will bear up under scientific scrutiny. Rather, that language works by helping us to identify and describe predictable, intelligible patterns of behaviour, of reasoning and of conscious experiences.

Schwitzgebel's central claim is expressed in the following passage:

> To believe that P. . . is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P. What respects and degrees of match are to count as "appropriate" will vary contextually and so must be left to the ascriber's judgment.
>
> (Schwitzgebel, 2002, p.253.)

Something needs to be said, then, about just what a dispositional stereotype is supposed to be and just how it is that the appropriateness of belief ascription is supposed to vary contextually.

Schwitzgebel defines a stereotype as 'a cluster of properties we are apt to associate with a thing, a class of things, or a property' (2002, p.250). By way of example - adapted from Putnam (1975, cited in Schwitzgebel, 2002, p.250) - he suggests that stereotypical properties of tigers include such things as *being striped* and *being four-legged*. A tiger that is not striped or four-legged deviates in those respects from the stereotype for being a tiger.

A dispositional stereotype for *having the belief that P*, then, is a cluster of dispositional properties we are apt to associate with the possession of that belief. (The 'are apt to' here is important, as we shall see in Chapter 6.) Crucially, dispositional stereotypes for beliefs do not include only behavioural dispositions. They also include *cognitive* dispositions, characterized as 'dispositions to draw conclusions entailed by the belief in question or to acquire new desires or habits consonant with the belief', and *phenomenal* dispositions, characterized as 'dispositions to have certain sorts of conscious experience' (Schwitzgebel, 2002, p.252). Taking the belief that there is beer in the fridge as an example, Schwitzgebel suggests that the following dispositional properties should be seen as belonging to the stereotype for the possession of that belief:

> the disposition to say, in appropriate circumstances, sentences like 'There's beer in my fridge;' the disposition to look in the fridge if one wants a beer; a readiness to offer beer to a thirsty guest; the disposition to utter silently to oneself, in appropriate contexts, 'There's beer in my fridge'; an aptness to feel surprise should one go to the fridge and find no beer; the disposition to draw conclusions entailed by the proposition that there is beer in the fridge (e.g., that there is something in the fridge, that there is beer in the house); and so forth.

> (Schwitzgebel, 2002, p.251)

Note that we do not need consciously to have made an association between the possession of the belief that P and the possession of some dispositional property in order for that property to belong to the dispositional stereotype for believing that P. It may never have occurred to us to associate believing that P with acting, thinking or feeling thus-and-so in such-and-such circumstances. It may never have occurred to us, after all, that someone who believes that P might find themselves in those circumstances; indeed, it may never have occurred to us that someone might believe that P in the first place. This is why the properties within a dispositional stereotype must be properties we are *apt to* associate with the possession of a particular belief, and not simply properties we *do* associate with it - properties included on a determinate mental 'checklist' we carry around in our heads.

Some of the properties belonging to a dispositional stereotype are, according to Schwitzgebel, more 'central' to it than others, in the sense that there would be wide agreement that the possession of those dispositional properties is part and parcel of believing that P. Others will be more marginal. For instance, it might seem obvious to almost everyone that someone who believes that there is beer in the fridge is thereby disposed to look in the fridge if he wants a beer; but it might seem less than obvious to some people that someone who believes there is beer in the fridge is thereby disposed to offer beer to a thirsty guest. (Perhaps they think people are generally selfish, or think it's usual to offer tea or coffee to a thirsty guest rather than beer.) In assessing whether or not someone has the belief that there is beer in the fridge, therefore, his having or lacking the former, more central dispositional property carries more weight than his having or lacking the second; we would be fairly quick to conclude that someone who fails to look in the fridge for a desired beer did not believe that there was beer in the fridge, but would hesitate to conclude that someone who offered a thirsty guest tea rather than beer lacked that belief.

Schwitzgebel acknowledges that 'the dispositions in belief stereotypes hold only *ceteris paribus*' (2002, p.253). A subject will in fact only be disposed to offer beer to a thirsty guest if various conditions hold: if he realizes that his guest is thirsty, if he is not a generally miserly individual, if his route to the fridge is not blocked, and so on. If those conditions *don't* hold, then his failure to offer beer to a thirsty guest cannot be seen as counting against his possessing the belief that there is beer in the fridge; rather, the non-manifestation of that disposition must be seen as being appropriately excused. The non-manifestation of that particular disposition on that particular occasion does not, after all, give us any reason to think that the subject deviates from the dispositional stereotype for believing that P in any systematic, general way.

Schwitzgebel resists the idea that we must be able to spell out just what are the 'excusing conditions' for the non-manifestation of a given disposition - the conditions in which 'the *ceteris paribus* clause is sprung' (2002, p.256) - suggesting that this is something best grasped intuitively. It is, after all, impossible in general fully to specify the conditions in which an ordinary or scientific generalization will hold, yet such generalizations can still reasonably and usefully be made. (To take an example from Schwitzgebel: 'Rivers erode their outside bank at a bend *if* the river is not frozen, *if* the bank is made of an erodible material, *if* there isn't a powerful fan in place preventing the river from touching the outside bank, etc.' (2002, p.254, Schwitzgebel's italics).) The central point here is familiar enough, and was touched on in Chapter 2: nobody has a particular, specifiable set of dispositions simply by virtue of having a particular, specifiable belief. What dispositions they have will depend on the complex relations between that belief and their other beliefs and desires, as well as their intellectual capacity, habits of thought, character traits and so on. As we have seen, this is one of the reasons for the failure of the behaviourist attempt to reduce all talk about the mind to talk about outwardly observable behaviour. Schwitzgebel (2002, p.258) emphasizes the point that phenomenal dispositionalism serves no such reductive agenda, and that in saying what it is for a subject to believe that P, appeal to other features of his mental life are therefore perfectly permissible.

The foregoing has hopefully served to clarify just what a dispositional stereotype is supposed to be. Now I want to say something about the grounds on which we might judge the ascription of a particular belief to a particular subject to be either appropriate or inappropriate.

Schwitzgebel emphasizes the fact that the grounds on which we make judgements about the appropriateness of belief ascription vary contextually - from

one set of circumstances to another. But before I go on to talk about how that is, I would like to make explicit a point which is made only implicitly in Schwitzgebel's paper (and in the foregoing discussion). The point is that the grounds on which we make judgements about the appropriateness of belief ascription also vary from one *subject* to another. We have already considered a case in which it seems equally appropriate to ascribe the very same belief to two subjects whose dispositional profiles differ greatly, and in ways that are clearly relevant to their possession of that belief.[2] (A subject's 'dispositional profile' is just the set of all her behavioural, phenomenal and cognitive dispositions. I shall argue in Chapter 6 that phenomenal dispositionalism is best formulated in terms of dispositional profiles themselves, rather than in terms of dispositional stereotypes against which subjects' dispositional profiles are judged; but let us not get ahead of ourselves.) And this case is by no means exceptional; clearly, subjects differ from one another with respect to what we might call their 'fine-grained' mental properties - their beliefs, desires and so on, as characterized in terms of their particular content - as a matter of course. When one considers in addition the sort of 'coarse-grained' differences there are between subjects belonging to different broad types - between a credulous child and a sceptical adult, a hot-tempered criminal and a cool-headed lawyer, a self-assured businessperson and a self-absorbed poet - it's abundantly clear that the dispositions we'd expect to go along with any given belief will vary enormously from subject to subject.

Moving on to the question of context: why might it be correct to ascribe the belief that P to a subject, S, in one context but not in another? If we're serious about analyzing belief in dispositional terms, after all - and hence about thinking of beliefs as things that stick around, waiting to manifest themselves, while their possessors go from one situation to another - surely we don't want to suggest that some long-held belief of a subject might be routinely popping in and out of his head as he goes about his business?

Well, no, we don't want to say that. Schwitzgebel's suggestion is not that S might believe that P in one context and yet fail to believe that P in another. Rather, the picture looks like this: S has some particular dispositional profile, and this doesn't change from one context to the next.[3] But the people with

---

[2] Relevant, that is, in terms of the degree to which, and the respects in which, they conform or fail to conform to the dispositional stereotype for that belief.

[3] This is a simplification, of course. No doubt subjects' dispositional profiles are changing all the time; someone might suddenly acquire a new disposition upon learning a new fact, or gradually lose a disposition as she cultivates different habits. But for present purposes, we may suppose that S's dispositional profile stays the same in all respects relevant to the ascription of the belief that P for some prolonged period of time.

whom S interacts - and who will hence sometimes be concerned to understand something about the way S can be expected to act, think and feel in this or that situation - just aren't in a position to know everything pertinent that there is to know about that dispositional profile. Indeed, even if they *were* in that position, it would hardly be practical to tip one another off about all the pertinent features of that profile one by one. ('Should you give the job to Bill? Well, you should bear in mind that in circumstances $C_1$, he's disposed to A; in circumstances $C_2$, he's disposed to B...') But as natural-born experts in interpreting one another's behaviour, with an instinctive grasp of the way certain dispositions go together, they have no trouble at all in identifying - in a rough-and-ready, provisional way - the sort of patterns into which S's behaviour seems to fit. That is to say, they have no trouble at all in identifying the dispositional stereotypes to which S more or less closely conforms.

It's clear from Schwitzgebel's account that a typical subject S will fail *perfectly* to conform to very many dispositional stereotypes. He might act, think and feel just as we would expect someone who believed that P to act, think and feel in circumstances $C_1$ and $C_2$, say, but deviate from expectations in circumstances $C_3$. His dispositional profile hasn't changed, though; he has the very same dispositions on each occasion. It's just that the dispositions which are manifested in circumstances $C_3$ are surprising in a way that those manifested in $C_1$ and $C_2$ are not.

It will be helpful, I think, to put some flesh on these bones in the form of an example. Suppose Bill is a civil servant advising the government on transport policy, and has all the relevant facts at his fingertips regarding the relative dangers of different forms of transport. On the basis of these facts, he affirms in written reports that plane travel is safer than train travel. Outside of work, too, if someone asks him what's the safest way to travel from London to Paris - train or plane - he'll reply, quite sincerely: 'plane'. And if he knows a family member is travelling by plane, he feels less anxious about their safety than he does if they're travelling by train. Still, when he's travelling by plane himself, he feels more anxious than he does when he travels by train; he finds himself imagining the plane crashing, and tries to distract himself by doing crossword puzzles and watching films. Sometimes he even *avoids* plane travel because it just doesn't feel safe, and takes the train instead.

Here, then, we have someone who in very many circumstances acts, thinks and feels just as we'd expect someone who believes that plane travel is safer than train travel to act, think and feel. Indeed, being an expert on these matters, he's more confident in this belief - in one sense - than many of the people with

whom he shares it. Yet in certain other circumstances, he doesn't act, think and feel as we'd expect someone with that belief to act, think and feel at all. In those circumstances, indeed, he might strike us as someone who conforms quite nicely to the dispositional stereotype for believing that train travel is safer than plane travel.

Now, we could just say that Bill's beliefs change from moment to moment; the belief that plane travel is safer than train travel simply pops out of his head sometimes, is temporarily replaced by a contrary belief, and then pops back in again. The trouble with this, though - as Schwitzgebel points out (2002, p.261) - is that we typically think of beliefs as persisting between their manifestations. We think that Bill believes *something* about the relative safety of travel by train and by plane even while he's asleep, or while his mind's on other things. Beliefs are usually seen, after all, if not explicitly as dispositions then at least as persistent mental states of some sort. And it's not as if Bill's dispositions are actually changing in relevant ways from one situation to the next; it's still true of Bill even while he's writing a report affirming the relative safety of plane travel that he has the disposition to become anxious when travelling by plane, to travel instead by train when possible, and so on. And it's still true of Bill, even while he's sitting on a plane fretting about the availability of parachutes, that he has the disposition to recommend to his daughter, when he talks to her later that month, that she take the plane back from Paris.

There are number of ways in which a well-informed interpreter of Bill might try to sum up what's going on in his case. She might say, for instance:

- He knows deep down that plane travel's safer than train travel, but he doesn't quite *believe* it.

- He believes that plane travel's safer than train travel, although sometimes he doesn't quite feel convinced of it.

- On a rational level, he believes that plane travel's safer than train travel, but on a gut level, he doesn't.

- When he thinks about it, he believes that plane travel's safer than train travel, but instinctively, he just 'knows' it isn't.

- He believes that plane travel's safer than train travel, but he only *partly* believes it.

- He's torn between believing that plane travel's safer than train travel, and believing the exact opposite.

. . . as well as:

- Sometimes he believes that plane travel's safer than train travel, and sometimes he doesn't.

Some of these might strike us as coming nearer the mark than others, and each of them could clearly be made more illuminating by being supplemented with some additional information about Bill's dispositional profile - about his disposition to feel anxious when flying, for instance. But what they all show, I think, is that it's simply not appropriate, when discussing Bill's case from a general perspective, either to ascribe to him or to refuse to ascribe to him - without qualification - the belief that plane travel is safer than train travel. His dispositional profile is such that he is 'in between' having that belief, and lacking it (Schwitzgebel, 2002, p.261 *et passim*; 2001).

So much for the general perspective. What about *other* perspectives? If Bill's boss wants to know his opinions on the relative safety of plane and train travel, should we hesitate to tell him quite straightforwardly that Bill believes the former is safer than the latter? No. Why would we? We can tell him everything he wants to know about Bill's dispositional profile - about the sort of advice he's disposed to include in written reports, etc. - just by making that simple, unqualified belief ascription. It's just not to the point to allude to certain dispositions of Bill's that are at odds with that belief. If, on the other hand, a friend of Bill's wants to know his views on the matter in order to begin planning a holiday together, we may very well want to refrain from making that belief ascription, at least in an unqualified form; in *that* context, after all, the respects in which Bill deviates from the dispositional stereotype for the belief in question are clearly relevant.

It's worth noting that there's nothing especially puzzling about the fact that Bill deviates from the dispositional stereotype for believing that plane travel is safer than train travel in the ways he does. We understand very well that people don't always act, think and feel as it appears they ought to act, think and feel if they believe what they say they believe (or what their behaviour in general suggests they believe). Indeed, Schwitzgebel identifies several familiar patterns of deviation from dispositional stereotypes, which I will briefly run through here:

- *Modularized believing* - as when a subject has procedural rather than declarative knowledge in a certain domain, so that she might fail to assent to the proposition that P even though some of her behaviour clearly conforms to the dispositional stereotype for believing that P. For instance, a person may speak in a way that suggests she grasps some rule of grammar, yet fail to assent to the proposition that such a rule obtains.

- *Unconscious beliefs* - as when aspects of a subject's behaviour (perhaps

even, in special circumstances such as hypnosis, verbal behaviour) are consonant with the belief that P, but the subject refuses in normal circumstances to acknowledge even to herself that she believes that P.

- *Low confidence* - when a subject is uncertain whether P, she can be expected to deviate from the dispositional stereotype for believing that P in certain ways; e.g. she may hesitate to act on the basis that P where the stakes are high, or feel little surprise when it turns out that not-P.

- *Self-deception* - cases of self-deception, Schwitzgebel suggests, may be a subset of cases of unconscious believing, with a subject refusing to assent to a proposition which certain of her dispositions suggest she takes to be true. It's not clear to me, though, that this is the only way in which someone might be self-deceived. It seems plausible, for instance, that a subject might convince herself that certain of her earlier actions were not based on the belief that P, when in fact they were based quite consciously on that belief at the time.

- *Unreflective inconsistency* - sometimes a subject just 'fails to put two and two together' (Schwitzgebel, 2002, p.264). Someone might sometimes say he believes, and in some respects act as if he believes, that birds are the only animals that lay eggs (say), and yet acknowledge when prompted that reptiles lay eggs too.

- *Peripheral ignorance*: if a subject just doesn't know certain facts related to the belief that P, she will deviate from the dispositional stereotype for that belief in predictable ways. For instance, someone who believes that Joe plays the clarinet but doesn't know the clarinet is a reed instrument will not have the disposition to assent to the proposition that Joe plays a reed instrument.

- *Developing beliefs*: while a subject is in the process of acquiring knowledge in some particular domain and fitting together the knowledge she has, she may deviate from the dispositional stereotypes for beliefs concerning that domain for the sort of reasons just discussed (in relation to unreflective inconsistency and peripheral ignorance).

- *Partial forgetting*: a subject may be part-way towards forgetting a once-familiar fact - the PIN on a debit card she no longer uses very much, say - such that she fails to conform neatly to the dispositional stereotype for believing that the PIN in question is 7701. Maybe she can no longer recall the number when she looks at the card, for instance, but is still disposed to enter it correctly when she puts the card into a cash machine.

Those who are inclined to reify propositional attitudes as causally efficacious 'token states' of subjects are seemingly obliged to insist that in all such 'in-between' cases, a subject either determinately has or determinately lacks the belief in question, and to go on from there to explain why that subject has certain surprising dispositions - presumably in terms of a causal story invoking other of their token mental states.[4] Schwitzgebel's response to such cases is quite different:

> Talk about belief is useful because people with some of the disposi-tions in a stereotype will tend to have many of the other dispositions in that stereotype. Such regularities allow us to make generaliza-tions and inductions on the basis of these stereotypes, and it is enormously convenient, even indispensable, to appeal to beliefs in describing our mental lives. Still, when there *is* a breakdown in the match between stereotype and the actual dispositional set of a subject, as will often happen in cases of the sort described above, simple belief talk may no longer be appropriate, and appeals to the stereotype may have to be replaced with more complicated ap-peals to specific dispositions or sets of dispositions or to recogniz-able patterns of deviation. On the [phenomenal dispositionalist] ac-count... once the dispositional profile of the subject is made clear, it is a mistake to think that there is still some further question to be answered, namely, what does the subject *really* believe?

(Schwitzgebel, 2002, p.266, Schwitzgebel's italics)

## 5.3 Phenomenal dispositionalism and externalism

There is a good sense in which phenomenal dispositionalism is an externalist position. On the view I am proposing, it is certainly not true that the contents of a subject's mental states are determined merely by the way things are 'in her head'; it matters too how things are in her environment.

However, Schwitzgebel (2002) notes an apparent tension between externalist and dispositionalist accounts of mental content. He expresses this tension by considering Wayne, an occupant of Earth, and Dwayne, his counterpart on Twin Earth - a planet that is identical to Earth save for the fact that in place

---

[4]These states might or might not be states our everyday language has a name for; in Chapter 9, I will consider Tamar Gendler's (2008a) argument for the recognition of a mental state type she calls 'alief', which she believes can explain a large range of behaviour that is not explicable in terms of familiar states such as belief and imagining.

of water ($H_2O$), it has twin water (XYZ) (Putnam, 1975, cited in Schwitzgebel, 2002, pp.266-267). Being molecule-for-molecule identical (if we ignore the difference between the water in Wayne's body and the twin water in Dwayne's), isn't it the case - assuming physicalism - that Wayne and Dwayne must have the very same dispositions? If so, a dispositional account of belief is surely bound to ascribe to Wayne and Dwayne beliefs with the very same contents, even when this runs contrary to externalist intuitions that their beliefs have different contents.

To put a little flesh on these bones: according to the dispositionalist, for Wayne to believe that water is potable is just for him to be disposed, in certain circumstances, to assent to the statement 'water is potable', to pour himself a glass of water, and in general to act, think and feel much as we would expect someone who had that belief to act, think and feel. But since Dwayne is physically identical to Wayne, he must also be disposed to do those same things in those same circumstances: if he were standing in Wayne's shoes, on Earth, he would act, think and feel just as Wayne would. So the dispositionalist is seemingly obliged to say that Dwayne, right now, sitting on his sofa on Twin Earth, believes that water is potable. And this runs contrary to the externalist intuition that Dwayne can't possibly have any beliefs about water, since he has never encountered it.

In response to such concerns, Schwitzgebel insists that there are, in fact, certain dispositional properties Wayne and Dwayne do not have in common. This is because 'dispositional properties may themselves be defined in part "externally," i.e., with reference to the organism's past or its environment' (2002, p.267). Schwitzgebel gives as examples Wayne's disposition to regard a present instance of water as an instance of the same sort of stuff Wayne drank as a child, and Dwayne's disposition to use the word 'water' with the intention of referring to the same stuff people in his Twin-Earth linguistic community refer to by using that word. He further notes that if one accepts *semantic* externalism (that is, externalism about the meaning of utterances as opposed to the content of mental states), there are other externally-individuated dispositions Wayne and Dwayne do not share: dispositions to utter sentences with certain meanings. Wayne isn't disposed to utter any sentences about twin water, and Dwayne isn't disposed to utter any sentences about water.

An objection that suggests itself here is that the role of indexicality in these examples is problematic. Dwayne does not, indeed, share Wayne's disposition to regard a present instance of water as an instance of the same sort of stuff Wayne drank as a child; but he does have the disposition to regard a present instance of water as an instance of the same sort of stuff *he*, Dwayne, drank

as a child, for if presented with a glass of $H_2O$, he would mistake it for a glass of XYZ. And surely (our imagined objector might insist) that means he has the 'same' disposition as Wayne *in the relevant sense;* each of them has the disposition to regard a present instance of water as an instance of the same sort of stuff he drank as a child.

Our objector might press the point as follows. Suppose Wayne and Dwayne are each disposed to tie their shoelaces before leaving the house. Following Schwitzgebel's example, we could identify two distinct dispositional properties of Wayne and Dwayne here: Wayne has the disposition to tie the shoelaces in Wayne's shoes (or: the same shoelaces Wayne tied last week), and Dwayne has the disposition to tie the shoelaces in Dwayne's shoes (or: the same shoelaces Dwayne tied last week). But surely it would be mere sophistry to suggest that this could have any bearing on such questions as whether Dwayne shares Wayne's belief that tying his shoelaces before leaving the house means he is less likely to trip on his front steps. When we ask such questions, it's simply beside the point that Wayne's shoes are *Wayne's* shoes and Dwayne's front steps are *Dwayne's* front steps. *In the sense that interests us*, Dwayne *does* share Wayne's belief.

Schwitzgebel's second example, our objector might suggest, involves a similar piece of sophistry or sleight-of-hand. Once again, we could indeed identify two distinct dispositional properties of Wayne and Dwayne here: Dwayne has the disposition to use the word 'water' with the intention of referring to the same stuff people in his linguistic community refer to by using that word, while Wayne has the disposition to use the word 'water' with the intention of referring to the same stuff people in *his* linguistic community refer to by using that word. But *for the purposes of belief ascription*, surely these ought not to be regarded as different dispositions any more than Wayne's disposition to tie Wayne's shoelaces and Dwayne's disposition to tie Dwayne's shoelaces ought to be regarded as different dispositions.

Finally, our objector might have this to say in response to Schwitzgebel's point about semantic externalism: suppose that on Earth, 'water is potable' means that water is potable, while on Twin Earth, 'water is potable' means that twin water is potable. And suppose that Wayne and Dwayne are each disposed, in certain circumstances, to say 'water is potable'. Is it true to say that Wayne is not disposed to utter a sentence that means that twin water is potable, and that Dwayne is not disposed to utter a sentence that means that water is potable? No, it is not. Each of them is in fact disposed to utter a sentence that means, on Earth, that water is potable and, on Twin Earth, that twin water is potable.

Perhaps there is a Wacky Earth where 'water is potable' means that water is poisonous, in which case Wayne and Dwayne are also disposed to utter a sentence that means (somewhere in the universe) that water is poisonous. It's only if one smuggles in an indexical element - supposing that 'utter a sentence with a certain meaning' means 'utter a sentence with a certain meaning *in one's own linguistic community*' - that one is entitled to describe Wayne and Dwayne as being disposed to utter sentences with different meanings.

I raise these objections because I think Schwitzgebel's argument on this point has an air of 'cheating' about it. It feels intuitively *wrong* in some way, and my imagined objector has had the task of putting his finger on *why* it feels wrong. The intuition driving his objections, I think, is this: if we're trying to specify what dispositions someone has - how she is disposed to act, think and feel in a range of circumstances - we ought to respect her point of view in a certain way. If the environmental stimuli to which Wayne and Dwayne are exposed, and their behavioural, cognitive and phenomenal responses to those stimuli, are *subjectively* indistinguishable, then we ought to say they have the same dispositions. We should not specify their dispositions in terms of the God's-eye-view of a thought-experimenter who knows the environmental causes of their subjective experiences to be different in some way. Hence if there *are* mental states the content of which must be specified in external, God's-eye-view terms, a dispositionalist account of these states is bound to fail.

What I think we must bear in mind here is that - as we saw in section 5.2 - phenomenal dispositionalism is supposed to be pragmatic in a certain way. In Schwitzgebel's view (and in mine), it is senseless to look for a binary yes-or-no answer to the question of whether a subject has a belief that P. What a subject has is a dispositional profile, and it may be that in certain contexts, ascribing the belief that P to that subject is a useful way to build up a picture of that dispositional profile.

Consider again the suggestion that Dwayne, right now, sitting on his sofa on Twin Earth, believes that water is potable. Ought we to ascribe this belief to Dwayne? One way to approach this question is to ask how his dispositional profile differs from that of Wayne (who certainly *does* believe that water is potable). If we allow ourselves to define dispositional properties partly in 'external' terms, it will certainly differ in some respects: for instance, as we have just seen, Dwayne does not have the disposition to regard a present instance of water as an instance of the same sort of stuff Wayne drank as a child, or as an instance of the sort of stuff referred to in his linguistic community as 'water'. However, there will be relevant dispositions Wayne and Dwayne *do* have

in common: for instance, just as Wayne would, if thirsty, drink a glass of water that was presented to him, so would Dwayne.

So, which similarities and differences between Dwayne's dispositional profile and Wayne's dispositional profile are the ones that *matter* for the purposes of deciding whether or not to ascribe to Dwayne the belief that water is potable? It depends, as we might expect, on the context of ascription. For as long as Dwayne is minding his own business on Twin Earth, never having encountered water and destined never to do so, it would serve no predictive or explanatory purpose to ascribe to him the belief that water is potable. We should therefore deny that he has that belief; his dispositional profile is not that of someone who believes that water is potable *in the respects that are relevant in that context*. However, we can imagine circumstances in which it would make good sense to ascribe to Dwayne the belief that water is potable. If Dwayne were to be tele-ported to Earth, for instance, we would struggle to predict and explain much of his behaviour (and many of his thoughts and feelings) if we were not prepared to ascribe to him various beliefs and desires concerning water. (I think this view is compatible with the common-sense view that in those circumstances, Dwayne's beliefs and desires concerning water would be explained by his mistaken belief that water was twin water. Dwayne would believe that water was potable, for instance, because he believed that twin water was potable and that water was twin water.)

In concluding this section, I want to consider how all this relates to non-mental dispositional properties such as solubility. Do we have a consistent picture here, or do mental dispositional properties require special treatment?

Consider a Twin Earth sugarcube. In terms of its intrinsic properties, it is identical to an Earth sugarcube. It would dissolve both if placed in twin water and if placed in water; so it has the dispositional properties *being water-soluble* and *being twin water-soluble*. It has these properties partly in virtue of the way it is intrinsically, partly in virtue of the way twin water and water are intrinsically, and partly in virtue of the way things are in Earth and Twin Earth environments generally.

Still, ascribing water-solubility to a sugarcube on Twin Earth is a useless thing to do in terms of predicting and explaining its behaviour. It's only if such a sugarcube were to be transported to Earth (or some water were to be transported to Twin Earth) that the ascription of water-solubility would serve any purpose.

Now consider Dwayne once again. In terms of his intrinsic properties, he is identical to Wayne (if we ignore the difference between the water in Wayne's body and the twin water in Dwayne's). He would (e.g.) drink a glass of wa-

ter if presented with one while thirsty, and also drink a glass of twin water if presented with one while thirsty. He therefore has the dispositional properties *being disposed to drink water when thirsty* and *being disposed to drink twin water when thirsty*. He has these properties partly in virtue of the way he is intrinsically, partly in virtue of the way twin water and water are intrinsically, and partly in virtue of the way things are in Earth and Twin Earth environments generally.

Still, ascribing the disposition to drink water when thirsty to Dwayne. an inhabitant of Twin Earth, is a useless thing to do in terms of predicting and explaining his behaviour. It's only if he were to be transported to Earth (or some water were to be transported to Twin Earth) that the ascription of that disposition would serve any purpose.

Fundamentally, then, we have here a consistent account of mental and non-mental dispositions. It's only when we consider the complexities of dispositional stereotypes and the practical purposes of mental state ascription that we become tempted to treat subjects differently from (mere) objects.

One final thought in response to the fundamental externalist intuition that it is flatly illegitimate to ascribe to subjects mental states that are 'about' things they have not had the right causal interactions with (so that, for instance, Davidson's 'swampman' (1987) - a randomly-generated physical duplicate of Davidson who acts just like him - cannot have beliefs about Davidson's friends since it has never met them). If we set sci-fi thought experiment scenarios aside, I think there are everyday cases we can point to in which it makes good sense to ascribe such states to subjects. Consider Bob, who (as we learned in Chapter 4) is afraid of spiders. Suppose he is also afraid of beetles, of centipedes, and so on. There is no 'natural kind' to which all and only such things as spiders, beetles and centipedes belong. Yet surely it make sense to ascribe to Bob a fear of creepy-crawlies in general - meaning he is afraid of a great many kinds of creature he has had no causal interactions with. (Perhaps he even has certain *beliefs* about creepy-crawlies in general - that they spread diseases, say.) If I notice a thumbnail-sized, hairy, many-legged creature crawling up Bob's arm, I don't need to know whether Bob has encountered a creature of the very same natural kind - species, genus, whatever - to know that Bob is afraid of such things; he's afraid of *all* things that are superficially like that one in certain respects. So why should we refrain from ascribing to Dwayne, upon his arrival on Earth, mental states that are 'about' water - a substance that is like twin water, a substance he encounters every day, in all respects but one? Similarly, why should we refrain from ascribing to him mental states that are 'about', say,

Wayne's mother? If the ascription of such states helps us make sense of other-wise inexplicable behaviour, thoughts and feelings on the part of Dwayne, we should go ahead and ascribe those states. (Caveat: the logic of expressions such as 'remembering that P' and 'recognizing person A' means that an exception must be made in these cases. Dwayne might *mistake* Dwayne's mother for his own mother and so *believe* that she plays the flute, or he might *realize* that Wayne's mother does everything his own mother does and so *infer* that she plays the flute, but he cannot *recognize* Wayne's mother and *remember* that she plays the flute.)

## 5.4 Conclusion

In this chapter, I have set out Schwitzgebel's version of phenomenal dispositionalism, based as it is around the notion of dispositional stereotypes for the possession of particular mental states - stereotypes to which subjects might conform in various respects and to various degrees, making the ascription to them of those mental states more or less appropriate in various contexts. I have also tried to shore up Schwitzgebel's response to the anticipated charge that phenomenal dispositionalism is incompatible with externalism about mental content.

As I noted at the beginning of section 5.2, Schwitzgebel is not concerned to say anything about the metaphysics of belief, or of dispositions in general. Much of the discussion in Chapters 3 and 4 was intended pre-emptively to answer some of the metaphysical questions left unanswered in his setting out of the phenomenal dispositionalist position. So let me now try to sketch a provisional picture of phenomenal dispositionalism in the round, metaphysics and all - a picture that learns the lessons of earlier chapters and applies them to the account just set out.

Provisionally, then - pending some revisions to the position, which I shall make in Chapters 6 and 7 - what phenomenal dispositionalism claims is this:

- To believe that P is to have some set of behavioural, phenomenal and cognitive dispositions such that one conforms, to an appropriate degree and in appropriate respects, to the dispositional stereotype for believing that P. (Roughly: it is to have dispositions to act, think and feel much as we would expect someone with the belief that P to act, think and feel in a range of circumstances.)

- These 'dispositions' one must have in order to be said to believe that P are not to be identified with their categorical bases. Dispositional properties are not categorical properties; nor are they role properties with corresponding categorical realizer properties; nor are they intrinsic properties of the subjects to which they are ascribed. They are relational properties; to have such-and-such a disposition is just to stand in such-and-such a relation to certain actual or possible objects and events.

- The holding of this relation in a given instance has counterfactual implications. For S to have the dispositional property *believing that P* is for certain counterfactuals about how S would act, think and feel in a range of circumstances to hold true. This relational, dispositional property of S supervenes on certain intrinsic, categorical properties of S, but also on intrinsic, categorical properties of other things.

- Nonetheless, dispositional mental states of subjects are both real and causally relevant. To explain that S took action A in circumstances C because she was in such-and-such a mental state is causally to explain why S took action A, just as surely as to explain that Lehmann Brothers collapsed in circumstances C because it was in such-and-such a state of financial health is causally to explain why Lehmann Brothers collapsed. The fact that no internal state of S is, or realizes, her mental state, and that no internal state of Lehmann Brothers is, or realizes, its state of financial health, is neither here nor there, so long as something, somewhere is doing the causal work required to make the right counterfactuals hold true.

Two significant problems remain. Firstly, Schwitzgebel's account of the role of dispositional stereotypes in belief ascription does not - for reasons I will set out in Chapter 6 - bear up under scrutiny. And secondly - as will become clear during that discussion - the dispositionalist still has questions to answer about how mental states interact with one another. I shall address this issue in Chapter 7.

# 6 From Dispositional Stereotypes to Dispositional Profiles

## 6.1 Introduction

We saw in Chapter 5 that Schwitzgebel's notion of a *dispositional stereotype* is central to his phenomenal dispositionalist account of belief and other attitudes. 'To believe that P...,' according to Schwitzgebel, 'is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P' (Schwitzgebel, 2002, p.253).

In this chapter, I will argue that while this claim cannot be accepted just as it stands, a similar but broader claim can be preserved in a new formulation of phenomenal dispositionalism; a version of the theory that takes the question of what mental states a subject has to be a question not of which dispositional stereotypes she conforms to, but simply of what dispositional profile she has. To be a subject in such-and-such an overall mental state - a subject with all of *these* beliefs, *these* desires, *these* character traits, and so on - is (I suggest) just to be a subject with all of *these* behavioural, phenomenal and cognitive dispositions (that is, a subject with *this* particular dispositional profile); but there is just no saying what it is to be a subject who has any given mental state considered in isolation.

In my view, one of Schwitzgebel's central claims - about just what a dispositional stereotype includes - invites two different readings. In sections 6.2 and 6.3, I will argue that whichever of these readings one accepts, the overall account cannot be made to hang together: on one reading, it's simply implausible that beliefs may only be ascribed to subjects on the basis of their conformity to a dispositional stereotype, while on the other reading, there appears to be no sense to Schwitzgebel's idea that deviations from the stereotype for a particular belief may be excused in certain conditions. In section 6.4, I will suggest that the phenomenal dispositionalist should move away from the idea that having each of the mental states one has is a matter of conforming to each of a

series of dispositional stereotypes, and towards a more holistic approach of the sort suggested in the previous paragraph.

## 6.2 'Aptness to associate': first reading and associated problems

Recall that Schwitzgebel, in resisting the idea that there is any determinate 'checklist' of dispositions a subject must have in order to be said to believe that P, characterizes the dispositional stereotype for believing that P as a cluster of dispositional properties that we *are apt to* associate with the possession of that belief.

One could take that to imply that we ought to be 'ultra-liberal' in accepting dispositional properties as belonging to that stereotype. Consider again the belief that there is beer in the fridge. Certainly I am apt to associate with that belief the disposition to offer beer to a thirsty guest, in this sense: if I take a subject, S, be at least moderately generous, to believe that his guest is thirsty and that his route to the fridge is clear, and so on, I am apt to think that S will offer his guest a beer if he believes that there is beer in the fridge. Similarly, however, I am apt to associate with that belief the disposition to conceal from a thirsty guest the fact that there is beer in the fridge: if I take S to be a miserly individual, or his guest to be a recovering alcoholic, I am apt to think that S will conceal the fact that there is beer in fridge if he believes that there is beer in the fridge.

This exercise could be continued *ad infinitum*. I am, for instance, apt to associate with the belief that there is beer in the fridge the disposition to attempt to stand on one's head and sing *My Way*: if I took S to believe that a million-dollar prize was available to anyone who both had beer in the fridge and was able to stand on his head and sing *My Way*, I am apt to think that S will attempt to stand on his head and sing *My Way* if he believes that there is beer in the fridge.

As we shall see in section 6.3, this reading may in fact be close to what Schwitzgebel has in mind. However, I want to consider an alternative reading first, because I think it sits more comfortably alongside Schwitzgebel's comments about deviations from stereotypes and the ways in which such deviations may be excused in certain conditions. As such, it is arguably a more natural reading and a more promising way to make sense of Schwitzgebel's overall position.

This alternative reading takes its cue from the fact that Schwitzgebel wants to say that someone who offers beer to a thirsty guest conforms in that respect to the dispositional stereotype for believing that there is beer in the fridge, while someone who conceals from a thirsty guest the fact that there is beer in the fridge deviates in that respect from that dispositional stereotype (although that deviation may be excused). If we take the sort of 'ultra-liberal' approach just sketched, that claim looks false; on that view, after all, both the disposition to offer beer to a thirsty guest, and the disposition to conceal from a thirsty guest the fact that there is beer in the fridge, belong to the dispositional stereotype for believing that there is beer in the fridge (because under certain conditions, we are apt to associate *both* of those dispositions with that belief). In fact, then, it seems natural to conclude that Schwitzgebel conceives of a dispositional stereotype not as a cluster of properties that we are apt to associate with a given belief *simpliciter*, but as a cluster of properties that we are apt to associate with a given belief *by default* - in the absence of any supplementary information about a subject's character, their other beliefs and desires, enviromental factors, and so on.

Note that in the case of the belief that there is beer in the fridge, this cluster of properties looks fairly 'well-rounded'. If we make some reasonable assumptions about what people are generally like - e.g. that they are basically rational, that they are more likely to tell the truth than to lie, that they are not so miserly as to withhold beer from a thirsty guest, etc. - we have a pretty good idea of how the typical believer-that-there-is-beer-in-the-fridge is going to act, think and feel in a range of circumstances.

But I think Schwitzgebel's choice of example tends to obscure the fact that there are many beliefs for which the range of 'default' assumptions about how someone with that belief is typically going to act, think and feel will be much more restricted. The problem, I think, is this: in all the respects that matter for the purposes of deciding whether or not it's appropriate to ascribe to someone the belief that there is beer in the fridge, pretty much everyone is going to act, think and feel pretty much the same way. So long as someone is not mute, pathologically miserly or inclined to lie, etc., it doesn't matter if they're male or female, under fifty or over fifty, politically left-leaning or right-leaning, religious or non-religious, somewhat dimmer than average or somewhat brighter, in the bottom half of the income distribution or in the top half, and so on; they are still going to be disposed to offer beer to a thirsty guest, etc.

By way of contrast, consider the belief that Jill is single. What does the dispositional stereotype for this belief look like? What dispositions are we apt to

associate, by default, with that belief? Well, there's the disposition to assert, when prompted, that Jill is single; the disposition to feel surprise upon learning that Jill is in fact married; and so on. No doubt we could go on expanding this list indefinitely, with a little imagination. Still, it seems to me that the dispositional stereotype for that belief is doomed always to look rather *thin*, in a way that the dispositional stereotype for the belief that there is beer in the fridge is not. If we ask ourselves how a subject, S, who believed that Jill was single, would typically act if Jill was coming round for dinner, how they would typically feel if they became aware that Jill's hand was resting on their thigh, or what they would typically infer if Jill mentioned how much she had enjoyed their company, answer comes there none. It all depends, we want to say. How old is Jill? Twenty-five? Seventy-five? How old is S? Is S a man or a woman? Is S married? Without at least a little such background information, there's just no saying how S would 'typically' act, think and feel in a whole range of circumstances.

To approach this same point from a different angle: suppose that just as 'Joe believes that there is beer in the fridge' means something like '*ceteris paribus*, Joe is disposed to assent to the proposition that there is beer in the fridge, to offer beer to a thirsty guest, to look in the fridge if he wants a beer, etc.', so 'this vase is fragile' means something like '*ceteris paribus*, this vase is disposed to break if dropped, if thrown against a wall, if struck with a heavy object, etc.'. Now suppose that on some particular occasion, a fragile vase is dropped but fails to break because there happens to be a large cushion on the floor just at the point of impact. On Schwitzgebel's account, as I understand it, the presence of the cushion excuses the non-manifestation of the vase's disposition to break if dropped. In just the same way, according to Schwitzgebel, we can see the presence of a desire to keep all one's beer to oneself as excusing the non-manifestation of Joe's disposition to offer beer to a thirsty guest. Just as the vase would have broken had the cushion not been there, so Joe would have offered his thirsty guest a beer had that desire not been there.

I think Schwitzgebel overlooks an important disanalogy here. If a vase is dropped but does not break - if, instead, it does something else, such as coming to rest or bouncing away or turning into a mermaid - one would never seek to explain what had happened, or failed to happen, by appealing to the vase's fragility. Its disposition to come to rest upon landing on a cushion, to bounce upon landing on a trampoline, or to turn into a mermaid upon landing in a magic lake might be relevant, but its fragility certainly would not be. Its behaviour in those cases, after all, is no more consistent with its being fragile than with its

being robust; so nothing about its behaviour on that occasion would justify the ascription of fragility to that vase. But if a thirsty friend visits Joe and Joe fails to offer him a beer - if, instead, he does something else - one might nonetheless seek to explain Joe's actions, thoughts and feelings by appealing to his belief that there is beer in the fridge. Why does Joe tell his friend he's out of beer, reassure himself that the risk of his lie being discovered is small, and feel anxious when his friend walks towards the fridge as if to open it? Precisely because, we might think, he believes that there is beer in the fridge and desires to keep it all to himself. His actions, thoughts and feelings are more consistent with his having the belief that there is beer in the fridge than with his lacking that belief; and so they justify the ascription to him of that belief.

It seems rather odd, then, to talk in terms of the non-manifestation of a disposition to offer beer to a thirsty guest being 'excused' by the presence of Joe's desire to keep all his beer to himself, and in terms of Joe 'deviating' from the way we would expect somebody with that belief to act, think and feel. If we know that Joe is miserly where beer is concerned, we don't regard his failure to offer beer to a thirsty guest as an anomaly that cries out for explanation. We aren't puzzled by a failure of Joe to fit into a recognizable, comprehensible pattern of actions, thoughts and feelings.

Consider the case of a twenty-five-year old single man who is introduced to a twenty-five-year-old woman, Jill (whom, we may suppose, he finds attractive). Suppose we are concerned to decide whether or not the belief that Jill is single can be attributed to this man. We are satisfied that he has such dispositions as the following: the disposition to ask Jill if she would like to go out for dinner some time; the disposition to wonder if he is Jill's type; the disposition to get his hopes up when Jill returns a phone call. Intuitively, it seems to me, his possession of those dispositions counts in favour of his possessing the belief in question: his actions, thoughts and feelings are, in those respects, more consistent with his having that belief than with his lacking it. Yet so far as I can see, on our present reading of Schwitzgebel's account, this man's possession of these dispositions counts for nothing - because someone who possesses them conforms no more closely to the dispositional stereotype for that belief than someone who lacks them. Such dispositions, after all, cannot belong to that stereotype - which only includes those dispositions we would expect pretty much *any* person with that belief to have, and not dispositions that we would expect only people of a certain age or a certain gender, or having a certain 'type', to have.

One final example to drive the point home. Consider the belief that the winning

National Lottery numbers this week are 4, 7, 14, 34, 35 and 40. What sort of dispositions are typically associated with this belief, and so (on our present reading of Schwitzgebel's position) earn a place in the dispositional stereotype for that belief? The disposition to assert that those are the winning numbers, no doubt; but also, surely, the disposition to acquire the belief that once again, one has failed to win the lottery, to bin one's Lottery ticket, to feel slightly deflated, and so on. That's how people with that belief would typically act, think and feel. Roughly one person in fourteen million, though, will be disposed to act, think and feel quite differently: to experience a sense of elation, to phone their friends and tell them they've won the Lottery, to infer that they can afford to give up work, and so on. Doesn't it seem rather beside the point to note that the non-manifestation of a disposition to feel slightly deflated is excused, in the case of such an individual, by the presence of a belief that their own lottery numbers are 4, 7, 14, 34, 35 and 40? Doesn't it seem that they deviate from the dispositional stereotype for that belief in a profound, systematic, general way - and yet that we have perfectly good grounds for ascribing to them that belief? Ascribing that belief to them, indeed, is the only way to make sense of their actions, thoughts and feelings.

On this reading of Schwitzgebel's position, then, he faces two very serious problems. Firstly, in a great many cases it is just not possible to make the sort of assumptions we would need to make about the beliefs, desires, character traits etc. of a typical believer that P in order to have a useful dispositional stereotype available to us; there are too many pertinent respects in which too many people will differ too often. And secondly, it is not enough simply to allow that *lacking* some of the dispositions in the stereotype for believing that P does not always count *against* someone's having that belief; we also need to allow that sometimes, *having* dispositions *not* in the stereotype for believing that P counts *in favour* of someone's having that belief. We can't settle the question of whether or not somebody believes that P on the basis of how many dispositions they have in common with a typical believer-that-P, but only on the basis of whether, *given that they have such-and-such other (typical or atypical) beliefs, desires, character traits etc.,* they have the sort of dispositions we'd expect them to have if they also believed that P.

# 6.3 'Aptness to associate': second reading and associated problems

One response to the concerns expressed in section 6.2, a response suggested (in personal communication) by Schwitzgebel himself (2011), would be to adopt something like the 'ultra-liberalism' referred to at the beginning of that discussion. On *this* view, dispositions to act, think and feel in ways that are quite atypical of subjects who believe, say, that there is beer in the fridge, may nonetheless be seen as belonging to the dispositional stereotype for that belief, because the necessary conditionality is built in to those dispositions. For instance: the disposition to stand on one's head and sing *My Way, if one believes that a million-dollar prize is available to anyone who both has beer in the fridge and can stand on their head and sing* My Way, should be seen as being included in the stereotype.

Let me note a couple of consequences of adopting this view. Firstly, it means that the dispositional stereotype for having a given belief very plausibly includes dispositions to act, think and feel any way that's nomologically possible in a given set of external circumstances, since there's no limit to the variety of 'if' clauses we could generate involving more or less eccentric sets of desires, beliefs, character traits etc. that a subject might, in principle, have. There'd always be *some* story we could tell about the eccentric beliefs, desires, ways of thinking etc. that a subject could have alongside the belief that P, and that would mean we were apt to expect that subject, in such-and-such external circumstances, to act, think and feel thus-and-so. The story about the man who believes he can win a million dollars by standing on his head and singing *My Way* is just one example.

Secondly, it makes it very hard to see just what use we could have for the idea that there are conditions in which deviations from a stereotype might be excused. Recall that on this reading of Schwitzgebel's position, it appears to be the case that both a subject who offers a beer to a thirsty guest, and a subject who conceals from a thirsty guest the fact that he has beer, conform equally well to the stereotype for believing that there is beer in the fridge - *if* the former is moderately generous, and/or feels morally obliged to offer his guest a beer, or whatever, while the latter is somewhat miserly, and/or believes his guest to be a recovering alcoholic, or whatever. The failure of the former to manifest a disposition to conceal from his guest the fact that he has beer, and of the latter to manifest a disposition to offer his guest a beer, could (I suppose) both be seen as being 'excused' by their possession of certain beliefs and character

traits. But the things that are here being excused cannot be these subjects' *deviations* from the relevant stereotype, for on this reading of Schwitzgebel's position, neither of the subjects we've been considering *does* deviate from that stereotype: each of them acts, thinks and feels much as we would expect them to, given the other beliefs and desires they have. So far as I can see, on this reading of Schwitzgebel's position, a subject could only deviate from the stereotype for believing that P if he failed to act, think and feel as we would expect him to act, think and feel if he believed that P *and had all the other beliefs, desires etc. he does in fact ha*ve. But in that case there would be no question of his deviation being 'excused' by his other beliefs, desires etc.

Nor, I think, would there be any question of a subject's deviations being excused by external conditions. Someone who fails to offer a beer to a thirsty guest because his route to the fridge is blocked is, after all, surely acting as we would expect someone who believed that there was beer in the fridge, was moderately generous, etc., to act *if his route to the fridge was blocked*. So if such conditionals are built in to the dispositional stereotype for believing that P, there seems to be no reason to see that subject as deviating from that stereotype.

Essentially, it appears to be impossible, on this reading of Schwitzgebel's position, for someone who acts, thinks and feels as we would expect him to if he believed that P, if he had the other mental states he does in fact have, and if external conditions were as they in fact are, to deviate from the stereotype for believing that P. That seems fair enough, in a sense - it fits the overall picture of believing that P being a matter of having the dispositions we are apt to think someone ought to have if they believe that P[1] - but it seems to leave the idea of 'excusing conditions' without any work to do. On this ultra-liberal view, people who deviate from the stereotype for believing that P can, I think, straightforwardly be seen as failing to believe that P.

In another response (in personal communication) to the concerns raised in section 2, Schwitzgebel (2011) suggests that the picture we work with should be one of 'mini-stereotypes' for different types of subject. The idea here is that the dispositional stereotype for believing that P will include not just a general

---

[1]Something like Davidson's 'principle of charity' (1973) is at work here. Since we are in the business of making sense of people, we are hardly likely to refrain from ascribing the belief that P to someone whose actions, thoughts and feelings would strike us as being more rational and more coherent if we *did* ascribe that belief to them. Schwitzgebel is, of course, aware of the need to accommodate this sort of consideration, but seems to be torn between doing so by conditionalizing the dispositions within a stereotype (enabling us to regard more people as conforming to that stereotype), and doing so by allowing for 'excusing conditions' (enabling us to regard people who do not conform to a stereotype as having the belief associated with that stereotype nonetheless).

pool of dispositions we are apt to think *all* believers-that-P will have, but also clusters of dispositions that are specific to particular types of subject: in the case of the belief that Jill is single, for instance, we might be on the lookout for clusters of 'pursuer dispositions,' 'avoider dispositions,' 'envious woman dispositions,' 'flirter dispositions,' etc.

In pragmatic terms - in terms of how we might actually go about deciding whether such-and-such a subject has or lacks such-and-such a mental state - this suggestion doubtless has some merit. But at this point, I think it must be acknowledged that we are straying further and further from the idea that there is any such thing as a unique dispositional stereotype for believing that P, to which anyone who believes that P ought to conform unless their deviations from that stereotype are appropriately excused by the presence of certain unusual conditions (e.g. their physical limitations or possession of an odd belief, or some peculiarity of their environment). Instead we seem to be moving towards a picture on which there are a multitude of stereotypes for believing that P: one for believing that P and also having *this* set of other beliefs, desires, character traits, etc., one for believing that P and also having *that* set of other beliefs, desires, character traits, etc., and so on. And although it may be very useful to sort people into general 'types', what matters in ascribing beliefs to a subject, S, is ultimately not whether S acts, thinks and feels as a subject typical of some general type ought to act, think and feel if they believed that P, but whether S acts, thinks and feels just as *she herself* ought to act, think and feel if she believed that P and had all the other beliefs, desires, character traits etc. she actually has - with these being as specific, individual, and even eccentric as you please.

# 6.4 Dispositional profiles

In the course of the discussion in sections 6.2 and 6.3, the idea that a subject's possession of some particular mental state is a matter of his conforming (in appropriate respects, and to an appropriate degree) to a dispositional stereotype for the possession of *just that state* - an idea that is at the heart of Schwitzgebel's formulation of phenomenal dispositionalism - has come to appear less and less tenable. We have found ourselves edging closer and closer to the view that the dispositional stereotype a given subject must conform to in order to be said to have such-and-such a belief (say) is not a stereotype for the possession of that belief itself, but rather a stereotype for the possession of that belief plus whatever other beliefs, desires, character traits and so on we

take that particular subject to have. And this, surely, is no 'stereotype' at all; the concept has been stretched past breaking point.

So this talk of dispositional stereotypes is going to have to go, and with it the whole idea that mental state ascription can proceed on a one-at-a-time basis. As Stout argues (see Chapter 2), a successful dispositionalism must take the holism of the mental seriously. We must accept that there is no one general way of acting, thinking and feeling that is characteristic of having a given belief; rather, there are many particular ways of acting, thinking and feeling that are characteristic of having that belief together with other possible sets of beliefs, desires, character traits and so on.

In my view, then, phenomenal dispositionalism should be formulated not in terms of dispositional stereotypes, but in terms of dispositional profiles. A subject's dispositional profile, recall, is just his overall set of behavioural, phenomenal and cognitive dispositions. (It therefore excludes those of his dispositions which are not behavioural, phenomenal or cognitive: the dispositions to drown if submerged in water or to dissolve if submerged in acid, for instance. Such dispositions are real enough, but to seek to explain or predict events by appealing to facts about such dispositions is not to seek to explain or predict events by appealing to facts about subjects' mental states.) On Schwitzgebel's account, every subject *has* such a profile - of course - but these profiles are not really something we ever look at in the round. Since mental state ascription proceeds in a one-at-a-time fashion, based on the respects in which, and the degree to which, a subject is judged to conform to each of a series of stereotypes for discrete mental states, we are only ever interested in looking at certain aspects of a subject's dispositional profile at any one time. This makes sense in *pragmatic* terms, of course, since we're just not in a position to know everything there is to know about the dispositional profile of every subject we encounter; we have no choice but to make assumptions about what sort of person we're dealing with, what knowledge they're likely to have or lack, etc., in order provisionally to ascribe large sets of mental states to them (a process to which I will return in Chapter 7.) But it does not make sense in terms of the 'shape' of a theory about *what it is* to have mental states. A theory with that shape just is not able properly to accommodate the interdependence of mental states - the holism of the mental.

A version of phenomenal dispositionalism with dispositional profiles at its centre, however, has a quite different shape. Rather than asking whether a subject has the right dispositions for someone who believes that P - say - it invites us to ask whether a subject's dispositional profile is, all things considered, that

of someone who has that belief along with these other beliefs, these desires, these character traits, and so on. Such a theory claims a constitutive link not between a subject's possession of particular mental states and her possession of particular dispositions, but between her possession of all the mental states she has and her possession of all the behavioural, phenomenal and cognitive dispositions she has.

## 6.5 Conclusion

If phenomenal dispositionalism is to accommodate psychological holism, I think it will have to move decisively away from the idea that questions about what mental states someone has can be settled on a one-by-one basis by asking whether that person conforms to each of a series of stereotypes. What we are trying to do when we ascribe a mental state to a subject S is, in my view, to build up a picture of that person's dispositional profile. We may not be explicit about the other mental states we take them to have, but tacitly we make various more or less justified assumptions - either very general assumptions about the beliefs, desires, habits of thought etc. that pretty much everyone has in common, somewhat general assumptions about the beliefs, desires etc. that subjects of a general type to which S belongs have in common, or more specific assumptions based on what we know about S as an individual. Ultimately, to be a person with such-and-such a set of mental states is just to be a person with such-and-such a dispositional profile; but to be a person with any *particular* mental state is not to have any *particular* dispositions. In Chapter 7, I will argue that when it comes to accommodating psychological holism, the phenomenal dispositionalist can learn from Dennett's (2002) version of interpretationism.

# 7 Dispositional Profiles and the Intentional Stance

## 7.1 Introduction

In Chapter 6, I argued that a subject's having some particular belief (or desire, or character trait, or whatever) cannot be - as Schwitzgebel suggests - a matter of that subject's conforming to a *dispositional stereotype* for the possession of that particular belief. We should, I suggested, take a more holistic view, according to which a subject's having a certain overall set of mental states is a matter of his having a certain *dispositional profile*.

In thus seeking to accommodate the holism of the mental, phenomenal dispositionalism takes a step closer to interpretationism - already, in my view, its closest relative among contemporary positions in the philosophy of mind (especially in the avowedly Rylean form defended by Daniel Dennett). When we adopt what Dennett calls (2002) the 'intentional stance' towards a subject, we ascribe a whole system of beliefs and desires to that subject and seek to predict and explain their behaviour with reference to that whole system. There is no attempt to ascribe mental states to a subject one at a time, either on the basis that the subject has some determinate set of dispositions the possession of which constitutes his having that mental state (as per logical behaviourism), or on the basis that the subject more or less conforms to a dispositional stereotype for the possession of that belief (as per Schwitzgebel's version of phenomenal dispositionalism).

There are, however, important differences between interpretationism and phenomenal dispositionalism. In this chapter I want to spell out what the phenomenal dispositionalist can take from Dennett's interpretationism and what he must reject. I will argue that the interpretationist tendency to boil mental life down to practical reasoning based on beliefs and desires serves to obscure the important role of ascriptions of other mental states - including non-intentional states - in predicting and explaining subjects' behaviour.

## 7.2 The intentional stance

Dennett (2002) describes the intentional stance as an alternative to two other stances we can adopt towards systems when we seek to predict and explain their behaviour: the 'physical stance', and the 'design stance'. To seek to predict a system's behaviour from the physical stance is to 'determine its physical constitution... and the physical nature of the impingements upon it, and use your knowledge of the laws of physics to predict the outcome for any input' (Dennett, 2002, p.557). However, even if it's possible in principle to predict the behaviour of any system in any circumstance using this strategy, it's not possible in practice most of the time. In our day-to-day interactions with things, we just aren't in a position to acquire and apply all the knowledge we'd need to have about their physical constitution, about the physical nature of the impingements upon them, and about the laws of physics. We can't pause to dismantle the ticket machine at the train station on our way to work, examine all its parts in microsopic detail, and map out the detailed causal chain that would be initiated if one were to press the button marked 'Print ticket'. We would, instead, seek to predict that system's behaviour from the design stance, 'where one ignores the... details of the physical constitution of an object, and, on the assumption that it has a certain design, predicts that it will behave *as it is designed to behave* under various circumstances' (2002, pp.557-558, Dennett's italics). On that basis, we would predict that the ticket machine, having been designed to print a ticket when someone presses the "Print ticket" button, would, in those circumstances, do just that. The strategy can also work, Dennett suggests, for many biological systems - 'plants and animals, kidneys and hearts, stamens and pistils' (2002, p.558). If we assume that animals are 'designed' to preserve their own lives, to reproduce, etc., and that hearts are 'designed' to pump blood around the body as needed, we will very often be able correctly to predict what they'll do in various circumstances without having any deep understanding of their physical constitution.

The possibility of taking the *intentional* stance provides us with a third strategy for predicting systems' behaviour. This is how Dennett suggests the strategy works:

> first you decide to treat the object whose behavior is to be predicted
> as a rational agent; then you figure out what beliefs that agent ought
> to have, given its place in the world and its purpose. Then you figure
> out what desires it ought to have, on the same considerations, and
> finally you predict that this rational agent will act to further its goals

in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many - but not all - instances yield a decision about what the agent ought to do; that is what you predict the agent *will* do.

(Dennett, 2002, p.558)

Our starting point in deciding what beliefs and desires an agent ought to have - if that agent is a human being, anyway - is that people generally acquire true beliefs about things they have suitable experience of, and people generally have certain basic desires (for food, pleasure etc.) in common. So if, for instance, we encounter someone who hasn't eaten for a week and who has just been presented with a tuna sandwich just like the ones he has eaten a hundred times before, we will think he ought to believe that the sandwich is edible and desire that his hunger be satisfied, and so predict that he will eat the sandwich. We also start out by assuming that people are perfectly rational (Dennett, 2002, p.560). All this is only a starting point; we can accommodate the fact that people sometimes have false beliefs, or fail to see the implications of their beliefs, if there's a suitable explanation. But we have to assume that 'true believers mainly believe truths' (Dennett, 2002, p.559), and are in the main rational, or the whole strategy collapses (cf. Davidson's 'principle of charity' or 'principle of rational accommodation' (1973)).

So far, this could all be taken as a sort of 'how-to' guide on the practical application of a Standard View theory of mind. No Standard View theorist would want to deny, after all, that we use something very like Dennett's intentional strategy when we ascribe mental states to people in order to predict and explain their behaviour. But there are two features of Dennett's account that set his interpretationist view apart from Standard View theories like functionalism and representationalism.

Firstly, Dennett asserts that '*What it is* to be a true believer is to be an *intentional system*, a system whose behaviour is reliably and voluminously predictable via the intentional strategy' (2002, p.557, Dennett's italics). Being interpretable as a system with propositional attitudes is thus held by Dennett to be *constitutive* of having those attitudes. This is quite distinct from the claim that being a true believer is a matter of one's brain being in such-and-such a functional state, or containing such-and-such representations, even though it might, as a matter of fact, follow from such claims that true believers are amenable to interpretation via the intentional strategy.

Secondly (and relatedly), Dennett sounds a note of caution over an influential line of thinking that can be considered as a response to the question:

why does the intentional strategy work? (this being understood as a question about 'how the machinery which Nature has provided us works' (Dennett, 2002, p.566)). The line of thinking, which starts out generically functionalist and ends up recognisably representationalist, runs like this:

> the account of how the strategy works and the account of how the mechanism works will (roughly) *coincide*: for each predictively attributable belief, there will be a functionally salient internal state of the machinery, decomposable into functional parts in just about the same way the sentence expressing the belief is decomposable into parts - that is, words or terms. The inferences we attribute to rational creatures will be mirrored by physical, causal processes in the hardware; the *logical* form of the propositions believed will be copied in the *structural* form of the states in correspondence with them. This is the hypothesis that there is a *language of thought* coded in our brains, and our brains will eventually be understood as symbol manipulating systems in at least rough analogy with computers.
>
> (Dennett, 2002, p.566, Dennett's italics).

Dennett thinks some version of this view will probably turn out to be right (2002, p.566), but cautions that it only appears *obviously* or *inevitably* correct if one confuses two distinct empirical claims: the claim that 'intentional stance description yields an objective, real pattern in the world,' and the claim that 'this real pattern is *produced by* another real pattern roughly isomorphic to it within the brains of intelligent creatures' (2002, p.566, Dennett's italics). The first claim is plainly true, but its truth is independent of the truth of the second claim - which has yet to be established.

## 7.3 Beyond the intentional stance

It should be clear from the above that Dennettian interpretationism and phenomenal dispositionalism share a good deal of common ground. Both are broadly Rylean positions that deny a constitutive link between having such-and-such a propositional attitude and being in such-and-such an internal, physical state, and instead think of having such-and-such a propositional attitude as being more a matter of responding to situations in a way that fits a recognizable, interpretable pattern.

The crucial divergence between the two positions is over the question of what

*types* of responses are relevant to the question of what propositional attitudes a subject has. The interpretationist admits only the behavioural responses observable by an interpreter; in this, he is very much the heir to that semi-mythical figure, Ryle the Behaviourist. The phenomenal dispositionalist admits cognitive and phenomenal responses in addition to behavioural responses, viewing these as crucial to the characterization of the patterns into which interpreters attempt to fit subjects. There could, on a phenomenal dispositionalist view, be cases in which there were no good behavioural grounds to justify interpreting someone as having the belief that P, but in which her internal utterances (say), or the occurrence 'in her head' of certain inferential chains, nonetheless meant that her dispositional profile was, all things considered, that of someone who believed that P. (Of course, interpreters are not in a position directly to learn about subjects' cognitive and phenomenal responses, as they are in a position directly to learn about subjects' behavioural responses; but that they nonetheless interpret subjects as having such responses, predict and explain their behaviour on the basis of assumptions about these responses, and regard having such responses as being part-and-parcel of having mental states, could hardly be clearer (in my view) from a consideration of the use made by interpreters of mentalistic language.)

Interpretationism also has a narrower focus than phenomenal dispositionalism. The intentional strategy is a way of predicting and explaining people's behaviour by ascribing intentional states to them - propositional attitudes that are *about* something. As such, it reflects a distinctly rationalistic understanding of human thought and action: people are to be understood as systems that have beliefs about the way things are, desires about the way things should be, and some natural machinery or other that equips them to make rational decisions about how to pursue those desires in light of those beliefs. Phenomenal dispositionalism, on the other hand, seeks to offer an account not just of the *intentional* mental states, or propositional attitudes, that are the stuff of practical reasoning, but also of many *non*-intentional features of mental life. Character traits are a prime example. We regularly seek to predict and explain people's behaviour, and to figure out what's going on 'inside their heads', by appealing to facts about their beliefs and desires alongside facts about their possession of character traits such as being quick-witted, short-tempered, or lacking in empathy. Jack might very well desire a chocolate bar and believe that the way to get one is to queue at the counter, but if we know him to be terribly impatient we might be inclined to predict that he'll drop out of the queue before he gets as far as buying anything. Adopting the intentional stance towards Jack - treating him as a rational agent with beliefs and desires about the world and the

capacity to work out how to pursue his goals in light of his beliefs - is central to that attempt at understanding and interpreting him, but it is by no means the whole story. His impatience is not a belief or a desire; it is not *about* anything at all.

Perhaps it will be objected that in ascribing impatience to Jack in order to predict his behaviour, we are in fact doing nothing more than ascribing additional beliefs and desires relevant to his likely behaviour in that context: say, a desire not to waste time and a belief that to queue for longer than a minute would be to waste time. I think there are good reasons to reject this view, however. Firstly, it fails to capture the psychological reality that Jack's behaviour may not be the result of rationally weighing up the costs and benefits of queueing vs. not queueing; rather, it may be the impulsive result of a feeling of frustration that even Jack recognizes as standing in the way of his doing what he ought, rationally, to do. Secondly, it fails to fit Jack's behaviour on this occasion into a broader pattern of impatient behaviour. Why did Jack give up piano lessons after a few weeks, even though he is otherwise best interpreted as having a long-standing desire to play the piano and a sincere belief that nothing could be a more worthwhile use of one's time than nurturing one's musical abilities? No doubt we could dream up another *ad hoc* belief-desire explanation, but it would be highly unlikely to further our general understanding of Jack in the way that a simple ascription of a non-intentional mental state, *impatien*ce, does. To limit oneself, as an interpreter, to ascriptions of beliefs and desires - whether made explicitly in those terms or not - is to miss opportunities to flesh out one's understanding of subjects' dispositional profiles, and so to leave oneself blind to real patterns in the world of just the sort that a competent interpreter should be in the business of spotting.

As well as the paradigmatic intentional states we call 'propositional attitudes' (e.g. the belief that P), and states that we do not characterize with reference to any intentional, propositional content (e.g. the state of being impatient), we ascribe mental states to people that seem to have both intentional and non-intentional elements. If I seek to explain Jill's behaviour by telling you that she loves her cat, part of what I'm trying to get across to you is, very plausibly, that she desires that her cat should be healthy and content (or something along those lines). But I might also be trying to get across the fact that, say, Jill finds it upsetting to see her cat suffer, or finds it hard to be parted from her cat for any length of time, or enjoys spending time with her cat, or tends to show affection to her cat. What I'm doing is inviting you to make a whole family of assumptions about Jill's dispositional profile. I'm not simply asking you to assume that Jill

will act to further her desire for her cat's well-being in light of her beliefs about the prerequisites of feline flourishing.[1]

Even when it comes to propositional attitudes, we would do well to remember that the clear-cut beliefs that P and desires that Q invoked in philosophical accounts of practical reasoning are idealizations. When we seek to predict and explain one another's behaviour through the ascription of doxastic states and 'pro-attitudes', we invoke not just beliefs that P, but also hunches that P, mounting suspicions that P, uneasy feelings that P, and unshakeable convictions that P; not only desires that Q, but also deep-seated yearnings that Q, half-wishes that Q, and reluctant acknowledgements of the necessity that Q. This is not simply a matter of needless linguistic extravagance; talk conducted in these terms cannot be reduced, without loss of predictive and explanatory power, to talk conducted in terms of a manageable number of real, fundamental, canonical state types. The subtle distinctions made by ordinary interpreters, and the nuances of the expressions they use, are indispensable parts of the business of building up complex pictures of one another's dispositional profiles. Someone who feels in his gut that P really can be expected, in certain circumstances,

---

[1]Nor, of course, am I alerting you to the presence of an internal state of Jill playing a particular causal role. It's worth pausing briefly to reconsider this point. What sort of 'causal role' could we take Jill's love for her cat to play, and how could we individuate the internal state of Jill's brain playing that role? Presumably, the idea would be that there is some state of Jill's brain that occupies a certain position in the causal chains leading from certain sensory inputs to certain behavioural outputs. When Jill sees her cat limping and telephones the vet, what's happening is something like this: her perception of the limping cat activates her love for the cat, which activates her desire for her cat's well being, which looks around for relevant beliefs about what can be done to promote the well-being of limping cats, finds the beliefs that vets treat cats with problems such as limps and that vets can be contacted by telephone, and together with those beliefs causes Jill to telephone the vet. And when Jill's cat jumps up on to her lap, that activates her love for the cat, which activates a desire to show affection for the cat, which interacts with the beliefs that cats are animals and that stroking an animal is a way of showing it affection to produce cat-stroking behaviour.

On a phenomenal dispositionalist view, it's a matter of complete indifference whether or not cognitive science confirms the truth of such 'just so' stories by producing a taxonomy of brain states corresponding to the *dramatis personae* of those stories. If it turns out that there just isn't any indentifiable state of Jill's brain that occupies the causal role her love for her cat is supposed to occupy in both of those scenarios, that is no threat to the truth of the claim that Jill loves her cat - for her dispositional profile remains that of someone who loves their cat.

Dennett makes a similar point in his (1978), in which he notes that the behaviour of a chess-playing computer program might be successfully predicted and explained by ascribing to it the belief that it should get its queen out early, even though the program does not include any representation with the content 'I should get my queen out early' (p. 107). The point is made in response to representationalist theories in particular and so is made in terms of representational content, but I think it could equally well be made in terms of causal or functional role: the behaviour of a chess-playing computer program might be successfully predicted and explained by ascribing to it the belief that it should get its queen out early, even though the program does not include any line of code (or whatever) that does the job reserved for the belief that one should get one's queen out early in the sort of functionalist 'just so' stories just referred to.

to act, think and feel differently from someone who accepts, rationally, that P must be the case.

On the other hand, it would be a mistake to think that the richness of ordinary language induces commitment to the existence of a bewildering variety of mental state types crying out for investigation by cognitive psychologists or neurobiologists. No doubt there are salient neurological differences between someone whose dispositional profile is such as to justify the ascription to him of a gut feeling that P and someone whose dispositional profile is such as to justify the ascription to him of a rational acceptance that P, and these differences might well merit scientific investigation; but this does not mean that neuroscience must ultimately either identify tokens of *the gut feeling that P* and *the rational acceptance that P* in subjects' brains, or else conclude that there are no such phenomena as gut feelings and rational acceptances. The ordinary interpreter and the neuroscientist are in quite different lines of business and should not expect their different theories to map neatly on to one another; the former is trying to fit subjects into interpretable patterns of interaction with the world, while the latter is trying to work out what brain states and processes underpin the existence of those patterns.

## 7.4 Profile building

Hopefully the foregoing has given a sense of where I find Dennettian interpretationism helpful and where I find it wanting. Here, then, is a sketch of what I think is going on when we ascribe mental states to subjects.

Our starting point is, as Dennett suggests, to assume that subjects believe what they ought to believe and desire what they ought to desire, given their place in the world and their purpose. Crucially, however, there is much more going on here than the mere attribution of rationality to subjects. It is not merely *qua* rational being that someone acquires a belief that his mother has just walked into the room, but *qua* being with a certain array of sensory equipment and a certain range of tendencies and capacities to process and respond to certain sorts of sensory input in certain ways - e.g. to attend to certain details of one's environment while filtering out irrelevant 'noise', to recognize faces, to respond emotionally to the presence of loved ones.[2]Among the relevant

---

[2]It has been suggested (Ellis and Young, 1990) that the presence of such an emotional response is what makes the difference between a normal subject acquiring the belief that his mother has just walked into the room, and a sufferer of the Capgras delusion acquiring the belief that an imposter who looks just like his mother has just walked into the room.

tendencies and capacities are, of course, tendencies and capacities to make certain rational inferences. But at bottom, deciding what a subject 'ought' to believe and desire, for the purposes of predicting or explaining her behaviour, is a matter not of recognizing that reason demands she believe this or that, but of being familiar with the ways in which subjects typically think and feel and interact with their environment - of being familiar, that is, with the dispositional profiles of typical subjects.

We do our interpreting in the full knowledge that people's dispositional profiles are not those of perfectly rational beings. People are good at seeing what follows from what in some cases but not in others, so what inferences we as interpreters think they 'ought' to make will depend on our assessment of their capacities. In some contexts we will soon go wrong if we assume, for instance, that everyone has a solid grasp of probability.

Consider the famous 'Monty Hall Problem'. The problem goes like this: a contestant on a game show hosted by Monty Hall is asked to choose one of three doors. Behind one door, he is told, is a car; behind each of the other two doors is a goat. Once the choice has been made, round one of the game is over and Monty opens one of the remaining two doors to reveal a goat. (We are to suppose that Monty knows but will not reveal the location of the car; that if the contestant has chosen the door with the car behind it, Monty decides at random which of the other doors to open; and, of course, that the contestant grasps all this.) The contestant is then asked, in round two, if he would like to switch to the other remaining door, or stick with his original choice. If the contestant wants to maximize his chance of finding the car, should he switch or stick?

The answer is that the contestant should switch in order to double his chances of winning the car from one in three to two in three. And the reason for this is not, on the face of it, especially hard to grasp; it certainly doesn't require mastery of any advanced mathematics.[3] Assuming not much more than basic rationality, people really 'ought' to be able to see it for themselves. But as a matter of fact, most people - including many people *with* a mastery of advanced mathematics (vos Savant, no date) - can't. Most people think it makes no difference if the contestant switches or sticks, since his chances of winning must be fifty-fifty in either case: either the car is behind the door he has already

---

[3]Here's a simple, non-technical explanation: if the contestant chooses the door with a car behind it in round one, he'll win if and only if he sticks. But if he chooses a door with a goat behind it in round one, he'll win if and only if he switches. Since there are twice as many 'goat' doors as 'car' doors, he's twice as likely to have chosen a 'goat' door in round one and therefore to win by switching.

chosen, or else it's behind the other door - right?

So if we want to predict the behaviour of that contestant, it's no use crediting him with the beliefs he 'ought' to have if only he can, so to speak, put two and two together. Better to play the odds and assume he's one of the majority of people who will believe their chances are fifty-fifty either way, in spite of the fact that that belief is incompatible with other beliefs they hold (e.g. that the initial choice was twice as likely to have been a goat as a car).

And note that even if we make *that* assumption, we still risk being led astray if we reason that the contestant is as likely to switch as to stick (since he thinks he's equally likely to win in either case). In fact we need to bear in mind that the contestant is going to be asking himself: how would I feel if I stuck and then lost? How about if I *switched* and then lost? Would I feel worse if I 'had' the prize and let it slip through my fingers, than if I just never had it? We need to bear in mind that his thinking will be affected by the pressure he's under, by his emotional responses to imagined scenarios, by hunches he knows are irrational, and so on. My money would be on the contestant sticking with his original choice; but that is not a prediction easily arrived at simply by crediting the contestant with a desire to win a car, a (false) belief that his chance of doing so is one in two whether he switches or sticks, and a capacity to use practical reasoning to settle on a course of action. In making it, I draw on a host of intuited assumptions - available to me not because I have crunched the practical-reasoning numbers, but just because I know what people are like (and what being a person is like)[4] - about what will be running through the contestant's head.

A competent interpreter grasps the sort of distinctions mentioned in section 7.3 - between feeling in one's gut that P and accepting rationally that P, for instance. He understands that a subject who feels in her gut that horse riding is dangerous can be expected to act, think and feel differently from a subject who accepts (on the basis of statistical evidence, say) that horse riding is dangerous. (Just assuming that both of them believe horse riding is dangerous

---

[4]This is not intended to be an endorsement of simulationism (Goldman, 2006; Gordon, 1986); I don't mean to claim that my ability to put myself in other subjects' shoes (in some sense) provides my sole or primary route to successful explanation and prediction of their actions. Rather, I mean to claim that that ability serves to extend my understanding of subjects beyond the mere recognition 'from the outside' of behavioural patterns into which they can be fitted (plus inferences from their observed behaviour to their possession of beliefs and desires that make sense of it). In particular, it puts me in a much stronger position to make assumptions about subjects' phenomenal dispositions and about cognitive dispositions for which we may lack behavioural evidence. It is what enables me, in this case, to hypothesize that a contestant might choose to stick with their original door simply because her affective reaction to an imagined 'switch and lose' outcome is much more unpleasant than her affective reaction to an imagined 'stick and lose' outcome.

will not get him very far in predicting and explaining their behaviour.) And he is adept at sorting people into types; he understands that there are very general differences between the actions, thoughts and feelings of impatient, uninhibited, quick-witted, passionate people and those of serene, reserved, slightly dim and rather frosty ones. And no doubt he makes use of stereotypes, in the everyday sense of the word, too; by considering the accent, or age, or attire, or religious affiliation of the person he's interpreting, he can take some short-cuts to some provisional assumptions about their attitudes and character traits.

'Provisional' is a key word. Interpretation is a process of perpetual revision, as old assumptions get dropped and new ones get made in light of new information. In trying to predict and explain the behaviour of someone we know little or nothing about, we have little choice but to base our cautious predictions and explanations on the assumption that their dispositional profile is roughly that of a 'typical' person - or perhaps a 'typical' person of their sex, age, appearance, or whatever else we have to go on. But as we learn more about them - by paying attention to what they say and do in a range of situations - our assumptions about their dispositional profile are being refined all the time. We ascribe to them a new belief, desire, fear, hope, or character trait, and a ripple effect changes our whole picture of them in subtle or not-so-subtle ways. The holism of the mental is in this sense 'in the eye of the interpreter'. The right place to look for 'interactions' between mental states is, therefore, not in a causal system of brain states responsible for producing behaviour, or in an abstract practical-reasoning system for rationalizing or recommending courses of action; it is within the pictures of subjects' dispositional profiles we build up for ourselves, as interpreters, using the language of mental states.

## 7.5 Conclusion

In this chapter, I have endorsed the Dennettian interpretationist's insistence that the ascription of mental states to subjects must proceed holistically; making sense of a subject necessarily means ascribing whole sets of mental states to them at once. Someone might have or fail to have the dispositional profile of a subject with all of *these* beliefs, *these* desires, and so on, but (for the reasons set out in Chapter 6) there is no dispositional stereotype for the having of any particular belief or desire considered in isolation - *pace* Schwitzgebel - and so no question of a subject conforming or failing to conform to such a stereotype.

I have argued, however, that to adopt Dennett's 'intentional stance' towards subjects, by ascribing beliefs and desires to them and predicting that they will

generally act as practical reasoning dictates, is to frame the business of interpretation in unduly rationalistic terms. Interpretation must be grounded in matters of fact about the tendencies and capacities subjects actually have: to acquire or to fail to acquire certain beliefs about their environment, to make or to fail to make certain inferences, to act 'against their better judgement' under the influence of certain feelings, and so on. Interpreters who focus unduly on beliefs, desires and practical reasoning will leave themselves blind to the subtleties and complexities of individual subjects' dispositional profiles - subtleties and complexities that are intuitively grasped by interpreters who use everyday language to ascribe a richer range of intentional and non-intentional mental states to subjects.

# 8 Functionalism, Phenomenal Dispositionalism, and Theological Language

## 8.1 Introduction

In this chapter, I want to offer a 'case study' illustrating how a phenomenal dispositionalist account of mentalistic language might be applied to specialized questions in the philosophy of mind and related areas. The nature of the question being addressed also means we will be able to get an idea of what phenomenal dispositionalism has to say about the sort of atypical agents[1] that populate the philosophy of mind literature - Martians (Lewis, 1980), 'super-Spartans' (Putnam, 1963), 'Blockhead' systems (Block, 1981), and so on.

The question I want to address is this: how (if at all) is it possible to speak of God using terms derived from our everyday discourse about human persons? This question has long troubled theists. God, after all, is supposed to be wholly other than human persons; surely, then, it cannot be legitimate to conceive of him as having features in common with them?

In his (1989), William Alston proposes that it may indeed be legitimate so to conceive of God, provided that the 'common abstract features' (Alston, 1989, pp.66-67) he is thought of as sharing with human beings are *realized* in suitably different ways in the divine and in the human case. His central thesis is that if the functionalist analysis of mental state concepts is correct (or even *partly* correct (1989, p.70)) in taking them to be concepts of functional roles, then these common abstract features may include mental states. And if this is right, then we can apply psychological terms univocally to human beings and God without having to worry about failing to respect God's radical otherness.

---

[1] I use the term 'agents' rather than 'subjects' so as not to beg the question against someone who wonders whether some such system might be capable of acting even though it had no thoughts and feelings. (To be an agent, I take it, is just to be something that acts, while to be a subject is to be something that thinks and feels - that has a conscious mental life.)

## 8 Functionalism, Phenomenal Dispositionalism, and Theological Language

In this chapter, I will argue that while Alston's functionalist account of mental state concepts is ill-equipped to accommodate a number of important theistic convictions concerning the nature of God, a phenomenal dispositionalist account might do rather better. The univocal application of psychological terms to human beings and God may, I suggest, be possible on the basis that the 'common abstract features' these subjects share could include behavioural, phenomenal and cognitive dispositions.

I will begin, in section 8.2, by briefly considering the historical background to the problem under discussion. This is intended simply to give a sense of certain important currents in theistic thought, currents we will continue to feel pulling in this direction or that, as the main discussion proceeds. Section 8.3 will consist of an assessment of Alston's account, in the course of which it will become clear that it is vulnerable to a number of objections falling into two broad categories: those which take it to be repugnant to God's otherness, and those which take it to be repugnant to God's personhood. In section 8.4, I will argue that a phenomenal dispositionalist account of the meaning of mental state terms answers all of these objections - though not without inviting objections of its own from theists with certain convictions.

I should make it clear that it is not my intention in this chapter to take sides in pertinent theological controversies concerning whether or not God is atemporal, or perfectly simple, or possesses libertarian free will, and so on. I have no theological convictions of my own and no reason to favour one theist's convictions over another's (though I will assume it's the God of Christian theism we're talking about). Just as some theists would find Alston's functionalism a poor fit for the God they believe in, there is no escaping the fact that others would have similar reservations about phenomenal dispositionalism. What follows, then, should be seen as an exercise in comparing and contrasting an attempt to show how claims about God's mental states might be understood in functionalist terms, with an attempt to show how those claims might be understood in phenomenal dispositionalist terms - an exercise that will, I hope, help to illuminate the important differences between these positions. I shall not make it my business to *assess* such claims about God's mental states, beyond noting the respects in which they might strike someone as being concordant or discordant with this or that conviction concerning the nature of God.

It is also worth making the point that the question under discussion here is not only of interest to theists. Anyone who is concerned to understand theistic claims - including those who seek to challenge those claims - also has good reason to try to pin down the meanings of the terms in which those claims are

made.

## 8.2 Personhood and otherness in the theistic tradition

Let's begin by considering the problem under discussion in a little more depth.

In speaking of a personal God, the theist typically uses the very same terms she would use to speak of human persons; call them 'personalistic' predicates, or 'P-predicates' (a term Alston borrows from Strawson (1959)) (Alston, 1989, p.41). Thus God is said, for instance, to be *just* and *good*, to *love* his children, and to have *knowledge* and *desires* concerning his creation. But just what should we take such terms to mean, when applied to God? After all, God is supposed to be something quite different from a human person: where we are embodied, finite, and flawed, God is incorporeal, eternal, omnipotent, omniscient and morally perfect. God's mind must therefore be quite different from ours, and if this radical otherness of the divine is to be respected – if we are to avoid sliding into anthropomorphism – it seems we have reason to be wary of taking our P-predicates to have just the same meanings when applied to God that they have when applied to human persons.

But here the theist faces an immediate problem. For what meanings, other than the everyday meanings they have when we use them to speak about one another, are available for our P-predicates? Mill presses the point as follows:

> Language has no meaning for the words Just, Merciful, Benevolent, save that in which we predicate them of our fellow-creatures; and. . . if in affirming them of God we do not mean to affirm these very qualities, differing only as greater in degree, we are neither philosophically nor morally entitled to affirm them at all. . . If in ascribing goodness to God I do not mean. . . the goodness of which I have some knowledge, but an incomprehensible attribute of an incomprehensible substance. . . what do I mean by calling it goodness?. . . To say that God's goodness may be different in kind from man's goodness, what is it but saying. . . that God may possibly not be good?
>
> (Mill, 1999, pp.249-250)

Thus the theist is presented with an apparent dilemma: accept that P-predicates can be truly and univocally predicated of God and human persons (and hence

that God is not so radically other as all that), or deny that they can be so predicated (and hence accept that God is just *so* radically other that we have no way of saying what his attributes are).[2]

Those theists who have responded to this dilemma have typically opted for its second horn. Being unable to say anything about the positive attributes of God is, they have pointed out, perfectly consistent with being able to say all sorts of genuinely informative things about what God is *not*, and Maimonides (1999, pp.92-98) goes so far as to assert that talk about God should be conducted strictly within the limits of this *via negativa* (or negative path).

Still, the majority of theistic thinkers have been unwilling to go that far, and indeed have placed considerable weight upon the idea that God shares certain features with human beings. The teleological argument for the existence of God, for instance, rests on the idea that our orderly, efficiently-functioning world must have been caused to exist by something which reasons, desires, purposes and acts much as we do – something, that is, with 'a mind *like the human*' (Hume, 1998, p.35, Hume's italics). The fourth of Aquinas' 'five ways' to demonstrate the existence of God, meanwhile, rests on the idea that God and his creatures have certain properties – 'perfections' – in common, God having these most fully and causing his creatures to possess them in lesser degrees (1999, p.106). In a somewhat similar vein, Descartes (1996, pp.39-40) argues that knowledge of the divine nature can be acquired by recognizing the weakness and limitations of our own faculties – e.g. understanding, memory, imagination – and going on to form an idea of these same faculties as they are possessed in a perfected form by God. Nor can the philosophically inclined Christian theist afford to lose sight of the fact that the God of the Bible is a personal agent in a very strong sense: someone who commands and forgives, makes promises, exhibits jealousy and anger, and so on.

Of course, the history of theistic thought on this subject is not the history of two warring factions, 'mystics' and 'anthropomorphites' perhaps (to borrow the terms Hume's Cleanthes and Demea apply pejoratively to each other (Hume, 1998, p.28 *et passim*)), affirming God's radical otherness and his personhood, respectively. Rather, it is (by and large) the history of theists attempting to show that there can be a sense in which God is *both* a person *and* radically other than human persons. Aquinas, for example, maintains that while the *via negativa* does indeed offer one route to knowledge about God, the *via anologica* also makes possible the positive application of P-predicates to God, in senses

---

[2]cf. Soskice's (1987, pp.143-144) characterization of the empiricist's challenge to theism: 'the would-be theist must claim to know too much of God, or claim to know nothing of him at all'.

analogous to those in which they are used of human persons (2008, First Part, q.13, a.5; a.6). Other thinkers have emphasized the role of metaphor or symbolism in religious discourse.[3]In Alston's view, however, theistic thinkers have typically been too hasty in ruling out the possibility that terms may be applied univocally to human beings and God without compromising God's otherness (1989, p.67). It is this relatively neglected possibility[4] that he sets out to explore.

# 8.3 Alston's functionalist account of theological language

## 8.3.1 The account

In summarizing Alston's position, I will restrict myself to considering those aspects of his account which either motivate the objections we will be considering in subsections 8.3.2 and 8.3.3, or which I will later make use of in offering my own phenomenal dispositionalist account in section 8.4. I will also be restricting myself, at this stage, to considering the account just as Alston offers it in 'Functionalism and Theological Language' (1989, pp.64-80) rather than incorporating the modifications he makes to it in 'Divine and Human Action' (1989, pp.81-102). This will, I think, make for a more satisfying dialectic when we come to consider objections to Alston's position.

Since both functionalism and the divine nature may be conceived of in a number of different ways, it seems prudent to begin by identifying the conception of each with which Alston is working. Two features of Alston's conception of the

---

[3]Such thinkers have differed widely, however, in their assessment of just what the role of metaphor *is*. Swinburne (1992, pp.157-162) argues that the ability to convey information through the use of metaphor is one of the features of human language which makes possible the revelation of propositional truths about God, and gives a number of examples of how careful interpretation allows us to pin down the meaning of metaphorical statements concerning God. Soskice (1987, pp.148-153), on the other hand – taking experience of God rather than propositional revelation about God to be of prime importance to the theist – argues that the use of metaphor allows us to communicate something about such experience which could not otherwise be communicated, and hence to speak genuinely of God without having to pin down what we are saying to some sort of claim about how God is to be defined.

[4]The possibility has certainly not been wholly ignored; for instance, Swinburne (1992, pp.150-154) argues that P-predicates such as 'good' and 'wise' are univocally applicable to human beings and God. What makes Alston's account truly distinctive is not, therefore, the claim that God has general personality traits (such as being wise) just as we do; rather, it is the claim that this is so because God also has particular mental states with particular contents (such as beliefs or desires that P) just as we do.

divine nature are noteworthy. Firstly, he denies that God is perfectly simple. This is controversial, but (as we shall see) essential to getting a functionalist account of the divine mind off the ground. Secondly, he takes God to be atemporal rather than eternally enduring. This too is controversial, although in this instance (again, as we shall see) the implications for the applicability of functionalism to the divine mind are not so serious. Alston also, of course, takes God to be incorporeal, omniscient, omnipotent and morally perfect.

The version of functionalism with which Alston is concerned (1989, pp.67-68) is the familiar *analytic* functionalist position which takes our ordinary folk psychological concepts of mental states to be concepts of states occupying certain causal or functional roles. The great virtue of this position, from Alston's perspective, is that it takes our everyday psychological concepts, and hence the terms which express them, to be wholly neutral with regard to what sort of states actually play these functional roles. So even if the state playing some causal role in the divine psychology is constituted quite differently from the state playing that same role in the psychology of a human person, it will be perfectly legitimate for us to use the same term to refer to both. From this, Alston's central thesis follows quite naturally: insofar as the way in which such-and-such a mental state is realized in God may be quite different from the way in which it is realized in a human being, we can legitimately maintain that God is so radically other than human beings that we cannot know what it is (or, especially, what it is *like*) for God, say, to believe or desire that P; yet insofar as such mental states play the same functional role in the divine mind that they play in the human, we can maintain with equal legitimacy that P-predicates are applicable to God and to human beings in just the same, readily understandable sense[5] (Alston ,1989, p.69).

Of course, this line of reasoning will only bear fruit if the divine and human psychologies can in fact have functional properties in common. Alston therefore argues at some length that such commonality is not ruled out by the differences between God and human beings. Rather than follow his line of thought in detail, I will simply note some of the most important points he makes:

- Since the preeminent functional role played by mental states is the production of overt behaviour, and since in the human case this involves bodily movement, we need to make sense of the idea that an incorporeal being is capable of such behaviour (is capable, that is, of acting). This we

---

[5]Note, however, that this sense need not be just the established, everyday sense a P-predicate already has. Alston is arguing not for the possibility of that sort of 'straight univocity', but rather for the possibility of 'modified univocity', i.e. the possibility that terms may be devised which are applicable univocally to God and human beings (1989, pp.65-67; p.73).

can do simply by recognizing that what is central to our concept of action is not bodily movement, but the intentional bringing about of changes in the world. (Alston, 1989, p.72)

- Since performing a function involves temporal duration, and since causality may be thought to imply temporal succession, we should conceive of functional psychological concepts in terms of lawlike connections rather than causal roles in order to make them applicable to an atemporal God (Alston, 1989, pp.72-73).

- The divine psychology is vastly simpler than the human psychology. The whole range of human goal-setting states (desire-like states) can be replaced in the divine case with a a single state appropriate to a being which is not subject to things like 'biological cravings', 'gusts of passion', 'uncontrollable longing', or indeed any kind of desire which runs contrary to what is recognized to be right; this state could be termed 'Recognizing that it is good that *p*'. Similarly, with regard to cognitive guidance states (belief-like states), God will not have 'degrees of belief' as human beings do, or indeed any beliefs which do not count as knowledge. For any true proposition *p*, God will (since he is omniscient) have the 'intuitive' knowledge that *p*. Thus God's cognitive guidance states do not, as ours do, play a role in inferential processes. They do, however, have 'lawlike connections' to other such states, these connections being logical relationships. (Alston, 1989, pp.75-78)

- If God is omnidetermining – if, that is, his only overt action is the creation of a universe of which every detail is fixed – then cognitive guidance of his behaviour will consist solely in his evaluative apprehension of the differing possibilities he could choose to actualize. If God leaves some details of creation undetermined, however – viz. the choices made by free agents and the effects of their making those choices – there will be actions taken by God which are guided by his knowledge of states of affairs which do not obtain solely in virtue of his own creative activity. (Alston, 1989, pp.76-77)

Clearly, then, there will only be a limited degree of commonality between the divine mind and the human, and in this sense there will only be a 'core of meaning' (Alston, 1989, p.79) which is shared by psychological terms as they apply to God and to human beings. This core of meaning can be identified, according to Alston, by starting with a human psychological concept and then stripping away everything that has to do with human limitations, until what's left is a psychological term that is literally applicable to God. Does this mean that

God's mental life will appear enriched as we proceed to do for our concepts of love, hope, sorrow and so on what Alston has done for belief and desire? To think so would be to miss Alston's point about the simplicity of the divine psychology. The functionalist is, after all, only interested in identifying those mental state concepts required to explain the full range of an agent's possible behaviour; and we have already seen that the full range of God's possible behaviour can, on Alston's view, be explained entirely in terms of a single goal-setting state (recognizing that it would be best that *p*) and a single cognitive guidance state (knowing intuitively that *p*). So on Alston's view, any P-predicate which can legitimately be applied to God must - in the final analysis - be telling us something about *just those states*.

## 8.3.2 Objections from divine otherness

### 8.3.2.1 The objection from divine simplicity

Since Alston sets out to provide us with an account of how P-predicates can be applied to God in a way that respects both his radical otherness and his personhood, it seems appropriate to divide objections to his account into those which take him to have failed to respect one or other of these things. I will begin by considering three objections which take Alston's account to be repugnant to God's otherness.

The objection from divine simplicity is motivated by the following consideration. In order for a functionalist account to be applicable to an incorporeal being, some account must be available of what it is for such a being to be in this or that *state*. Now, for a *physical* thing like a brain to be in this state rather than that is, on a typical functionalist understanding, for it to have some particular set of physical features rather than another such set; to have some particular structure or composition, perhaps, or to exhibit some particular pattern of physical activity. For an incorporeal being to be in this state rather than that, then, is presumably for it to have some analogous set of nonphysical features, features that we can conceive of as playing a role in explanations of that being's mental activity and behaviour. Thus we might think of such a being as possessing some particular set of ideas, or mental pictures, or propositions, or whatever, which might be stored or accessed in a certain way or manifest themselves in certain patterns of thought (e.g. perhaps two particular ideas might be recalled from memory, 'held before the mind's eye' and compared). This sort of picture, though, is flatly inapplicable to a being which is perfectly

simple, and to which, therefore, no sort of structure or composition or arrangement of parts whatsoever may be attributed. Alston is perfectly well aware of this, and makes it clear from the outset (1989, pp.64-66) that his account is not intended to be applicable to a perfectly simple God. I mention it here simply to make clear that the theist who conceives of God in this way has reason to reject Alston's account outright.

### 8.3.2.2 The objection from divine omniscience

Central to a functionalist account of mind and action is the identification of two distinct types of mental state, goal-setting states and cognitive guidance states. Now, in the human case, cognitive guidance states – beliefs – are taken to be functionally related not just to behavioural 'outputs', but also to sensory 'inputs'; paradigmatically, a belief is acquired by a subject in response to information taken in through her senses, and then plays a role in the production of her behaviour. God, of course – since he just doesn't have senses as we do – can't be thought of as acquiring beliefs in response to sensory input. However, since an omniscient God knows that $p$ for any true proposition $p$, there will be a lawful co-variation between his cognitive guidance states and facts about the world. Alston proposes, therefore, that we can simply think of all the true propositions there are - or all the facts about the world - as constituting input to the divine mind (1989, p.96).[6]

Laura Garcia, whose (1992) includes a sustained critical assessment of Alston's position, rejects this proposal on the following grounds (1992, pp.207-208). For any given action taken by God, there will be certain facts about the world – certain contingently true propositions – which are true as a result of that action. And God's knowledge of those facts can't possibly play a role in the production of that very action; it would be absurd to suggest that God's knowledge that $p$ is (in part) responsible for his bringing it about that $p$. So in considering any given action of God, it just doesn't seem possible to think of all the true propositions there are as constituting input to the divine mind.

---

[6] I am sidestepping a terminological issue here. Alston's stated intention (1989, p.70) is to leave functional concepts having to do with inputs entirely out of the picture; hence he maintains that true propositions, or facts about the world, provide us only with an *analogue* of input in the divine case (1989, p.96). I will not persist with this distinction between inputs and analogues of inputs. Certainly the picture of some sort of perceptual fact being physically inserted into someone's head via the sense organs doesn't carry across from the human to the divine case; but if we're serious about adopting a functionalist account of the meaning of mental-state terms, and believe that we have identified something which plays essentially the same functional role in the divine case that input plays in the human, I think we may as well go ahead and call it 'input'.

This is only a problem, however, if Garcia is correct in supposing that on Alston's view, all the true propositions there are must be thought of *en masse* as constituting input to the divine mind in considering any given action taken by God. And it is easy to see why the theist might want to make just that assumption: an omniscient God can't strictly be *lacking* any items of knowledge when he acts. However, theists have often felt it important (in order to maintain that human beings are genuinely free agents, and that God is not culpable for creating individuals who do evil) to preserve the notion that God can, in some sense, acquire 'new' information stemming from the free choices we make. Thus Alston distinguishes clearly between God's original creative activity, which is not guided by his knowledge of the choices made by free agents and the states of affairs which obtain in virtue of their making those choices, and those aspects of divine activity which are guided by such knowledge – knowledge which God has only because he 'looks and sees' how things turn out (1989, p.77).

In light of Garcia's critique, it appears that God's knowledge of contingent truths resulting from his own actions must likewise be conceived of as 'new' information, knowledge God has because he looks and sees how the world is. (This seems a little odd, since God must know very well, in taking some given action, what the world will be like as a result of that action; but I think there is a legitimate distinction to be drawn between God's knowledge that his action will bring it about that *p*, and his knowledge that *p*.) This 'new' information, along with 'new' information stemming from the free choices made by human agents, may then be conceived of as playing a functional role in the production of a certain class of divine actions: those which consist in God's 'responses' to contingent states of affairs, rather than his original creation of the world *ex nihilo*.

If God is taken to be omnidetermining, however, the implications of Garcia's critique are much more serious. For if God performs just one overt action, the creation of a universe of which every detail is fixed, then there is no behaviour-production role which could possibly be played by his knowledge of contingent truths resulting from that action (as distinct from his knowledge that those contingent truths *will result* from that action). And since on a functionalist account, for an agent to know something *just is* for her to be in some state which plays a functional role in the production of behaviour or at least *would* play such a role in certain circumstances - it appears that on such an account, God just doesn't have that knowledge. Since God is necessarily omniscient, therefore, a functionalist account of cognitive guidance states is flatly inapplicable to an omnidetermining God.

### 8.3.2.3 The objection from divine freedom

In offering an account of mentalistic language on which mental events and states are conceived of as figuring in a causally cohesive, broadly naturalistic picture of the world – a picture within which everything that happens can be explained in terms of lawlike regularities – functionalism has often been thought to advance a physicalist agenda (Armstrong, 1993; Lewis, 2002). But perhaps it need not do so at all. Perhaps the physical realm is not causally closed, and there are nonphysical things that can interact with physical things in lawlike ways.[7] In that case there could be, as Alston proposes, states of an incorporeal being which play a functional role in a causal chain leading from facts about the world ('inputs') to changes intentionally effected in the world ('outputs'). Still, the theist may very well find himself wondering just where this gets us – for this is not at all the way that God, who is supposed to be a supernatural, perfectly free being who creates natural laws rather than being subject to them, fits into the theistic world view.

And indeed, in his 'Divine and Human Action', Alston acknowledges (1989, p.91) that his account in 'Functionalism and Theological Language' is defective precisely in that it makes God subject to natural laws, thus removing his freedom of choice.[8] Garcia (1992, p.206) draws attention to Alston's attempt to modify that account by bringing 'agent causation'[9] into the picture, his proposal being that an agent's mental states do not strictly determine that she will act thus-and-so, since her free volition may intervene between those states and her behaviour (1989, pp.92-95).

---

[7]Thus 'functionalism is compatible with the sort of dualism that takes mental states to cause, and be caused by, physical states' (Levin, 2013, section 1) – i.e. with the sort of *interactionism* proposed by Descartes (1996). Many have wondered, of course, just how things as fundamentally different as the dualist supposes mental and physical states to be *could* interact causally; this is a question which has been pressed upon the interactionist since Descartes' own time, when Princess Elisabeth of Bohemia was among the first to identify the problem (Descartes, 1978, p.106; pp.111-112).

[8]I am here following Alston (1989, p.91) and Garcia (1992, p.206) in assuming that on a standard functionalist account, these natural laws are to be thought of as deterministic.

Traditionally, concerns over divine freedom have been motivated not by any suggestion that God is *subject to natural laws*, but by the suggestion that God may be *constrained by his own nature* to act in some particular way – that he may, in particular (given his perfect goodness) lack the freedom to refrain from creating the best possible world (see Rowe's (2004) for an extended discussion of this problem). It is characteristic of Alston's attempt to apply a model of the human mind to God that the problem he has in accommodating divine freedom is more akin to the traditional problem of reconciling human freedom and determinism than to the traditional problem of reconciling the perfect freedom and the perfect goodness of God.

[9]Agent causation is distinguished by its proponents from *event causation*, the idea being – roughly – that an agent's actions can be free (in a libertarian sense) insofar as they are caused by *the agent herself* rather than by events in her mind (Pink, 2004, pp.111-112).

As Garcia notes (1992, pp.206-207), and as Alston himself admits (1989, pp.94-95), this is to insert an irreducibly non-functionalist notion – libertarian free volition – into his account. But this should come as no surprise, since Alston clearly has no interest in preserving the naturalistic world view of traditional functionalism except insofar as it is embedded in those meanings of P-predicates he himself proposes to devise. Of course, if those meanings strike the naturalist as chimeras, crude attempts to graft disreputable talk of libertarian free will onto an otherwise respectable way of speaking about agents, this matters to Alston not one jot. But the conception of God embedded in the sort of meanings he proposes to devise for our P-predicates still has to be coherent and tenable.

Garcia maintains that it is not, for the following reason. Alston's view rests fundamentally on the idea that it is legitimate to speak of God as having 'tendencies' (Alston, 1989, p.91) to act in this way or that, and that when we do so we are not just talking about statistical laws or counterfactual truths; rather, we are attributing to God states that play a causal role in the production of his behaviour. Seemingly, then, these tendencies are to be thought of as pulling an agent in this direction or that, with the strongest tendency winning the day and deciding what action is taken unless checked by a free volition. But if this is correct, then Alston is quite wrong to suggest that God's free agency is causally responsible for his actions; rather his actions are the automatic result of those of his tendencies which he elects not to block. As Garcia suggests, God's will here is in a position analogous to that of a bomb specialist who can through her free volition *allow* a bomb to go off or not, but cannot through her free volition *cause* that bomb to go off. Just as the bomb's being set will cause an explosion unless the specialist chooses to intervene, so some state of God will cause him to act thus-and-so unless he chooses to intervene. This is a quite unacceptably passive role for the divine will – God, surely, must be thought of as having the power actively to bring things about through the exercise of his will, not just the power to block tendencies to which he finds himself subject. Nor will it help Alston's case to suggest that God's tendencies can instead be thought of as consisting in his actively pushing in this direction or that, actively seeking or striving towards some goal, for this would involve his actively pursuing multiple incompatible goals at once[10] – a clearly irrational thing to do (Garcia, 1992, pp.208-211).

These points are, I think, sufficiently powerful to suggest that Alston's account

---

[10]This is so because if God didn't have tendencies to pursue incompatible goals, there would be no question of his choosing to pursue one goal rather than another by choosing to act in accordance with one tendency rather than another.

of the roles played by God's will and by his mental states in the production of behaviour is not satisfactory just as it stands. However, the fundamental difficulty Alston faces here – squaring the evident importance of mental states to any satisfactory account of an agent's behaviour, with the intuition that freedom of the will entails that her behaviour is not simply *determined* by such states – is hardly a new problem thrown up by his attempt to apply functionalism to the divine mind. On the contrary, it is an all-too-familiar problem which faces any account of agency, whether divine or human. If Garcia is right and Alston fails to give an adequate account of just how libertarian free will is supposed to work – well, he is in good company. But here the defender of Alston could simply help himself to a 'get out of jail free' card – appealing, perhaps, to Kant's distinction between the causality of nature and the causality of freedom (1993, p.376), or to Colin McGinn's (1993) 'mysterian' view that the problem of free will is among those philosophical stumpers which are simply insoluble for beings with our cognitive abilities – and choose to respect the intuition that agents, and especially God, just must *have* libertarian free will whether we can see how it works or not. So long as God's freedom poses poses no *special* problem, that strikes me as a reasonable line to take. And we must be careful not to lose sight of the fact that Alston is attempting only to show how meanings could be devised for P-predicates which are applicable to God, not to give a complete and accurate account of the workings of the divine mind. If 'God is in mental state S' must be taken to mean something a bit vague and woolly-sounding like 'God is in a state which plays a causal role in his tending to take this sort of action in this sort of circumstances, but doesn't strictly *determine* that he does so since he has free will', maybe that will have to do. (Very plausibly, this leaves us no worse off in the divine case than in the human, which seems like a reasonable benchmark against which to judge the success of an account of God's mental states.)

So, one option for the defender of Alston's position in the face of Garcia's critique is just to insist that the sort of meanings he proposes be devised for P-predicates as they apply to God may be legitimately so devised even in the absence of a complete and satisfactory account of the divine mind, and to take it on faith that propositions about God expressed in those terms will not then be flatly false (since God's actions really must be the result of some such interaction between his mental states and his will as is being presupposed). A radical alternative would be to plump for a rationalist compatibilist account of freedom, according to which God is perfectly free precisely because he is perfectly rational (see Pink, 2004, pp.48-49). This would satisfy neither Garcia nor Alston, both of whom insist on some form of libertarian freedom for God,

but it would give us a very neat functionalist model of the divine mind. On this view, God's mental states could be thought of quite straightforwardly as being causally responsible for his behaviour, just as functionalism seems to require. To the objection that this makes God into a mere automaton, a rationalist compatibilist advocate of functionalism about the divine mind could offer two replies. Firstly, he can simply affirm that rationalist compatibilist freedom – with an agent's actions being determined rationally by his cognitive guidance and goal-setting states, unhindered by interference from environmental factors and irrational impulses – is precisely the sort of freedom we should expect to find exemplified in a perfectly rational, omniscient being. Secondly, he can observe that in the divine case, there is no place for the sort of anxieties raised by compatibilist accounts of *human* freedom, anxieties stemming from the apparent possibility of tracing a causal chain back from an agent's goal-setting states to the causes (perhaps biological, cultural or physical) of her being in those states; for God would himself be the ultimate source of all his goal-setting states.[11]

## 8.3.3 Objections from divine personhood

### 8.3.3.1 The objection from the phenomenal character of mental life

Let's now move on to consider the respects in which Alston's account might be thought repugnant to God's personhood.

Alston openly acknowledges that the concepts of divine mental states we are left with in light of his functionalist account 'are very thin, to say the least' (Alston, 1989, p.98). We have an idea of the sort of relations in which God's mental states stand to his actions and to each other, but that's all we have; in particular, and in contrast to the human case, we lack any idea of what it is like to be in those states. This observation serves to put the spotlight firmly on one of functionalism's major perceived shortcomings: its failure to capture important facts about the phenomenal character of mental life.

As we have seen, one of the virtues of the functionalist approach to mental states in general is that it is not 'chauvinistic' as far as such states are concerned – it takes seriously the idea that things quite different from human beings might have the same mental states we do, but differently realized. This

---

[11] I am assuming here that it is no threat to divine freedom if some of God's cognitive guidance states have an 'ultimate source' outside God himself, e.g. if God's knowing that $p$ is a consequence of its being the case that $p$.

enables us to understand how a Martian, say, or a robot, or of course an incorporeal being, might be said to have the very same types of mental states that we have, even though it didn't have a physical structure anything like ours. But what if something had neither a physical structure anything like ours, nor an 'inner life' anything like ours? What if something with no capacity for conscious experience whatever – a machine, a vegetable, an organizational structure – appeared nonetheless to be capable of entering states having the same sorts of functional role as our mental states (e.g. registering states of affairs, setting goals and effecting changes in the world)? Would we not then find ourselves in the absurd position of attributing mental states to things which self-evidently don't have minds?[12]

More specifically, is this the position Alston leads us to with regard to God? It is all very well, one might argue, for Alston to say that on his account, we don't know what it is like to be God; but couldn't we go further and say that on his account, it seems that there is *nothing* that it is like to be God? After all, it is not as if we don't know what it is like to be God in just the way that Nagel (1974) suggests we don't know what it is like to be a bat. While a bat differs from us in ways that make it impossible to imagine just what its conscious experiences might be like, it is also similar to us in ways that make it reasonable to suppose that it has broadly the same *types* of experience we do – experiences of things like physical sensations, affective responses and mental processes. But on Alston's account, God just does not have things like physical sensations or affective responses, and does not go through mental processes of any sort. So what is there for him to have conscious experiences *of*? In what could his inner life – the subjective aspect of his existence – possibly consist? In the absence of any answer to questions like these, one would surely be justified in concluding that Alston's God is best characterized not as a personal agent, but as a sort of moral equivalent of gravity – an impersonal, mindless force guiding the universe. And that is a bit *too* radically other for most theists.

### 8.3.3.2 The objection from religious belief and practice

A related, but more general problem is this. Although God may, on Alston's account, be 'personal' in some sense – Alston characterizes him as 'a timeless "personal system" of functionally interrelated psychological states' (Alston, 1989, p.101) – he is plainly not the sort of person one could relate to in the way that much of our religious discourse takes human beings to relate to God. He is

---

[12]The suggestion here, that in avoiding chauvinism functionalism ends up committing the contrary sin of liberalism, is closely associated with Ned Block; see for example his (1991).

not, that is, the sort of thing 'with which we can coherently conceive ourselves to be in dynamic personal relations of dialogue, support, love or instruction' (Alston, 1989, p.101). Alston is forced to acknowledge, therefore, that while that sort of discourse may essential to the practice of religion, it must strictly be interpreted in non-literal terms (1989, pp.101-102). But this seems likely to have many theists tearing their hair out in despair, on two counts. Firstly, it means that while Alston's account may offer some insight into what it might mean for God to know or to desire that *p,* it leaves our original dilemma perfectly intact as regards the vast bulk of our talk of God. What does it mean to say that God loves us, hears and answers our prayers, gets angry with us, is merciful or just in his dealings with us, and so on? We still don't know. Worse than that – and this is the second reason one might have to despair at Alston's conclusion – it looks all too possible that such talk doesn't really mean *anything*, because it's based on a view of God that is flatly mistaken. It's all very well, one might respond to Alston, saying that we need to conceive of God in dynamic, personal terms for religious purposes, but if we know perfectly well that *that's not what God is like*, then natural theology and religion look to be in danger of pulling away from each other in a potentially alarming way.

## 8.3.4  Conclusion of section 8.3

Given the wide variation in theistic conceptions of God, none of the foregoing objections is flat-out fatal to Alston's account. That account might, in fact, appear perfectly acceptable to someone with the following theistic commitments: that God is not perfectly simple; that God is not omnidetermining; that the perfect freedom of God does not (assuming the truth of rationalist compatibilism) entail the falsity of determinism; that God's being a person with a mind does not entail his having conscious experiences of the sort we have when we perceive the world via our senses, go through cognitive processes, and have affective responses; and that God is not literally the sort of being much of our religious discourse takes him to be – a dynamic, personal agent who interacts with his creatures – and so those portions of our religious discourse must be interpreted metaphorically or symbolically. Nor might that account appear wholly unacceptable to someone who broadly took that view but had reservations about signing up to rationalist compatibilism.

However, very many theists think that God *is* wholly simple, *is* omnidetermining, and (most importantly) *is* a dynamic, personal agent - one whose interactions with our forebears are faithfully recorded in sacred texts, and one with

whom we may interact today. It seems to me that many theists might think Alston has done more harm than good by leading us to a conception of a God who is a personal agent in only a terribly thin sense, if at all. Hence it is worth our while to ask whether some account of the meaning of P-predicates is available that allows for their univocal application to human persons and to the sort of God in which these theists believe. In section 8.4, I will ask whether a phenomenal dispositionalist account might fit the bill.

## 8.4 Phenomenal dispositionalism and theological language

Just as Alston concedes from the outset that functionalism is flatly inapplicable to a perfectly simple God, I must begin this section by conceding that phenomenal dispositionalism is flatly inapplicable to an atemporal God. This is simply because conscious experiences are - so far as I can see - necessarily temporal; they are something one 'goes through'. Thus it does not appear coherent to suppose that a being with no experience of time whatever could, as phenomenal dispositionalism requires, be disposed to have conscious experiences. What follows is, therefore, an attempt to apply phenomenal dispositionalism to a God who is *temporally* eternal as well as incorporeal, omniscient, omnipotent and perfectly good.

Of course, this means that those theists who are committed to God's atemporality will judge me to have fallen at the first hurdle. So be it. No account of the meaning of P-predicates as applied to God could hope to please all of the people all of the time; and if a phenomenal dispositionalist reading of theistic claims leads us to conclude that those claims are incoherent - to conclude, for instance, that God could not be both atemporal *and* personal - we can always opt to revise those claims rather than to revise that reading.

A very welcome feature of phenomenal dispositionalism, from the theist's perspective, is that it is relatively untainted by any physicalist or, more broadly, naturalist agenda. Hence I can be spared the sort of agonies Alston goes through in trying to integrate nonphysical states and undetermined volitions into a functionalist picture which is typically seen as advancing such agendas. In other respects, however, I can afford to follow Alston's line of thinking quite closely. A phenomenal dispositional account of divine mental states, like his functionalist account, will hinge on the possibility that God may have certain abstract features in common with human beings; it is just that where Alston

takes these features to include functional roles played by mental states, I will be taking them to include phenomenal, cognitive and behavioural dispositions. Like Alston, then, what I need to do is to establish that God could in fact share the abstract features in question with human beings.

And of course, Alston has already done much of this work for me. His suggestion that a nonphysical being may be thought of as acting if it is capable of intentionally bringing about changes in the world will serve perfectly well to explain how behavioural dispositions may be ascribed to an incorporeal God. Similarly, his identification of the lawlike connections (in the form of logical relationships) which hold between God's cognitive guidance states provides us with a useful way of thinking about God's cognitive dispositions: there is no barrier to our saying that just as a human being who believes that *p* will, if *p* entails that *q*, typically be disposed to believe that *q*, so will God (though since God is unique, and certain to believe everything that is entailed by every true proposition, the 'typically' will drop out in the divine case).

With regard to God's phenomenal dispositions, however, I will have to stand on my own two feet. The question facing us, then, is whether the mental states of an incorporeal, infinite and eternal being could have a phenomenal character similar to that of human mental states. Could God be disposed to feel sorrow at the suffering of someone he loves, for instance, or to feel anger towards the person who has intentionally caused that suffering?

It might be argued that God's having such emotional responses is ruled out by the doctrine of *divine impassibility*, according to which God is 'not subject to suffering, pain, or the ebb and flow of involuntary passions' (Johnson, 2000). Here we must note, first of all, that there are many theists who simply deny this doctrine.[13] But we can also observe that this doctrine has not in any case universally been thought to entail that God does not have 'real and powerful feelings'; rather, it has been thought by some to entail only that his feelings are never 'involuntary, irrational or out of control' (Johnson, 2000). And since phenomenal dispositionalism is not in the business of assigning 'feelings' any position in a causal chain – since it is perfectly neutral on such questions as whether they are caused directly by external events or under the control of the will, or whether they can overrule or be overruled by reason – this point can be accommodated quite straightforwardly.

A second, related question has to do not with whether God is *capable* of hav-

---

[13]The denial of divine impassibility is especially characteristic of *open theism*, a position which in general takes God to be a dynamic personal agent in a very strong sense (see Hasker, 2004, p.97 *et passim*).

ing mental states with a phenomenal character similar to that of human mental states, but whether it is legitimate for us to suppose that he *does*. Alston makes it very clear that he thinks it is important to 'preserve the point... that we can form no notion of what it is like to be God' (Alston, 1989, p.71). However, if a being shares (and indeed, exceeds) our capacity to love and care for people, to be merciful and compassionate towards them, to understand their weaknesses, and so on, then there is surely at the very least a strong *prima facie* case for supposing that that being has an inner life which is somewhat like ours with regard to its phenomenal character. That being may, it seems to me, legitimately be presumed to *feel* compassion for people, *feel* sorrow concerning the suffering of those it loves, and so on. So while the theist should no doubt stop short of saying that he knows *just what it's like* to be God - not least because an omniscient and omnipresent God has an awful lot of things to have feelings about at any given moment - it it is less than clear that he must refrain from attributing to God dispositions to have conscious experiences having a phenomenal character somewhat similar to our own.

Forming a rounded picture of what it would be for God to be in this or that mental state will be a matter of thinking about what sort of person God is and what sort of behavioural, phenomenal and cognitive dispositions a person of that sort ought to have, given what we take to be true about the world and about moral values. (Note that whatever dispositions God *ought* rationally to have, he *will* have. The caveats I have raised elsewhere about taking an overly rationalistic approach to the characterization of mental states in human beings - given our propensity to fail to draw correct inferences, to act on the basis of irrational fears, and so on - don't apply in God's case.) According to the theist, we know rather a lot about the sort of person God is; we know that he is a perfectly rational person, a perfectly good person, a perfectly loving person, a perfectly just person, a person of infinite power and knowledge, and so on. (Just how far this list can be extended will depend on the precise status of biblical and other reports concerning God's 'inner life' and overt actions.) So for instance – and here we find ourselves, once again, following Alston's line of thinking quite closely – we know perfectly well that if *p* is the case, and if *p* entails that *q*, then God will know that *q*; that if it would be best that *p*, then God will desire that *p*; that if God desires that *p*, and if the most rational way for him to bring it about that *p* would be to take action *a*, then he will take action *a* (cf. Alston 1989, pp.75-76 *et passim*). And if the foregoing account of God's phenomenal dispositions is on the right lines, we are also entitled to suppose that, for instance, God's loving someone or his being angry with someone entails his having certain sorts of feelings.

This last point serves to highlight a further important difference between Alston's account and its phenomenal dispositionalist rival. On Alston's account, we must apparently regard God's mental states as wholly generic, incapable of being characterized in any terms other than their functional role in setting goals or providing cognitive guidance. Thus Alston is forced to the conclusion that claims about God's mental life which go beyond the attribution to him of knowledge concerning (possible) states of affairs and their moral value must be interpreted metaphorically or symbolically. Phenomenal dispositionalism, on the other hand, makes possible the literal application to God of a rich psychological vocabulary – one which attributes to him mental states which vary with regard to their phenomenal character.

Thus all of the objections we have seen levelled at Alston's functionalist account of divine mental states are answered by switching to a phenomenal dispositionalist account. Since we are no longer taking mental states to be internal structures of a complex system, there is (so far as I can see) no bar to conceiving of God as perfectly simple; a perfectly simple being may be conceived of as having dispositions of the sort phenomenal dispositionalism takes our mental state terms to ascribe to subjects. And since 'causal chains' simply do not figure in our account of dispositions, there is no question of a problem arising as regards the compatibility of free will and determinism. Phenomenal dispositionalism is perfectly neutral on this issue, entailing no claims about the nature of causal processes that might or might not explain *why* certain counterfactuals hold true of the subjects to whom we ascribe dispositions.

Since phenomenal dispositionalism does not conceive of knowledge in functional terms, there is also no problem with attributing knowledge to God that plays no functional role. Nonetheless, there is a lingering question here: just what sort of dispositions are we attributing to God when we describe him as having this or that belief concerning contingent facts which are the result of his creative action? The obvious candidates now become cognitive and phenomenal dispositions. Even if there are no behavioural dispositions whatsoever associated with God's believing that *p*, he will nonetheless be disposed to believe everything that is entailed by *p* and perhaps to have conscious experiences having to do with *p* (such as feeling sorrow or anger concerning *p*).

And of course, phenomenal dispositionalism is particularly well suited to answering the objection that the God described by Alston is unrecognizable as the dynamic, personal God whose interactions with human beings are at the heart of much religious belief and practice. Phenomenal dispositionalism insists that God's being a person with a mind entails his having a temporal exis-

tence, and his having dispositions to act, think and feel thus-and-so in different circumstances – meaning that he may very well be just the sort of dynamic, personal agent much of our religious discourse takes him to be. Hence there is no need for us to set off down the road of interpreting all such discourse in metaphorical or symbolic terms – a road which looks to many theists like a slippery slope terminating in *de facto* atheism.

## 8.5 Conclusion

Because I presented the material in section 8.4 as a sort of parallel to Alston's functionalist account of divine mental states, it perhaps read rather like a pitch to theists intended to encourage them to sign up to a rival phenomenal dispositionalist account. So let me take a step back from this exercise in applying phenomenal dispositionalism to a particular question, place that exercise in a wider context, and then say what I think this chapter has really been about.

Analytic functionalism is a theory about the meanings of our mental state terms. If this human, this animal, this robot, this deity has states that play such-and-such causal roles - well then, yes, that human, that animal, that robot, that deity can truthfully be said to have such-and-such 'beliefs', 'desires', 'hopes', 'fears', and so on, since that is what we *mean* when we use those terms. Pinning down the meanings of those terms in this way means we can pin down just what it is we're asking when we ask (say) if animals have beliefs, if a computer could have desires, or if God could have mental states in common with human persons. By making certain assumptions about what sort of thing X is, and then trying to figure out whether a thing of that sort has (or could have) states playing such-and-such functional roles, you can decide whether or not it would be legitimate to ascribe such-and-such mental states to that thing.

As we have seen, however, one of the major objections to functionalism is that this line of thinking licenses the ascription of mental states to things that patently do not have minds - things that functionally resemble genuine subjects in that they generate the right sort of behavioural outputs in response to the right sort of environmental inputs, but that wholly lack the sort of conscious mental life that makes genuine subjects *subjects*. One plausible objection to Alston's functionalist account of the divine mind is that it does precisely this - licensing the ascription of mental states to something that we should not take to have a mind at all, since it has no conscious experiences. (Worse than that, Alston's God does not even go through any *unconscious* processes

that could be described, however loosely, in psychological or cognitive terms - which plausibly makes him a poorer candidate for mental state ascription than a pocket calculator.)

Phenomenal dispositionalism is also a theory about the meaning of our mental state terms, and similarly enables us to pin down just what it is we're asking when we ask about the mental states of non-human agents. It enables us to address those questions by trying to figure out whether a thing of such-and-such sort has (or could have) the right sort of phenomenal, behavioural and cognitive dispositions to make legitimate the ascription of such-and-such mental states to that thing. As such, it steers well clear of the liberalism that plagues functionalism. It does not rule out the possibility that computers or deities could have minds (say), but it invites us to consider more than their functional organization before concluding that they do. In the case of a computer, we might want to ask whether they are made of the right sort of stuff to give rise to conscious experiences. In the case of God, we might want to ask whether he is the sort of thing that could display intelligence. We can't just bundle phenomenal and cognitive dispositions in with behavioural dispositions and declare that whatever *acts* like a thinking, feeling thing thereby *is* a thinking, feeling thing.

In terms of the material in this chapter, then, the moral of the story is not supposed to be that, yes, God *is* the sort of thing that could have beliefs and desires, get angry, feel jealous, love people, and so on. The moral of the story is supposed to be that phenomenal dispositionalism provides us with a way of understanding *claims* about the mental states of God or of any other human or non-human agent - claims which can then be defended or refuted, reconciled with one another or dismissed as inconsistent, and so on.

On final point in relation to what has become a recurring theme in this thesis: Alston's attempt to apply functionalism to the divine mind represents an interesting case study in the boiling down of mental life to practical reasoning based on beliefs and desires. At the end of this process, the God he ends up describing is a person in a terribly thin sense, if at all. This being's 'mental life', if it has one, certainly has vanishingly little in common with the mental lives of human beings. Of course, Alston goes to the extremes he does only because he is seeking to characterize the mind of a being that, unlike human beings, is perfectly rational, is all-knowing, etc. But I think it is worth asking how far we can go down this road in characterizing human minds without getting that characterization drastically wrong.

In Chapter 9, I consider one response to concerns that belief-desire psychology is not up to the task of characterizing the mental lives of human beings:

Gendler's argument for the recognition of a state type she calls *alief*.

# 9 A Phenomenal Dispositionalist Case Against Alief

## 9.1 Introduction

As I noted in the conclusion to Chapter 8, a recurring concern in this thesis has been that practical-reasoning-based, belief-desire psychology is inadequate to the task of capturing the realities of mental life (and so to the task of predicting and explaining behaviour in mentalistic terms). When combined with the assumptions that this belief-desire psychology is folk psychology, and that folk psychology is a theory about the functional organization of the mind, it is easy to see how this concern could motivate eliminativist arguments that folk psychology will ultimately be discarded in favour of a theory that does a better job of describing the workings of the mind, in part by identifying the states that are *really* involved in those workings.

One response to such arguments - the response I have been advocating - is to adopt a dispositionalist account of mental state ascriptions. The appeal of this move is that it allows one to concede to the eliminativist that there may indeed be no such 'things' as beliefs and desires in a Standard View sense - that these terms may not pick out any individuable, physical, scientifically respectable states of subjects - without conceding that ascriptions of beliefs and desires to subjects must therefore, in the final analysis, be flatly false. Rather, the dispositionalist argues, such ascriptions can be true regardless of what states of a subject have what content or play what causal roles - just so long as the subject has the right dispositions.

In this chapter, I want to consider a different response to the sort of concerns that motivate eliminativism. In her (2008a), Tamar Gendler concedes that belief-desire psychology fails to pick out the states that actually produce much of our behaviour. But rather than agreeing with the eliminativist that folk psychological mental states must therefore be consigned to the dustbin of history, or with the dispositionalist that the Standard View must therefore be so

consigned, Gendler argues that we should simply incorporate into our existing taxonomy of mental states a new state type operating alongside belief and desire: *alief*. This hitherto-unidentified state type, Gendler argues, occupies an important causal role in generating much of our behaviour.

What I want to argue in this chapter is, essentially, that Gendler is attempting to lead us down a philosophical rabbit hole - one that we need not enter if we adopt phenomenal dispositionalism. In section 9.2, I will introduce Gendler's concept of alief, drawing attention to the sort of mental and behavioural phenomena she believes alief can help to explain and to certain important characteristics of alief as she defines it. In section 9.3, I will consider Muller and Bashour's argument that 'aliefs simply do not constitute a legitimate psychological category' (2011, p.371), and suggest that their criticism of Gendler (unwittingly, no doubt) invites us to continue Gendler's project of fleshing out and refining a philosophically respectable taxonomy of mental states (including alief-like states). In section 9.4, I will argue that the concept of alief is so badly flawed that to attempt to refine it would be futile, and that attempts to devise a philosophically respectable taxonomy of mental states are, in any case, fundamentally misguided.

## 9.2 Gendler on alief

In introducing the concept of alief, Gendler draws our attention to a number of cases like the following (2008a, pp.635-637):

- A subject walking on a transparent walkway high above the Grand Canyon (the 'Skywalk') feels as if she is in danger of falling, and as such proceeds more cautiously than she would if she were on an opaque walkway in the same location; but proceed she does, and she would not hesitate to affirm that the walkway is safe.

- A subject is disinclined to eat a piece of fudge shaped to look like dog faeces; but he would readily acknowledge that it would taste just the same and be just as safe as a normally-shaped piece of the same fudge that he would not hesitate to eat.

- A subject borrows some money because she has left her wallet at home; but upon taking the cash, she looks in her bag for her wallet (to put the money in).

Gendler insists that such subjects really do believe what, if asked, they would say they believe: that the walkway is safe, that the fudge is delicious, that the

wallet is at home. But they have *aliefs* that are out of line with those beliefs.

As defined by Gendler,

> Alief is a mental state... with associatively linked content... That is,
> a cluster of contents that tend to be co-activated... In paradigmatic
> cases, an activated alief has three sorts of components: (a) the
> representation of some object or concept or situation or circum-
> stance, perhaps propositionally, perhaps nonpropositionally, per-
> haps conceptually, perhaps nonconceptually; (b) the experience of
> some affective or emotional state; (c) the readying of some motor
> routine... Notwithstanding [that] characterization... I do not want to
> rule out the possibility of there being aliefs that involve the mental
> activation of a different sort of associative cluster. Perhaps there are
> cases where the activation occurs at a sufficiently low level to render
> the notion of representation inapplicable. Perhaps there are states
> that lack an obvious affective ingredient, or that do not include the
> clear activation of a motor routine, but that nonetheless sufficiently
> resemble our paradigm cases that we want to count them as aliefs.
> Perhaps there are cases where the most noticeable associations
> are not easily subsumed under the three categories offered—cases
> that primarily involve the heightening or dampening of certain sorts
> of attention, or the heightening or dampening of certain perceptual
> sensitivities.

> (Gendler, 2008a, pp.642-644)

In the three cases described above, the subjects' relevant aliefs have, suggests
Gendler, the following representational-affective-behavioural or R-A-B content
(respectively): 'Really high up, long long way down. Not a safe place to be! Get
off!!' (2008a, p.635); 'Filthy object! Contaminated! Stay away!' (2008a, p.636);
'Bunch of money. Needs to go into a safe place. Activate wallet-retrieval motor
routine now.' (2008a, p.637).

Aliefs, says Gendler, can be either dispositional or occurrent:

> A subject has an occurrent alief with representational-affective-behavioral
> content R-A-B when a cluster of dispositions to entertain simulta-
> neously R-ish thoughts, experience A, and engage in B are acti-
> vated—consciously or unconsciously—by some feature of the sub-
> ject's internal or ambient environment. A subject has a disposi-
> tional alief with representational-affective-behavioral content R-A-B
> when there is some (potential) internal or external stimulus such

> that, were she to encounter it, would cause her to occurrently alieve R-A-B.

> (Gendler, 2008a, p.645)

So, when a subject steps on to a transparent walkway high above the ground (for instance), what happens is that her dispositional alief with the content 'Really high up, long long way down. Not a safe place to be! Get off!!' is rendered occurrent.

Kriegel suggests that 'Gendler's case for the alief/belief distinction is a sort of argument from serviceability: the distinction is indispensable for explaining a host of mental phenomena that would be utterly perplexing without it' (2012, p.476). I agree that that is how Gendler's argument goes, but in assessing that argument, I think it's important to understand just who is supposed to find these phenomena perplexing, and why.

It's interesting to note, first of all, something Gendler says about the transparent-walkway case: 'Though some readers may find this story politically or aesthetically disturbing, none - I take it - find it perplexing' (2008a, p.634). Clearly, then, she accepts that the failure of a person calmly and confidently to walk along a transparent walkway high above the Grand Canyon, in spite of her avowed belief that the walkway is safe, is not going to leave the average person mystified in the way that - say - the failure of a brick to fall to the ground in spite of its lacking any visible means of support would leave him mystified. The latter event cries out for explanation in terms of some force, hitherto unrecognized by the average person, being at work; the former event does not.

Why not? Simply because we understand very well why the person stepping onto the transparent walkway acts as she does:

> The reason, of course, is that [stepping onto the walkway] activates a [certain] set of affective, cognitive, and behavioral association-patterns. . . input to her visual system suggests that she is striding off the edge of a cliff. This visual input activates a set of affective response patterns (feelings of anxiety) and motor routines (muscle contractions associated with hesitation and retreat), and the visual-vestibular mismatch produces feelings of dizziness and discomfort, leading to additional activation of motor routines associated with hesitation and withdrawal.

> (Gendler, 2008a, p.640)

In describing the phenomena that are her focus, then - cases of belief-behaviour mismatch - Gendler does not in fact expect her readers to find them perplex-

ing. On the contrary, she expects us to recognize the sort of psychological processes at work in these cases straight away.

And yet Gendler goes on to say this:

> I will argue that any theory that helps itself to notions like belief, desire, and pretense needs to include a notion like alief in order to make proper sense of a wide range of otherwise perplexing phenomena. Without such a notion, I will contend, either such phenomena remain overlooked or misdescribed, or they seem to mandate such a radical reconceptualization of the relation between cognition and behavior that traditional notions like belief seem quaint and inadequate. In short, I will argue that if you want to take seriously how human minds really work, and you want to save belief, then you need to make conceptual room for the notion of alief.
>
> (2008a, pp.641-642)

Just what is going on here? After conceding that these phenomena are not perplexing, Gendler now tells us that they are. After describing them in a way she takes to be both familiar and correct, she tells us that they are at risk of being overlooked or misdescribed. After explaining, in familiar and seemingly adequate terms, how one's automatic responses to stimuli can lead to behaviour that's out of line with one's avowed beliefs and intentions, she suggests that these cases may mandate a radical reconceptualization of the relation between cognition and behaviour. Why?

Gendler's worry is that these cases present us with an apparent dilemma. If we stick to a simple model based around belief-desire explanations of behaviour - which she takes to be a folk psychological model - we just won't be able to account for these phenomena; but if we adopt an alternative psychological model based around things like 'visual inputs' and 'motor routines' - a model that seemingly fares much better in terms of accounting for these phenomena - we concede to the eliminativist that an empirically respectable account of the causal processes underlying behaviour just is not going to be based around interactions between states corresponding to folk psychological 'beliefs' and 'desires'.

Gendler's response is to make a 'thus far and no further' concession to the eliminativist. We can recognize the importance of arational mental processes that have nothing to do with beliefs, she proposes, without denying the importance of rational mental processes that *do* involve belief. And we can continue to explain subjects' behaviour in terms of states rather than digging down a

level and describing detailed processes. All we need to do is divide mental labour between belief on the rational side and a different state, alief, on the arational.

Having set out her stall, Gendler proceeds to defend the view that the sort of mental phenomena she's interested in can't be accounted for using just the resources of belief or imagination. From there she makes the leap to the conclusion that we had better acknowledge the role of alief in producing behaviour. Presumably this is because Gendler takes a narrow view of the taxonomy of folk psychology: it recognizes beliefs, desires and imaginings as legitimate state types, but not much else. Certainly it does not deal in the *gut reactions*, *feelings as if P* and so on that are part and parcel of ordinary mentalistic discourse. Hence once belief and imagining have been tried and found wanting, we have to look outside the existing folk psychological taxonomy of state types for a philosophically respectable solution to the problems Gendler identifies.

## 9.3 Muller and Bashour's case against alief

In their (2011), Hans Muller and Bana Bashour argue that 'aliefs simply do not constitute a legitimate psychological category' (p.371). In this section, I want to consider their reasons for thinking so.

Muller and Bashour note (2011, pp.374-375 *et passim*) that belief is a normative rather than merely descriptive concept, on Gendler's view. Belief is not simply a mental state that, as it happens, is typically formed and revised through rational processes in response to evidence; rather, it is a mental state the nature of which is to be thus formed and revised. Mental states that don't live up to that standard just don't qualify as 'beliefs', and so if we don't act as we ought to act if we were rational and attending to the evidence, then our actions must be due not to beliefs, but to aliefs formed through sub-rational mechanisms.

Gendler recognizes (2008b, p.566n.26) that things can't really be this clear-cut, since there are clearly cases in which people's beliefs don't change as they apparently ought to change in response to evidence. Some of our deepest-seated beliefs, about religion and politics for instance, don't routinely shift in response to evidence, and Gendler concedes that 'the matter requires further thought' (2008b, p.566n.26). Muller and Bashour argue that this is a bigger problem for Gendler's account than she seems to realize:

> In fact, when we compare the ways that many of our beliefs and what she calls aliefs are shaped, we see that the alief/belief distinction collapses. Many of our deep-seated beliefs can be (and probably have been) formed through habituation, and many of what Gendler calls aliefs may be changed by changes in evidence.

> (Muller and Bashour, 2011, p.375)

In proposing his famous Wager, Muller and Bashour note, Blaise Pascal is acutely aware of the role played by habituation in giving rise to religious faith:

> Endeavour then to convince yourself, not by increase of proofs of God, but by the abatement of your passions. . . Learn from those who have been bound like you, and who now stake all their possessions. . . Follow the way by which they began: by acting as if they believed, taking the holy water, having masses said, etc. Even this naturally will make you believe and deaden your acuteness.

> (Pascal, 1958, p.68, quoted in Muller and Bashour, 2011, p.376)

In that case, the suggestion is that habituation can *help* a subject to acquire a belief in the absence of supporting evidence she finds compelling. But there are other cases in which habituation seems to *hinder* the revision of a subject's beliefs, even when that subject is presented with evidence. For instance, note Muller and Bashour (2011, p.375), a staunch homophobe is not likely to give up his belief that homosexuality is wrong just because he is shown that his reasons for thinking so are inadequate.

The other side of the coin is that the states Gendler calls aliefs often *do* change in response to evidence:

> Consider, for example, finding out that your favourite clothing store manufactures its products in sweat shops in underprivileged countries. As soon as you are aware of this information, your gut reaction when walking past that store turns from one of joy or temptation to one of disgust and anger. . . Here it is not simply the belief that changes, but also something more affective, and it involves your intuitive reaction. . . and this is surely the sort of phenomena [sic] Gendler intends the term 'alief' to cover.

> (Muller and Bashour, 2011, p.377)

For Muller and Bashour, however, the problem is not just that Gendler invites us to see a sharp distinction between alief and beliefs where no such distinction exists; it is also that she invites us to avert our eyes from important and genuine

distinctions between states she lumps together under the heading 'alief'. In particular, they argue:

> Lumping habits and instincts together obscures the role that affect plays in the explanation, prediction, and the phenomenology of action and reaction. A brief survey of Gendler's own examples brings out the salient fact that emotions are intricately involved in the instinct cases and wholly absent from the habit cases.

(Muller and Bashour, 2011, p.381)

In the case of the subject stepping on to the Skywalk, Muller and Bashour note (2011, p.382), clearly the feelings of fear she instinctively experiences are crucial to explaining her behaviour. Yet in the case of the subject reaching for her wallet through force of habit, the explanation for this behaviour seems to have nothing to do with her affective response to having money in her hand. Muller and Bashour make a similar point about the case of a subject who sets his watch five minutes fast, in order (as Gendler would have it) to induce in himself the alief that it is five minutes later than it really is:

> it is quite a stretch to say that fear of falling and a concern that one might be late count as "affective" in quite the same sense. And this is because the examples of instinct involve genuine emotional engagement in the situation while the case that involve habituation do not.

(Muller and Bashour, 2011, p.382)

Muller and Bashour are clearly on to something here: phenomenally speaking, there is a world of difference between acting a certain way through sheer force of habit, as if on autopilot, and acting a certain way because powerful, instinctive, affective responses to a situation have kicked in. However, I don't think the sort of sharp distinction proposed by Muller and Bashour will hold up, because it's just not the case either that 'genuine emotional engagement' is present in every case where instincts are at work, or that in no case where habituation is at work are affective responses involved 'in quite the same sense'. When I unthinkingly take a sip of water in response to a mild sensation of thirst, that action is instinctive if any is; but the strength of the affective response that prompts it is a good deal closer to that of the man who feels a slight sense of urgency upon looking at his watch than to that of the man who finds himself overwhelmed by fear as he stands on the Skywalk. Conversely, the 'thirst' of the alcoholic is clearly the result of habituation; yet in this case we *do* have 'genuine emotional engagement' and a powerful affective response.

This is not to say that there is not a genuine and important distinction to be made between cases in which affective response plays a key role in driving behaviour and cases in which it does not, or that the lumping together of these cases under the heading 'alief' does not serve to obscure this distinction. Nor is it to say that the related but different distinction between instinctive and habitual behaviour is itself not worth making, or is not similarly obscured by Gendler's alief/belief distinction. The point I want to make is just that Muller and Bashour end up repeating the very mistake they identify their opponent as having made: just as the cases Gendler lumps together under the heading 'alief' include cases in which emotional engagement/affective response is present and cases in which it is absent, so too do the cases Muller and Bashour invite us to lump together - both under the heading 'habit' and under the heading 'instinct'. And so the temptation is to dig down another level and acknowledge not two, but four meaningfully different types of mental phenomena where Gendler had just one: affective-response instincts and non-affective-response instincts, plus affective-response habits and non-affective-response habits.

I suggested in the introduction to this chapter that Gendler is attempting to lead the philosophy of mind down a rabbit hole - and here we find ourselves peering into it. The philosophy of mind has traditionally been in the business of commenting on folk and (to a lesser extent) scientific psychology, asking questions like: just what claims are being made when we ascribe mental states to subjects? In virtue of what are those claims true or false? What, if anything, are things like beliefs and desires supposed to be? What is the relationship between mental phenomena and physical phenomena? But Gendler invites us to stop musing from the sidelines and get stuck in to the business of psychology itself - describing 'how human minds really work' with reference to a proprietary philosopher's taxonomy of mental states.

In truth, though, this rather overstates the difference between what Gendler consciously sets out to do and what philosophers of mind have, perhaps less consciously, long found themselves doing. For the belief-desire psychology beloved of philosophers, with its emphasis on practical reasoning as the driver of human behaviour, is itself a proprietary theory built from certain raw materials furnished by folk psychological language and practice; folk psychology itself, as reflected in our everyday use of mentalistic language, is an altogether more baroque affair.

Consider this comment by Jennifer Nagel, made in the context of Gendler's discussion of alief:

> Perhaps in part because folk psychology makes belief attribution

> feel so natural to us, it is hard to come to see ourselves as driven
> by some messier packages of representational, affective and motor
> signals
>
> (Nagel J., 2012, p.786)

The suggestion here is that, *qua* folk psychologists, we naturally explain one another's behaviour primarily in terms of our beliefs; we assume that it's our beliefs (in conjunction with our desires, presumably) that drive our behaviour. And that leaves little room for the recognition of cases in which belief is wholly or partly overruled, or perhaps just bypassed, by things like arational affective responses to sensory inputs.

This suggestion is, in my view, quite false. *Qua* folk psychologists, whose understanding of the mind shapes and is shaped by our everyday use of mentalistic language, we do not have the slightest difficulty in recognizing and describing the operation of arational mental processes. Here, for instance, is a folk psychological explanation for the behaviour of the subject stepping on to the Skywalk:

> Her eyes tell her she's standing on thin air, so she can't shake the
> feeling that she's about to fall. Even though she knows deep down
> that the walkway's safe, her instinct is to walk with extra care or get
> off it altogether.

As for the subject who turns his nose up at the faeces-shaped fudge, well, *just the sight of it's enough to make him feel ill*. These explanations come naturally to a folk psychologist. It's the philosopher of mind, with his preference for practical-reasoning explanations of behaviour based on a stripped-down belief-desire psychology, who struggles to accommodate these cases and who is the target of Gendler's appeal for the inclusion of alief in his mental-state taxonomy.

It should come as no surprise that my own inclination, in responding to Gendler and to Muller and Bashour, is to counsel the abandonment of this restrictive philosopher's model of psychology not in favour of a revised version including alief, but in favour of a dispositionalist interpretation of folk psychology proper. But lest anyone should be tempted to enter the rabbit hole - to embark, that is, on the research program unwittingly opened up by Gendler and her critics, a program based around expanding the taxonomy of belief-desire psychology by identifying and incorporating alief-type states - I want to say why I think the prospects for success in this area are so poor.

# 9.4 Content and predictive/explanatory value

It is worth noting explicitly that when Gendler describes the rendering occurrent of aliefs, what she describes is the going through of a mental process: representations of one sort or another (e.g. visual impressions of something or thoughts about something) give rise to affective responses (like emotions) which give rise in turn to the activation of certain motor routines. On the face of it, processes and states are very different things. Nonetheless, Gendler thinks that we can bring these processes into the mental-state fold by defining a new state type - alief - that serves to *contain* that bundle of representations, affective responses and motor routines.

In section 9.3, we considered an important objection to that proposal: that a focus on aliefs, as defined by Gendler, tends to obscure important differences between distinct mental phenomena (for instance, between habit and instinct, or between cases in which affective responses are or are not relevant to the explanation of subjects' behaviour). However, this leaves open the possibility of tinkering with the concept of alief - perhaps identifying different types of alief, or defining new state types modelled after alief. I want to suggest that the concept of alief is so badly flawed that this would be a futile exercise.

The fundamental problem with Gendler's account, in my view, is that the concept of mental state *content* with which she is working is thoroughly obscure. Just what does it mean to say, for instance, that a dispositional 'alief has representational-affective-behavioral content that includes... the visual appearance as of a cliff, the feeling of fear and the motor routine of retreat' (Gendler, 2008a, p.641)? This state must, on Gendler's account, be possessed right now by billions of people who have never so much as seen the view from a cliff (since most of the people in that category would, I think it's reasonable to assume, nonetheless react to the Skywalk in much the same way as the subject described by Gendler). How is it, then, that they each possess a token mental state containing 'the visual appearance as of a cliff'? How can someone who is right now feeling no fear be said to possess a token mental state that contains 'the feeling of fear'? What is it for a mental state to contain a motor routine?

The idea can only be, I think, that for a subject to have a dispositional alief with content R-A-B *just is* for her to be disposed 'to entertain simultaneously R-ish thoughts, experience A, and engage in B' (that is, to alieve R-A-B occurrently) in response to certain stimuli. If so, however, Gendler's project is radically different from the project of a Standard View theorist who thinks that the content of my belief that tea tastes best with two sugars is stored in my

brain in something like the way that it might be written in a book or burned on to a disc, and that my tea-making behaviour can be explained and predicted on the basis that I have a belief with that representational content. Gendler does not - surely - think that *right now* there is a representation of the view from the Skywalk in my brain, just waiting to be activated, and that because we know that that representation is there (alongside an associatively linked affective response and motor routine), we can predict my behaviour upon stepping on to the Skywalk. Rather, she thinks that when the view from the Skywalk does force itself upon me, as I step out on to it, I can be expected to have a certain sort of affective response that in turn makes me more likely to behave in a certain way. What she is reaching for here appears, in fact, to be a broadly dispositionalist strategy for predicting and explaining my behaviour: to ascribe to me a dispositional alief with the content 'Really high up, long long way down. Not a safe place to be! Get off!!' is just to say something about the way I am disposed to react to certain stimuli.

Or perhaps I should say: it is to *attempt* to say something about the way I am disposed to react to certain stimuli. For if you are told that I dispositionally alieve 'Really high up, long long way down. Not a safe place to be! Get off!!', all you are entitled to infer is that there is at least one situation in which I would judge myself to be really high up, feel that this was not a safe place to be, and find myself inclined to get off whatever I was on at the time. Just what situation(s) that would happen in, though - who knows? It could be when I step on to the Skywalk, when I climb a ladder, when I sit on a fairground ride - or in all three of those situations, or in none of them. Similarly: what do we know about the likely behaviour of someone with the dispositional alief 'Filthy object! Contaminated! Stay away!'? To all intents and purposes, nothing at all. So in terms of predicting subjects' behaviour, the ascription to them of (dispositional) aliefs appears to be a thoroughly useless practice.

A possible reply to this objection might go as follows: the ascription of an alief such as 'Filthy object! Contaminated! Stay away!' might indeed be useless outside of a meaningful context. But suppose the fudge experiment is described to us and we're asked to predict the behaviour of experimental subjects. Might it not be useful to say: "well, assuming that they're disposed to alieve 'Filthy object! Contaminated! Stay away!' in those circumstances, I reckon they're likely to behave thus-and-so"? To this I would reply: indeed that would plausibly be a useful thing to do (for the same reason that it would plausibly be useful to explain subjects' behaviour after the fact by ascribing to them an occurrent alief with that content: in both cases, a claim is being made about the occurrence of

a mental process underlying subjects' behaviour). But that looks to me like an ascription to subjects of a disposition to *acquire* an alief, not an ascription of an alief subjects already have. It is analogous to the following prediction of subjects' behaviour in terms of belief: 'assuming that they're disposed to believe, upon seeing a lion in front of them, that there's a lion in front of them, I reckon they're likely to behave thus-and-so'. That is an ascription of a disposition to acquire a belief with the content 'there's a lion in front of me' - not an ascription of a belief with that content that subjects already have, only dispositionally rather than occurrently.

It's worth dwelling on the analogy with belief-based predictions and explanations of behaviour here. What *justifies* my ascription to a given subject of the disposition to acquire, upon seeing a lion in front of her, the belief that there is a lion in front of her? Well, this fits a general pattern. The subject almost always comes to believe that what she sees in front of her really is in front of her (outside of certain specific contexts, e.g. when she's looking at a television or a mirror). We might say that she believes the evidence of her visual sense can be trusted. Since she has *that* belief, of course, she'll acquire the belief that there is a lion in front of her if she sees a lion in front of her. If she also has the belief that lions are dangerous, then it's reasonable to expect that upon acquiring the belief that there is a lion in front of her she will acquire the belief that she is in danger, and feel afraid. Because that fits a pattern too - she almost always feels afraid when she believes she's in danger. And when she feels afraid of this sort of immediate physical danger, she almost always takes instinctive action to evade the threat. So the practice of belief ascription is about spotting patterns and using them to predict and explain the ways subjects act, think and feel in specific circumstances. The same goes for the practice of ascribing desires, and in my view - leaving behind the attenuated mental-state taxonomy of belief-desire psychology for a moment - for the practice of ascribing fears. prejudices, personality traits, and so on.

So when we see someone run screaming from a lion, and someone asks why she did that, our explanation of her behaviour does not have to begin and end with a description of the mental process that immediately preceded her running away - she believed there was a lion in front of her, felt afraid and initiated a 'run away' motor routine. Rather, it can draw on what we know about her dispositional profile in general - as expressed in mentalistic language such as 'she knows lions are dangerous,' 'she's got more sense than to try to fight an animal much stronger than her,' 'she's scared of anything with big teeth' and so on - and in so doing fit her behaviour into an intelligible pattern, reflecting and

perhaps enhancing our understanding of her as a person.

Compare that with the alief model of explanation. Suppose I explain that someone ran away from a lion because she occurrently alieved 'Lion! Danger! Run away!', and you press me for an explanation of *why* she went through that process. The obvious answer - the answer that Gendler's account appears to invite - is that she possessed that alief dispositionally and the appearance of the lion in front of her was what it took to render it occurrent. But this fails to make any kind of *sense* of the subject's behaviour. It amounts to nothing more than a trivially true ascription to her of the disposition to go through just the process she went through in just the circumstances that she went through it. It doesn't leave us feeling as if we know anything about what makes the subject tick in general. It doesn't leave us better able to see how the subject's behaviour on that occasion fits any more general pattern of behaviour, or reflects her personality or her attitudes.

The foregoing should not be read as a defence of the adequacy, in all cases, of belief-desire explanations of behaviour. Gendler is quite right that belief-based explanations of behaviour - 'She ran away because she believes lions are dangerous' - are a poor fit for many cases. But the way belief ascription works is better than the way alief ascription works, because it aims to tell us something more than just: 'in that particular circumstance, this particular response is what we can expect (or: was what we got)'. In fleshing out belief-desire psychology, then, what we need is to enhance our ability to ascribe general dispositions to subjects - to build a rounded picture of the patterns of behaviour (and thought, and affective responses) of particular subjects, so that we get an idea of how they'd react to finding themselves in a whole range of circumstances.

And as I noted in section 9.3, and hinted again above, this does not require us to cast around for hitherto undiscovered mental state types that can be integrated into folk psychology. It simply requires us to take advantage of the richness of folk psychological language as it is encountered in the wild. If we allow ourselves to talk about subjects' hopes and fears, prejudices and habits of thought, character traits and so on, as well as about their beliefs and desires, we can not only accommodate the sort of cases that so trouble Gendler, but integrate our understanding of those cases with our understanding of more everyday cases.

Take the Skywalk case again. Suppose I ask why that man hesitated to walk across the Skwalk and I'm told he alieved 'Really high up, long long way down. Not a safe place to be! Get off!!' Plausibly that leaves me with a rough idea of the mental process driving his behaviour on that occasion - but that's it. How-

ever, suppose I am instead told that the man has a fear of heights. Suppose it is then explained to me that the man, like most people, has a tendency to experience affective responses to visual stimuli based on mere appearances. Now I'm beginning to understand *why* he went through that process. (And it's not because he dispositionally alieved 'Really high up, long long way down. Not a safe place to be! Get off!!' an hour before he reached the Skywalk.) I can even start to predict how he'd feel and act when riding a Ferris wheel, how likely he'd be to go skydiving, or how he'd respond to seeing something he finds scary (or sexy, or sad, or whatever) depicted in a film. By building up a dispositional profile of this man, I begin to understand why he acted the way he did on that particular occasion, and how he can be expected to act (and think, and feel) on other occasions.

## 9.5 Conclusion

In the end, it is precisely in terms of 'serviceability' that alief falls down. The ascription of circumstance-specific dispositional aliefs is a positively useless exercise in terms of predicting and explaining subjects' behaviour compared to the building up of dispositional profiles through ordinary-language ascriptions of beliefs and desires, hopes and fears, prejudices and gut feelings, habits of thought and intellectual blind spots. And without *dispositional* aliefs - persistent states involving associative links between certain sorts of representations, affective responses and motor routines, and underlying subjects' reactions to particular stimuli - the language of *occurrent* aliefs seems to be nothing more than an eccentric form of shorthand for descriptions of certain mental processes.

The problem Gendler identifies is real enough: a psychological model that recognizes beliefs, desires and imaginings but not much else in the way of mental states just is not up to the task of predicting and explaining human behaviour, since it does not capture the realities of mental life. However, her response to this problem is, in my view, desperately misguided. The way to defend folk psychology against eliminativist arguments is not to refine it into a proprietary philosophical theory of mind, complete with 'new' state types to fill the causal-explanatory gaps left by the excision of all but a few canonical state types; it is to challenge the assumption driving those concerns, that is, the assumption that folk psychology is a theory about the functional organization of the mind.

# 10 Conclusion

In this thesis, I have tried to make the case that the failure of logical behaviourism does not entail the failure of dispositionalism. On the contrary, a liberal, holistic dispositionalism of the sort I call (following Schwitzgebel (2002)) *phenomenal dispositionalism* can avoid the pitfalls both of logical behaviourism on the one hand, and of the Standard View theories that have superseded it on the other.

It is uncontroversial, I think, that to ascribe things like propositional attitudes and character traits to subjects is to say something about how those subjects are likely to respond when they find themselves in a range of different circumstances. In that sense and to that degree, then, it is uncontroversial that many of our mental-state concepts are dispositional concepts. It is similarly uncontroversial that the ascription of these dispositional mental states to subjects plays a central role in everyday, folk psychological explanations and predictions of subjects' actions. On these points, I think the logical behaviourist, the Standard View theorist and the phenomenal dispositionalist can all agree.

It is when it comes to *accounting* for the evident success of mental-state ascription in everyday predictive-explanatory practice that these three traditions part company. The logical behaviourist argues that folk psychological practice depends for its success upon the one-at-a-time reducibility of sentences ascribing mental states to corresponding assertions that the subject of ascription exhibits, or *would* exhibit, such-and-such observable behaviour in such-and-such a range of circumstances. The Standard View theorist argues that folk psychological practice depends for its success upon a one-to-one correspondence between the particular mental states ascribed to subjects and particular, categorical, internal states of those subjects' brains, playing particular causal roles in mediating - together with other, similar states - between sensory inputs and behavioural outputs. The phenomenal dispositionalist argues that folk psychological practice depends for its success upon our ability to use everyday mentalistic language to build up pictures of subjects' dispositional profiles and so to understand both how they might be expected to act, think and feel in a range of circumstances, and how their actions, thoughts and feelings on

particular occasions fit into broadly intelligible patterns.

The implausibility of the logical behaviourist position has long been evident. I have tried to undermine the plausibility of the Standard View position by attacking it at its root: a line of thinking according to which the success of disposition ascription in general as an explanatory-predictive practice - not just dispositional mental state ascription - depends upon a one-to-one correspondence between the particular dispositional properties ascribed to objects and particular, categorical, intrinsic properties of those objects, playing particular causal roles in bringing about certain manifestation events in response to certain stimuli. I have argued that the ascription of dispositional properties just does not, and cannot, work like this. The claims that sugar is disposed to dissolve in water in circumstances C and that water is disposed to dissolve sugar in circumstances C are *precisely the same claim*: the claim that in circumstances C, the mixing of sugar and water would result in the formation of a sugar-water solution. The causal relevance of certain particular, intrinsic, categorical properties (of sugar on the one hand, and of water on the other) to the production of this manifestation event is not in doubt; but the idea that these properties correspond one-to-one with certain dispositional properties (of sugar on the one hand, and of water on the other) is just untenable. For the purposes of explanatorily and predictively successful disposition ascription, it doesn't matter what categorical properties of what objects play what causal roles in the production of manifestation events; all that matters is that something, somewhere does the causal work required to bring those events about in the right circumstances.

So if we are concerned to preserve our everyday folk-psychological taxonomy of mental states, we need not wait anxiously for neuroscientists to come up with a taxonomy of brain states that maps neatly onto it. Nor need we reassure ourselves that such a brain-state taxonomy just *must* exist, since our everyday mental-state taxonomy reflects a theory about the functional organization of the mind that just cannot be wrong. Nor, finally, need we set about revising our everyday mental-state taxonomy - as Gendler (2008a) suggests - in order to make up for perceived shortcomings.

All we need do in order to preserve our everyday mental-state taxonomy, in fact, is to look around and see that people's actions, thoughts and feelings really do fit into intelligible patterns and that our everyday language of mental states allows us to describe these patterns. Discoveries in neuroscience will no doubt shed light on the particular roles played by particular brain states in particular mental processes, just as discoveries in chemistry have shed light

on the particular roles played by particular categorical states of substances like sugar and water in particular chemical processes; but these discoveries represent no threat whatsoever to the truth of claims made about subjects' dispositional profiles in the ordinary language of mental states.

I should say what I mean when I talk about our 'everyday mental-state taxonomy'. It seems to me that there is a considerable but barely acknowledged degree of tension between two influential lines of thinking in the philosophy of mind. On the one hand, we have the line of thinking according to which everyday mental-state language is a reliable guide to a powerful and true theory about the functional organization of the mind - a theory wielded by every user of that language and usually called 'folk psychology'. On the other hand, we have the line of thinking - rarely made explicit, but everywhere evident in the philosophy of mind literature - according to which much of our everyday mentalistic discourse is little more than 'noise' that needs to be tuned out, for the purposes of serious inquiry, so that we can focus on a handful of reputable, canonical state types (belief and desire foremost among them) that can be neatly defined in terms of their role in cognitive processes (and especially practical reasoning). Gendler is to be commended for daring to state what ought to be obvious: that the taxonomically austere belief-desire psychology beloved of philosophers is woefully inadequate to the task of expressing our common-sense understanding of the role in human psychology of many phenomena that don't fit a neat practical-reasoning model: habit, instinct, affect and so on. But the alief-belief-desire psychology she recommends is only slightly less austere and only slightly more adequate to the task. In my view all this austerity is needless; our everyday mental-state taxonomy, consisting of all the mental state types users of everyday mentalistic language find it useful to cite in order to capture the complexities and nuances of subjects' dispositional profiles, should be embraced and defended *in toto*. Since the phenomenal dispositionalist does not endorse the reification of mental states and does not consider everyday mentalistic language to express a theory about the functional organization of a causal system, this is a matter of linguistic sophistication and not of ontological extravagance or theoretical complexity.

In my view, then, the philosopher of mind should leave the psychologist and the neuroscientist to do their work without worrying about whether their cognitive-state or brain-state taxonomies will map on to the mental-state taxonomy of folk psychology, and without presuming to alert them to the existence of hitherto unrecognized state types discovered through a process of philosophical reflection. Similarly, she should respect the use made of mental-state concepts

by true *folk* psychologists - ordinary language users - and resist the temptation to distil from their abundant linguistic raw materials a more austere and philosophically respectable theory of mind. Ordinary-language psychology, with its emphasis on understanding subjects' dispositional profiles rather than their functional organization, is perfectly respectable just as it is.

# Bibliography

Alston, W.P. 1989. *Divine Nature and Human Language: Essays in Philosophical Theology*. Ithaca, New York: Cornell University Press.

Anonymous [no date]. *Why does sugar dissolve in water?* [Online]. [Accessed on 19 March 2014]. Available from: http://answers.yahoo.com/

Aquinas, T. 1999. Whether There is a God. In: Helm, P. ed. *Faith & Reason*. New York and Oxford: Oxford University Press, pp. 101-106.

Aquinas, T. 2008. *Summa Theologica*. [Online]. [Accessed 19 March 2014]. Available from: http://www.newadvent.org/

Armstrong, D.M. 1981. *The Nature of Mind*. Brisbane: University of Queensland Press.

Armstrong, D.M. 1993. *A Materialist Theory of the Mind*. London and New York: Routledge.

Baker, L.R. 1995. *Explaining Attitudes: A Practical Approach to the Mind*. Cambridge: Cambridge University Press.

Baker, L.R. 2001a. Practical Realism Defended: Replies to Critics. In: Meijers, A. ed. *Explaining Beliefs: Lynne Rudder Baker and Her Critics*. Stanford: CSLI Publications, pp.183-217.

Baker, L.R. 2001b. Practical Realism Defended: Replies to Critics. [Online]. [Accessed on 21 March 2014]. Available from: http://people.umass.edu/

Blackburn, S. 1991. Losing Your Mind: Physics, Identity, and Folk Burglar Prevention. In: Greenwood, J. D. ed. *The Future of Folk Psychology*. Cambridge: Cambridge University Press, pp.196-225.

Block, N. 1981. Psychologism and Behaviourism. *The Philosophical Review*. **90**(1): pp.5–43.

Block, N. 1991. Troubles With Functionalism. In: Rosenthal, D.M. ed. *The Nature of Mind*. New York and Oxford: Oxford University Press, pp. 211-228.

Carnap, R. 2002. Psychology in Physical Language. In: Chalmers, D.J. ed. *Philosophy of mind: classical and contemporary readings*. New York and Oxford: Oxford University Press, pp.39-44.

*Bibliography*

Chisholm, R.M. 1957. *Perceiving*. Ithaca: Cornell.

Choi, S. 2006. The Simple vs. Reformed Conditional Analysis of Dispositions. *Synthese*. **148**(2), pp.369–379.

Choi, S. 2008. Dispositional Properties and Counterfactual Conditionals. *Mind*. **117**(468), pp.795–841.

Choi, S. and Fara, M. 2014 Forthcoming. Dispositions. In: Zalta, E.N. ed. *The Stanford Encyclopedia of Philosophy*. Spring 2014 Edition. [Online]. [Accessed on 19 March 2014]. Available from: http://plato.stanford.edu/

Churchland, P.M. 1981. Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*. **78**(2), pp.67-90.

Davidson, D. 1973. Radical Interpretation. *Dialectica*. **27**(1), pp.314-28.

Davidson, D. 1987. Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*. **60**(3), pp.441-458.

Dennett, D. 1978. *Brainstorms*. Cambridge, MA: MIT Press.

Dennett, D.C. 2002. True Believers: The Intentional Strategy and Why It Works. In: Chalmers, D.J. ed. *Philosophy of mind: classical and contemporary readings*. New York and Oxford: Oxford University Press, pp.556-568.

Descartes, R. 1978. *Descartes: His Moral Philosophy and Psychology*. New York: New York University Press.

Descartes, R. 1996. *Meditations on First Philosophy.* Revised Edition. Cambridge: Cambridge University Press.

Ellis, H.D and Young, A.W. 1990. Accounting for delusional misidentifications. *The British Journal of Psychiatry*. **157**(2), pp.239-48.

Elugardo, R. 2001. Brain States, Causal Explanation, and the Attitudes. In: Meijers, A. ed. *Explaining Beliefs: Lynne Rudder Baker and Her Critics*. Stanford: CSLI Publications, pp.91-117.

Fara, M. 2005. Dispositions and Habituals. *Noûs* 39: 43–82

Garcia, L. 1992. Divine Freedom and Creation. *The Philosophical Quarterly*. **42**(167), pp.191-213.

Gazzaniga, M.S. 2000. Cerebral specialization and interhemispheric communication: does the corpus callosum enable the human condition? *Brain* **123**(7), pp.1293-1326.

Geach, P. 1957. *Mental Acts: Their Content and Their Objects*. London: Routledge & Keegan Paul.

Gendler, T.S. 2008a. Alief and Belief. *Journal of Philosophy*. **105**(10), pp.634-663.

Gendler, T.S. 2008b. Alief in Action (and Reaction). *Mind and Language*. **23**(5), pp.552-585.

Goldman, A.I. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Gordon, R. 1986. Folk Psychology as Simulation. *Mind and Language*. **1**(2), pp.158–171.

Gundersen, L. 2002. In Defence of the Conditional Account of Dispositions. *Synthese.* **130**(3), pp.389–411.

Harman, G. 1982. Conceptual Role Semantics. *Notre Dame Journal of Formal Logic.* **23**(2), pp. 242-256.

Hasker, W. 2004. *Providence, Evil and the Openness of God*. London: Routledge.

Hume, D. 1998. *Dialogues Concerning Natural Religion*. Second Edition. Indianapolis and Cambridge: Hackett Publishing Company.

Jackson, F. and Pettit, P. 1988. Functionalism and Broad Content. *Mind.* **97**(387), pp.381-400.

Jackson, F. and Pettit, P. 1990. Program Explanation: A General Perspective. *Analysis.* **50**(2), pp. 107-117.

Johnson, P.R. 2000. *God Without Mood Swings*. [Online]. [Accessed on 19 March 2014]. Available from: http://www.spurgeon.org/

Johnston, M. 1992. How to Speak of the Colors. *Philosophical Studies.* **68**(3), pp.221–263.

Kant, I. 1993. *Critique of Pure Reason*. London: Everyman.

Kriegel, U. 2012. Moral Motivation, Moral Phenomenology, and the Alief/Belief Distinction. *Australasian Journal of Philosophy*. **90**(3), pp.469-486.

Levin, J. 2013. Functionalism. In: Zalta, E.N. ed. *The Stanford Encyclopedia of Philosophy*. Fall 2013 Edition. [Online]. [Accessed on 19 March 2014]. Available from: http://plato.stanford.edu/

Lewis, D. 1980. Mad Pain and Martian Pain. In: Block, N. ed. *Readings in the philosophy of psychology, vol. 1*. Cambridge, MA: Harvard, pp.216–222.

Lewis, D. 1997. Finkish Dispositions. *The Philosophical Quarterly*. **47**(187), pp.143–158.

*Bibliography*

Lewis, D. 2002. Psychophysical and Theoretical Identifications. In: Chalmers, D.J. ed. *Philosophy of mind: classical and contemporary readings*. New York and Oxford: Oxford University Press, pp.88-94.

Locke, J. 1997. *An Essay Concerning Human Understanding*. London: Penguin.

Logue, H. 2010. *Email to Gregory Owens*, date not recorded.

Maimonides, M. 1999. Language and Creation. In: Helm, P. ed. *Faith & Reason*. New York and Oxford: Oxford University Press, pp. 92-98.

Manley, D. and Wasserman, R. 2008. On Linking Dispositions and Conditionals. *Mind*. **117**(465), pp.59–84.

McGinn, C. 1993. *Problems in Philosophy: the Limits of Inquiry*. Oxford, UK and Cambridge, USA: Blackwell.

Mellor, D.H. 1995. *The Facts of Causation*. London: Routledge Press.

Mellor, D.H. 2000. The Semantics and Ontology of Dispositions. *Mind*. **109**(436), pp.757–780.

Mill, J.S. 1999. The Infinite Goodness of God. In: Helm, P. ed. *Faith & Reason*. New York and Oxford: Oxford University Press, pp. 249-250.

Muller, H. and Bashour, B. 2011. Why Alief is Not a Legitimate Psychological Category. *Journal of Philosophical Research*. **36**, pp.371-389.

Mumford, S. 1998. *Dispositions*. Oxford: Oxford University Press.

Nagel, J. 2012. Gendler on Alief. *Analysis*. **72**(4), pp.774-788.

Nagel, T. 1974. What Is It Like to Be a Bat? *The Philosophical Review*. **83**(4), pp.435-450.

Pink, T. 2004. *Free Will: A Very Short Introduction*. Oxford: Oxford University Press.

Prior, E. 1985. *Dispositions*. Aberdeen: Aberdeen University Press.

Putnam, H. 1963. Brains and Behavior. In: Butler, R.J. ed. *Analytical Philosophy, Second Series*. Oxford: Basil Blackwell, pp.211-235.

Rowe, W. 2004. *Can God Be Free?* Oxford: Oxford University Press.

Ryle, G. 2000. *The Concept of Mind*. London: Penguin Classics.

Schwitzgebel, E. 2001. In-Between Believing. *Philosophical Quarterly*. **51**(202), pp.76-82.

Schwitzgebel, E. 2002. A Phenomenal, Dispositional Account of Belief. *Noûs*. **36**(2), pp.249-275.

Schwitzgebel, E. 2011. *Email to Gregory Owens*, 18 October.

Schwitzgebel, E. 2013. A Dispositional Approach to Attitudes: Thinking Outside of the Belief Box. In: Nottelmann, N. ed. *New Essays on Belief: Constitution, Content and Structure*. Basingstoke: Palgrave Macmillan, pp.75-99.

Schwitzgebel, E. 2014 Forthcoming. Belief. In: Zalta, E.N. ed. *The Stanford Encyclopedia of Philosophy*. Spring 2014 Edition. [Online]. [Accessed on 19 March 2014]. Available from: http://plato.stanford.edu/

Skinner, B.F. 1953. *Science and Human Behaviour*. New York: Macmillan.

Soskice, J.M. 1987. *Metaphor and Religious Language*. Oxford: Clarendon Press.

Steward, H. 1997. *The Ontology of Mind: Events, Processes and States*. Oxford: Oxford University Press.

Steward, H. 2010. *Conversation with Gregory Owens*, date not recorded.

Stich, S. 1983. *From Folk Psychology to Cognitive Science*. Cambridge, MA: Bradford Books/MIT Press.

Stout, R. 2006. *The Inner Life of a Rational Agent: In Defence of Philosophical Behaviourism*. Edinburgh: University of Edinburgh Press.

Strawson, G. 2008. The Identity of the Categorical and the Dispositional. *Analysis*. **68**(4), pp.271-82.

Swinburne, R. 1992. *Revelation: From Metaphor to Analogy*. Oxford: Clarendon Press.

Tanney, J. 2009a. Rethinking Ryle: A Critical Discussion of The Concept of Mind. In: Ryle, G. *The Concept of Mind*. 60th anniversary edition. London: Routledge, ix-lvii.

Tanney, J. 2009b. Gilbert Ryle. In: Zalta E.N. ed. *The Stanford Encyclopedia of Philosophy*. Winter 2009 Edition. [Online]. [Accessed on 19 March 2014]. Available from: http://plato.stanford.edu/

Watson, J.B. 1913. Psychology as the Behaviourist Views It. *Psychological Review*. **20**(2), pp.158-177.

Watson, J.B. 1930. *Behaviourism*. New York: W. W. Norton and Company, Inc.

vos Savant, M. [no date]. *Game Show Problem*. [Online]. [Accessed on 19 March 2014]. Available from: http://marilynvossavant.com/