THE UNIVERSITY
of EDINBURGH

Internalism and the Explanatory Role of Narrow Content

Sam Baird

PhD Philosophy

The University of Edinburgh

2015

Abstract

The central problem addressed by this thesis is how narrow content can be genuinely representational and relate to the individuation of mental representations. A second problem relates to the explanatory role of narrow content. The challenge facing the proponent of narrow content is that externalism is the standard position with regard to representation, individuation, and explanation, and it is often held that narrow content is either incoherent or explanatorily redundant. To this end, I defend a version of a two-component theory of content which accepts that wide content is relatively uncontroversial but holds that narrow content has an important role to play.

The thesis is divided into two main parts. In part one I focus on the problem of representation and the individuation of mental representations. In Chapter 1 I consider two of the main accounts of wide content from Putnam (1975) and Fodor (1994) and show that in each case there is a notion of narrow content that is compatible with the externalist approach to representation. In particular, I discuss versions of narrow content associated with two-dimensionalism (Chalmers 2003; Jackson 2003a) and radical internalism (Crane 1991; Segal 2000), among other views (e.g. Prinz 2000; Kriegel 2008), and highlight important objections to these existing accounts of narrow content. In response to these problems, I defend a pluralistic view of narrow content according to which there are two main forms of narrow representation, and contend that this account can be systematically supported by understanding the supervenience claim made by the internalist, the kinds of properties represented by mental representations, and the nature of the underlying mental representations that bear narrow content. I then show that this account of narrow content has the resources to respond to a common objection to narrow content (Lepore & Loewer 1986; Block & Stalnaker 1999). In support of this approach to narrow content I show in Chapter 2 that internalism is consistent with

recent work on concepts, and in particular with concept pluralism (Laurence & Margolis 1999; Weiskopf 2009) – the view that concepts are comprised of a variety of conceptual structures. I argue that such conceptual structures can be coherently individuated by the narrow contents outlined in Chapter 1, giving rise to an internalist version of concept pluralism. I support the internalist construal of concept pluralism by identifying problems with externalist versions (Laurence & Margolis 1999) and responding to objections to narrow content (Weiskopf 2007). In Chapter 3 I develop an extended critique of social externalism (Burge 1979; Goldberg 2002), and accounts of perceptual psychology (Egan 2009; Burge 2010), both of which purport to have anti-internalist consequences for the individuation and nature of mental representations and provide a challenge to the account of narrow content and mental representations developed in Chapters 1 and 2.

In part two, I defend the explanatory role of narrow content. In Chapter 4 I contend that the truth conditions of propositional attitude ascriptions may relate to the narrow contents expressed by attitude ascribing sentences. To show this, I defend a Fregean theory of attitude ascriptions and consider two objections (Soames 2002; Travis 2008) to Fregean accounts that would block the relevant semantic role of narrow content. In addition, I show how this account provides a basis for rejecting two externalist arguments (Putnam 1975; Burge 1979) that rely on assumptions about the nature and role of attitude ascriptions. I contrast this account with existing internalist accounts (Segal 2007; Chalmers 2007), and show how it provides a response to Soames' (2002) Perfect Earth objection. In Chapter 5 I critically assess two existing accounts of the role of narrow content in psychological explanations of behaviour (Jackson & Pettit 1988; Fodor 1991), focusing on objections raised in connection with proximal causes (Burge 1995) and explanatory generality and causal relevance (Yablo 2003). In response to these objections, I develop an alternative approach which appeals to the rational role of reasons (Davidson 1963; McDowell 2006), and contend that only narrow content can provide a suitably rational explanation of behaviour. I show how this provides a response to the objections raised, and how the appeal to reasons provides a basis for rejecting Fodor's (1994) more recent externalist account of behaviour.

Contents

Declaration

I hereby declare that this thesis is entirely my own work, except where indicated by quotation, reference, and acknowledgement. The work has not been submitted for any other degree or professional qualification.

Signed: Date:

Acknowledgements

I would like to thank Jesper Kallestrup, my primary supervisor, for his support throughout my time at Edinburgh. His work in this area served as a key starting point and a reference throughout. The thesis has benefited greatly from the many discussions we had, and the drafts of chapters that Jesper patiently read through and provided feedback on. Thanks also to Mark Sprevak who agreed to be my secondary supervisor part way through the project.

I am indebted to my family for the support they have provided me over the past few years. Val and Dugald provided crucial financial support for which I am extremely grateful. I thank Fei for keeping me grounded throughout the project. Special thanks go to my parents, June, who provided much needed encouragement, and Neil, to whose memory this thesis is dedicated.

Introduction

In this thesis I develop and defend an account of narrow content and its explanatory role. There are two questions that a proponent of narrow content must address. First, is whether there is a viable notion of narrow content. The varieties of narrow content in the literature have been subject to much criticism, thus a central challenge facing the proponent of narrow content is to articulate an account that withstands the sorts of objections raised. Second, is to say exactly why narrow content is needed, if, as is commonly held, wide content is relatively uncontroversial. Externalism is often seen as the default position in the philosophy of mind and language, given the general acceptance of the main conclusions stemming from Kripke (1980) and Putnam (1975) on names and natural kind term meaning, as well as from naturalised accounts of representation from Dretske (1981) and Fodor (1994) (among others). In addition, while many of the insights of externalism are taken on board by proponents of narrow content, externalism is often taken to show that there is no plausible basis for narrow content, and/or that narrow content plays no significant explanatory role in philosophy and psychology since wide content alone is sufficient. Despite the current externalist orthodoxy, many proponents of narrow content aim to show that narrow content is both viable and explanatorily central to philosophy and psychology. The version of narrow content I defend is motivated in part by the problems facing existing accounts, and is thus intended to contribute to the rehabilitation of the internalist tradition by developing a novel approach to both the nature and explanatory role of narrow content that avoids the most common objections in the literature.

There are nearly as many accounts of narrow content as there are proponents, all of which make different claims about the form of representation or representational framework, the properties that are represented, and the basis of such content in

mental representations. Three approaches exemplify the range of different stances in the literature. The first approach aims to reject externalism and wide content entirely, thereby arguing that all mental content is narrow (for instance Crane 1991; Segal 2000; Farkas 2008b). Such an approach takes on the unnecessary burden of having to show that all varieties of wide content fail to relate fundamentally to mental nature. This is not my view. A second approach seeks to identify a special, limited, notion of narrow content that accounts for certain phenomena, allowing wide content to account for others. On this approach, wide content is usually of central importance for representation and explanation, however, narrow content is sometimes understood as central to understanding the nature of mind, especially in terms of reflecting the subject's way of representing the world (Dennett 1987; Prinz 2000; Loar 2003). A worry with this account is that it fails to appreciate significant insights from the internalist literature, especially with regard to the broader representational nature of narrow content. Furthermore, the foundations of these views are often contentious, for instance Prinz (2000; 2002; 2005) relies on a radically empiricist theory of concepts, and Loar (2003) appeals to the notion of phenomenal intentionality, and both notions face substantial criticism. A third approach contends that narrow content is full-blown representational content and central to psychological explanations, whilst also endorsing wide content. The most prominent version of this approach is associated with two-dimensional semantics (Chalmers 2003; Jackson 2003a). However, again, this view is subject to an extensive critical literature, with a number of criticisms of Chalmers' framework.

Given these problems, there is scope for developing an alternative, comparatively ambitious, account of narrow content that doesn't rely on the contentious positions outlined above. To this end, I argue for a pluralistic account of narrow content, according to which narrow content relates to a variety of forms of representation, and contend that we can understand narrow content as relating to different kinds of mental representations as understood in the psychological literature on concepts. In particular, I show that narrow content is consistent with concept pluralism, which is the thesis that concepts are comprised of a variety of conceptual structures as discussed in the psychological literature. This comprises the topic of part one of this

thesis, which develops and defends the account of internalism and narrow content. In Chapter 1 I develop the pluralistic account of narrow representation and narrow content in detail. I show how this view avoids a common objection to narrow content, and I highlight the key assumption relating to pluralism about mental representation. This sets up Chapter 2 which focuses on the account of concepts that supports pluralistic narrow content. Drawing on the philosophical and psychological work on concept pluralism, I show that concepts may be individuated by the kinds of narrow content developed in Chapter 1. Chapter 3 rounds-out the account of narrow content by defending it against two approaches to the individuation of mental representations that purport to have anti-internalist consequences.

A strategy pursued by some externalists, most notably Fodor (1994), is to develop a thoroughgoing externalism by arguing that wide content alone is sufficient for accounting for both mental representation and psychological explanations that draw on mental content. The central claim is that wide content is all the content that psychology needs to recognise, and hence there is no need for narrow content in psychology. Furthermore, other externalists (for instance Soames 2002) argue that narrow content plays no semantic role in ascribing propositional attitudes like belief, which relate to the nature of the mental representations of the ascribee. In light of this, in the second part of this thesis I aim to show how the account of narrow content I defend in part one is also central to psychological explanation, and, in addition, plays an important semantic role in propositional attitude ascriptions. With regard to the semantic role of narrow content, in Chapter 4 I show how narrow content plays a role in the semantics of propositional attitude ascriptions in so far as the ascription of narrow content is consistent with a Fregean theory of attitude ascriptions. Narrow content is often taken to play such a role, however, I show how my account has important advantages over two exiting accounts. Finally, in Chapter 5 I develop a novel account of the role of narrow content in psychological explanations of behaviour. The view provides a way to avoid objections that face existing internalist accounts which rely on the explanatory role of narrow states, rather than narrow content. The view I defend appeals to the role of reasons in explaining behaviour, and I make the case that only narrow content of the sort developed in part one can

play this role. I then show how this view provides a response to the objections raised against existing internalist accounts.

In showing that narrow content is fully representational, relates to the individuation of mental representations, and plays an explanatory role in propositional attitude ascriptions and psychological explanations of behaviour, this thesis provides a partial rebuttal to the kind of univocally externalist approach of Fodor and others. However, the aim is not to reject wide content, but to address the imbalance within current philosophy of mind and make the case that narrow content has an important role to play within philosophy and psychology.

Chapter 1

Representation and Narrow Content

1. Introduction

It is common to hold that mental content is wide, that is, that the content of thoughts supervenes on external factors such as the social and physical environment of the thinker. Whether thoughts also have narrow content is a contested issue. Among proponents of narrow content there is a wide variety of proposals, many of which involve fundamental points of disagreement. Furthermore, there are both internalists and externalists who deny that it provides a genuine form of representational content, and contend that the role of narrow content should be restricted to cognitive content. To this end, there are two main aims of this chapter: to show why narrow content is genuinely representational and should be accepted by proponents of wide content, and to develop a robust account of the nature of narrow content. The chapter is structured as follows. In Section 2 I provide an outline of the wide/narrow content debate, and a critical review of existing accounts of wide and narrow content. I consider two main forms of externalism and show that in each case there is a kind of narrow content available to the internalist. I begin by considering Putnam's (1975) Twin Earth scenario and the case for wide content based on natural kind terms. I then consider Fodor's (1994) account of wide informational content. In both cases I identify notions of narrow content that are consistent with the general approaches to representation, and highlight key limitations with, and objections to, these accounts of narrow content. This discussion sets up the basis for developing an alternative approach to narrow content. In Section 3 I develop and defend a pluralist account in which narrow content represents a range of properties, through different representational mechanisms, and which relates to different kinds of mental representations. In Section 4, I consider a common objection to narrow content

which holds that there is no coherent way to identify or express narrow content, hence that narrow content is fundamentally incoherent. In response, I show that the objection can be avoided by appealing to the account developed in Section 3.

2. Externalism, internalism and representational content

In Section 2.1 I introduce the notions of mental representation and representational content, and the distinction between wide and narrow content.

2.1 Mental representation and the wide/narrow content debate

Mental representations represent such things as individuals (such as my pet cat), and natural objects (like water). The representational content of a mental representation is typically understood in terms of what the mental representation represents or is about, so the representational content of a mental representation like CAT would be cats, and the content of the mental representation WATER would be water. In addition to representational content, mental representations are often taken to have cognitive content that reflects how things are represented as being. The representational content of a thought, such as the thought expressed by the sentence 'cats like water', is composed of the contents of each of the mental representations expressed by an utterance of that sentence. Such thoughts are semantically evaluable for truth and falsity, so the representational content can be understood as being truth conditional content. We can express the truth condition of a thought expressed by 'cats like water', as 'cats like water' is true if and only if cats like water. The truth conditions tell us how the world must be for the thought to be true. Therefore, we need to know what the terms 'cat' and 'water' (or the representations CAT and WATER) represent in order to determine if the thought is true or false.

On one view of representational content, mental representations for natural objects like water, or artefacts like chairs, represent properties, for instance the property associated with being water or the property associated with being a chair. It is possible to frame this view within the possible worlds representational framework. On this approach, representational content is understood in terms of the properties represented at possible worlds, so that the content of the representation WATER is

understood in terms of a property at a relevant possible world or set of possible worlds[1]. Truth conditions for thoughts can also be understood in terms of possible worlds. For instance, the truth conditional content of 'cats like water' can be captured in terms of the set of possible worlds where cats like water.

There is an extensive literature on mental representations and representational content. An important question in this area relates to the nature of representation, such as why CAT represents cats. Other questions relate to the nature of mental representations. Representations like CAT are typically understood in terms of concepts, and there is a debate about the nature of concepts in the philosophical and psychological literature. In Chapter 2 I will look at the nature of mental representations in more detail. In this chapter I am concerned with the nature of representation and the individuation of representational content. In particular, I am concerned with whether representational content is wide, narrow, or perhaps a combination of both (more on these notions below). While there is a near-consensus that mental representations have wide content, there is some debate about whether they also have narrow content, and what such content is like. Thus, I shall be concerned with two issues: motivation for narrow content, and the question of its proper nature. But first, we need to be clear about what wide and narrow contents are.

Externalists about mental representation contend that the content of representations are determined, at least in part, by certain external, typically causal, conditions relating the subject to what is represented. Consequently, the content of a representation is wide because it is said to supervene, at least in part, on facts about the subject's environment. Supervenience is here understood as a relation that holds between two kinds of properties. If A supervenes on B, then any change to A must correspond to a change in B. According to externalists, representational content can

---

[1] There are other ways of understanding content, such as in terms of extensions at possible worlds, so that 'water' would represent the actual instances of water at a possible world. However, I will presume a property-view of representational content as this allows for the distinction between representing, say, the property H_2O and representing particular instances of H_2O. For instance, if S's mental representation expressed by 'H_2O' represents the actual instances of H_2O, this doesn't allow for a distinction between a general thought of H_2O, and particular thoughts of instances of H_2O, since in both cases the content would be the same, i.e. the set comprising all instances of H_2O.

change even while holding intrinsic properties of subjects constant. Thus, content does not supervene on intrinsic properties (the significance of this claim will be shown below).

Internalism is typically understood to be the thesis that content supervenes on intrinsic properties of subjects. Thus, internalists hold that intrinsic duplicates, in radically different environments, may possess a representation with the exact same content. When intrinsic properties are fixed, the content will also be fixed (subject to further caveats to be outlined below). Internalists therefore hold that changes in content will relate to changes in intrinsic properties, whilst externalists hold that content can change even holding intrinsic properties constant.

Given this characterisation of the debate, the guiding question to be addressed in the following sections is whether there is a viable notion of narrow representational content. In Sections 2.2 and 2.3 I consider two of the main externalist forms of representation from Putnam (1975) and Fodor (1994) that serve to motivate the notion of wide content. I then show that in each case there is a viable notion of narrow content and hence that there is a strong case for endorsing a two component theory of content. However, I also show that there are problems with existing accounts of each variety of narrow content. In particular, I highlight specific problems with accounts from Segal (2000), Prinz (2000), Chalmers (2003), and Jackson (2003a). In addition, I show that there are broadly two distinct accounts of narrow content in the literature, and that we have reason to accept versions of both. I then outline, in Section 2.4, how this critical assessment of narrow content provides the basis for a pluralistic form of narrow representation.

2.2 Putnam on natural kind representation

The first basis for externalism about representational content I will look at is Putnam's (1975) Twin Earth thought experiment. The thought experiment relies on the notion of natural kind terms that are expressions in natural language that aim to

represent natural kinds[2]. Natural kinds are usually understood as the kinds in nature of interest in the sciences (such as water, tigers, and gold) distinguished by their underlying microphysical, chemical, and biological properties. A central tenet of Putnam's thesis is that 'water' is a natural kind term and functions to pick out H_2O exclusively. Turning to the Twin Earth scenario, we are to consider Earth and a superficially identical planet Twin Earth, the only difference being that on Twin Earth the stuff they call 'water' is not H_2O and is instead composed of a novel chemical kind abbreviated to 'XYZ'. We are then to consider two intrinsically identical subjects, Oscar on Earth and Twin Oscar on Twin Earth, and the meaning and reference of the term 'water'. By assumption, Oscar and Twin Oscar are ignorant of the chemical composition of the substances in their respective environments. Furthermore, Putnam holds that Oscar and Twin Oscar may share all their psychological states, for instance, they might both think of the different substances in terms of a general kind of watery stuff, where 'watery stuff' is shorthand for the superficial properties common to H_2O and XYZ, such as being a clear, thirst-quenching, liquid that falls from the sky and fills the lakes and oceans. Despite this, Putnam holds that the meaning of 'water' as used by Oscar and Twin Oscar will be different. The difference in meaning is taken to result from the fact that on Earth Oscar refers to H_2O, so that the meaning of Oscar's term 'water' relates to H_2O, and not XYZ. On Twin Earth, Twin Oscar refers to H_2O, thus the meaning of Twin Oscar's term 'water' relates to XYZ, not H_2O. One consequence of this is that speakers of English must use the expression 'twin water' to talk about XYZ, since 'water' refers to H2O. This results in an externalist account of natural kind term meaning since Oscar and Twin Oscar are intrinsically identical, and may share their psychological states, but the representational content of the term 'water' when used by each is different. Consequently, the representational content of natural kind terms fails to supervene on intrinsic properties.

Although Putnam accepted that Oscar's and Twin Oscar's psychological states were identical, and hence the thesis was restricted to the reference and meaning of the

---

[2] Whether natural language has such terms is an important question. If not, then much of Putnam's position could be rejected. I consider this point below.

natural language term 'water', it is possible to run the same argument for mental representations and representational content. Thus, on this externalist view, Oscar's mental representation WATER is about H_2O and Twin Oscar's is about XYZ. The contents of their representations are wide since content fails to supervene on intrinsic properties, that is, the content of the mental representation WATER can change whilst holding fixed the subject's intrinsic properties. Properly speaking, we should distinguish Oscar's representation WATER from Twin Oscar's representation TWIN WATER. Consequently, internalism is held to be false where the thesis is understood as being that representational content supervenes on intrinsic properties, since the content of Oscar's and Twin Oscar's mental representations are different, despite them being intrinsic duplicates.

An important assumption here is that differences in the natural kinds present on Earth and Twin Earth entail that Oscar and Twin Oscar don't share a kind of mental representation or any representational contents. Putnam's account of the meaning of natural kind terms provides a way of supporting this representational difference between intrinsic duplicates (again, a similar point could be made for mental representation). Putnam intended to show that Fregean descriptivism about meaning and reference was incorrect. According to Frege (2000), expressions have both sense and reference. Sense relates to the mode of presentation of the referent, which provides a rule for determining reference or provides a condition that something must satisfy to be the referent. On the Fregean view, the sense of the expression 'Hesperus', given by 'the evening star', determines its referent (the planet Venus). Two expressions can have different modes of presentation but the same referent (for instance the mode of presentation associated with 'Phosphorus' is given by 'the morning star', while the referent is also the planet Venus). Through the Twin Earth scenario Putnam intended to show that meaning (sense) alone cannot determine reference. To see this, note that the sense of 'water' might be given by 'the clear thirst-quenching liquid'. However, since 'water' refers to different kinds on Earth and Twin Earth, then sense can't determine reference in a strict Fregean manner.

On Putnam's account, the meaning of 'water' is composed of two parts: a stereotype and essence. The stereotype is a descriptive component that relates to a cluster of superficial properties associated with the term, such as being a *clear, thirst-quenching liquid*. Since, by assumption, Oscar and Twin Oscar's terms have different referents, the stereotype can't determine what is represented on its own. The essence component is introduced in order to explain the representational difference and relates to what Putnam calls a 'hidden indexical' component of natural kind term meaning. This indexical component relates to a potentially unknown hidden essence of the relevant kind and requires that speakers intend to refer to the local kind that has the same internal structure as paradigmatic instances of the kind the speaker is acquainted with. Thus, the indexical component allows Oscar to refer to H_2O and Twin Oscar to XYZ, since they both intend to use 'water' to pick out the local natural kinds of which they are acquainted. Consequently, Putnam's account can be understood as allowing that meaning (or sense) determines reference, but where meaning itself is determined in part by what is referred to. Thus, since the meaning of Oscar's term 'water' relates in part to what it refers to, i.e. $H2O$, then meaning can determine reference where we understand meaning as involving a hidden indexical component.

Putnam's thought experiment has engendered a wide range of responses that focus on problems with different aspects of the view. I will not cover all of these here. Instead, I will consider two responses to Putnam's view that relate to the existence of narrow content and its possible nature. In particular, I consider two apparently conflicting internalist responses which aim to show that there is available a notion of narrow representational content, and I identify problems for each account.

2.2.1 Descriptive representation of natural kinds

The first account accepts with Putnam that Oscar and Twin Oscar represent different natural kinds, but contends that this representational difference is consistent with a kind of narrow content that functions to descriptively refer to the relevant natural kind at the subject's actual environment. There are a variety of such descriptive

accounts of narrow content in the literature[3], however, I will consider two of the most prominent views associated with two-dimensional semantics (Jackson 2003a 2003b; Chalmers 2002; 2003).

Two-dimensional semantics aims to account for both wide and narrow content through an adaptation of the possible worlds representational framework where wide and narrow contents are associated with distinct ways of considering possible worlds. The narrow content of Oscar's representation WATER is determined by considering various possible worlds that are superficially identical to Earth as being Oscar's actual world and then determining what thing at that world counts as being represented by WATER. For instance, if we consider Twin Earth as being Oscar's actual world, then XYZ would turn out to be what is represented by Oscar's mental representation WATER since it is XYZ that is the clear, thirst-quenching liquid that fills the lakes and rivers, whereas if we consider Earth as the actual world then it is H_2O. Wide content is associated with considering worlds counterfactually, so that given Oscar's actual world is Earth, where WATER represents H_2O, we then consider what would be represented by Oscar's WATER at various worlds considered counterfactually. For instance, on Twin Earth Oscar's WATER would not represent XYZ, since the representation's wide content is fixed by Oscar's actual world and represents that property (H_2O) at any world considered counterfactually.

The two versions of this view provided by Jackson and Chalmers accept the basic model outlined here, but they differ in the fine details. Jackson's (2003a) account holds that Oscar's WATER represents *the actual watery stuff of our acquaintance*, where 'our' picks out the subject and his colinguals who use 'water' to refer to the natural kind they are acquainted with. Importantly, 'watery stuff' is intended to

---

[3] On Searle's (1983) view the descriptive content of 'water' picks out the kind which has the same structure as the cause of the subject's visual experience. Such a view makes the term 'water' function to pick out whatever shares the same structure as instances of which the subject is acquainted with through visual experience. More recently, Mendola (2008) has argued for a demonstrative account where clusters of descriptions refer via a 'dthat' operator which functions to demonstratively refer to the thinker's local environment and the object there which satisfies most of the descriptions. Both views aim to show that in some way Putnam's general approach is consistent with a notion of narrow content. This appears to be common to the various descriptive accounts of narrow representational content.

reflect the way in which H_2O and XYZ appear the same to the subject, and is equivalent to Putnam's stereotype which provides a set of properties that relate to the superficial features of water and twin water. The narrow content is therefore a reference-determining function, since we consider what the relevant narrow content would pick out at various contexts, i.e. the worlds considered as actual. Thus, for Oscar, *the actual watery stuff of our acquaintance* picks out H_2O, as this is what he is acquainted with, whereas it picks out XYZ for Twin Oscar.

Chalmers (2002; 2003) understands narrow content differently. On his account, we understand narrow content in terms of an abstract specification of a subject's epistemic space which is an idealized reconstruction of the subject's a priori knowledge concerning how they take the world to be. What is represented by such narrow content is then associated with the notion of epistemic possibility, so that when we consider what Oscar's thoughts represent, we consider the set of epistemically possible worlds that correspond to how things could be given the subject's a priori knowledge. The main difference between these views is that Jackson considers a single space of possible worlds, rather than introducing a special notion of epistemic possibility. This gives Jackson's approach the benefit of not having to articulate and defend epistemic possibility, which is especially pertinent given extensive criticism of the relevant apparatus required to construct the set of epistemically possible worlds. For instance, some have argued against the construal of the a priori (Byrne & Prior 2006), and others argue that there is no idealized language with which to construct the relevant epistemic spaces (Stalnaker 2001; Sawyer 2007). What is important for present purposes is that narrow content descriptively determines what natural kind is represented by a natural kind representation like WATER (or term 'water') by considering what natural kind at worlds considered as actual possess the relevant watery properties. Crucially, the account of representation relies on an element of indexicality, so that the relevant descriptions (e.g. *the actual watery stuff of our acquaintance*) indexically determine what is represented at the relevant context or world.

Aside from the concerns about the notion of epistemic space, there are two initial problems for this view of narrow content. First, an important question that arises here is whether subjects ever have mental representations that relate to this account of narrow content. In particular, it is not clear that subjects have the relevant intentions to pick out a unique natural kind, especially subjects like Oscar who, we are told, is not aware of the microphysical properties of water. On Jackson's account it is held that Oscar intends to represent a natural kind with WATER (i.e. 'the actual watery stuff' relates to a unique kind), however, this is a significant assumption to make as it seems possible that mental representations may fail to neatly align with natural kinds as they occur in the environment. For instance, where Oscar thinks of watery stuff, one approach might be to distinguish the mental representation of water from that of watery stuff, where the former relates to H_2O, and the latter relates to a more generic kind (I will consider a version of this approach in more detail in Section 2.2.2 below).

A second issue is why we should think of representations as representing natural kinds indexically. This sort of objection runs as follows[4]. Natural kind terms don't function like typical indexical terms like 'I' and 'now' which differ in their content when spoken by different individuals at different times. The content of 'I' changes when uttered by Oscar and when uttered by Twin Oscar because the content relates to the speaker. The term 'water', so the objection goes, has the same content when uttered by different speakers at different times, and hence does not seem to meet the criteria to be an indexical expression. A quick response to this concern is that the claim regarding the hidden indexical component is not that natural kind terms are the same kind of term as indexicals. The claim, instead, is that the indexicality of natural kind terms relates to the issues raised by Twin Earth scenarios which lead to the possibility of terms like 'water' representing different natural kinds than those on Earth. The objection can therefore be disregarded since the debate is not whether natural kind terms are indexicals in the standard sense, but whether terms for natural

---

[4] See Burge (1982) for an example of this kind of objection to allowing natural kind terms to represent indexically.

21

kinds should be evaluated across different contexts (or possible worlds), which merely leads to a form of descriptive representation.

2.2.2 Representation of heterogeneous kinds

An objection to the descriptive model of representation that an internalist might raise takes issue with the claim that Oscar intends to represent a unique natural kind. The descriptive model of narrow representation shares Putnam's intuition that Oscar and Twin Oscar represent different natural kinds. An alternative view, noted above, is that Oscar and Twin Oscar share a form of mental representation that represents something more generic than a natural kind property, reflecting the fact that neither subject is aware of the fundamental nature of water and twin water. If there were a way to show that Oscar and Twin Oscar represented the same property, this would avoid the need for descriptive representation and the assumption that all subjects' WATER representations represent natural kinds. This provides the basis for the second response to Putnam which rejects the assumption that Oscar and Twin Oscar necessarily represent different natural kinds. On this view, it is questioned whether it is appropriate to hold that Oscar represents a natural kind at all, and instead it is suggested that Oscar and Twin Oscar represent the same heterogeneous kind, a kind that, in effect, subsumes a range of natural kinds. Crane (1991) and Segal (2000) have articulated versions of this view, and both hold that narrow content is just ordinary content, so that they deny that there is a need to introduce wide content and a special notion of narrow content to account for the Twin Earth scenario[5]. The central claim is that 'water', for Oscar and Twin Oscar, effectively represents both H_2O and XYZ since their mental representations represent whatever Oscar and Twin Oscar would classify as falling under 'water'. I will say more about Segal's (2000) particular account of representation in Section 2.3.2. For now, I want to outline reasons that support the idea that there can be descriptive representation of heterogeneous kinds.

---

[5] This position is sometimes called 'radical internalism' in that it aims to show that all content is narrow, and/or that the arguments for externalism fail.

Support for the heterogeneous kind view against the position of Putnam and two-dimensionalists, can be provided by putting pressure on the assumption that subjects typically possess natural kind terms or natural kind mental representation. Segal (2004) and Wikforss (2005) (among others) have noted that it is an empirical question whether 'water' is a natural kind term (and a similar point applies to mental representations). For instance, if we found that speakers have different intentions relating to these terms, such as intending to only refer to a more generic or heterogeneous kind, then much of Putnam's thesis would fail. Putnam in fact concedes that some kind terms refer to heterogeneous kinds and not natural kinds, such as 'jade' which refers to both jadeite and nephrite. Furthermore, since many of our kind terms like 'sofa' or 'computer' relate to non-natural kinds, which appear to be the same across possible worlds (more on this below), the proponent of this view should not be worried by the intuition, often expressed in the literature, that 'water' just is a natural kind term.

One worry with this view might be that if we understand watery stuff as a non-natural kind property, then the introduction of this property would be ad hoc, since the property is in some way strange or problematic, and hence we should prefer either wide content or descriptive representation that relates to different natural kind properties, rather than being committed to a variety of non-natural kind properties. I will say more about heterogeneous properties in Section 3, however, it's worth pointing out that there is no problem in principle with taking terms in natural language, or mental representations, to represent non-natural kinds. Consider once more 'computer', which represents an artefactual kind. Any object that instantiates this artefactual kind will be represented by 'computer' or COMPUTER, thus, if we are to question the nature of heterogeneous properties because they are non-natural kinds, then we would need to do so for all other non-natural kind terms.

Setting aside this issue for now, there are two objections that can be raised against the internalist view of Crane and Segal. First, wide content is relatively uncontroversial, and there are apparently good reasons to accept some form of externalism. Second, it's not clear that the heterogeneous kind model of narrow

content is always appropriate, especially for subjects who are aware of the microphysical properties of water, or who explicitly intend to represent a unique kind. Thus, we should be willing to hold that there are a variety of mental representations that relate to the representation of both natural kinds and non-natural heterogeneous kinds. While natural kinds can be represented by the descriptive model, the latter may be also be represented by a form of representation where 'the watery stuff' picks out a heterogeneous property common to Earth and Twin Earth. What would support this view is if there is an account of mental representations that can independently motivate the idea of distinct representations of kinds of natural objects, especially since some externalists will want to hold that there is just one natural kind representation WATER. In Chapter 2 I provide the details of such an account and show how this supports the position suggested here.

Finally, it should be noted that the two-dimensional model doesn't seem right for non-natural kinds like sofas and computers, since intrinsic duplicates may plausibly represent the same properties at different possible worlds. Such kinds are different from the heterogeneous kinds that relate to H_2O and XYZ, so we shouldn't think of the content relating to different natural kinds, but instead to the relevant non-natural kind property that is identical at different possible worlds. For instance, the terms 'sofa' and 'computer' appear to pick out the same non-natural kind on Earth and Twin Earth without necessarily relating to whatever specific objects instantiate these properties. Differences in the underlying microphysical properties of particular sofas would not affect Oscar and Twin Oscar's SOFA representation which function to represent the general artefact type sofa.[6]

To summarise the discussion so far, descriptive representation can provide a form of narrow representation, and there are a variety of properties that subjects' mental representations represent. However, existing views raise important questions about 1) the nature of the description and its basis in mental representations, and 2) what the

---

[6] At this point it is important to note that Burge's (1979) arthritis argument purports to show that all such terms in natural language have wide content, and hence that appealing to non-natural kind terms does not lend support to internalism. I set this argument aside for now, but will consider it in some detail in Chapter 3.

resulting property represented is, such as specific natural kind properties at the subject's actual environment, or more generic heterogeneous properties that don't neatly align with natural kinds. I will consider this form of narrow representation in more detail in Section 3. Before that, I want to consider a second form of narrow representation.

2.3 Fodor on wide informational content

The second basis for externalism and wide content relates to the most prominent approach to representation which understands representation in terms of information. Several versions of informational semantics have been developed in the literature from Dretske (1981), Millikan (1989), Papineau (1993), and Fodor (1994). Here I focus on Fodor's view, however, similar points could be made against the other views. Informational accounts of content hold that a mental representation, *M*, represents what M carries information about. Typically, the relation between M and what M represents is given in causal terms, and Fodor's (1994) account is a version of a causal co-variation theory of content. The basic idea is that M represents what it causally co-varies with. On this view, M would represent whatever object, *X* (an individual, kind, etc,), would cause M to trigger in normal circumstances. For instance, CAT can be triggered by cats and hence would represent cats. However, this can't be the whole story because of things that merely look like cats, and Twin Earth cases involving cats and twin cats. Suppose on Earth there are cat lookalikes that trigger CAT in normal circumstances but which we don't want to say are represented by CAT, or on Twin Earth twin cats are creatures with fundamentally different properties. This problem is known as the 'disjunction' problem..

The problem is where we cannot say what unique X a concept in Twin Earth-type cases represents, and must instead appeal to disjunctive contents where a concept represents, say, X's or Y's. For instance, when we consider deviant causes which may trigger S's HORSE concept such as cows on a dark night, the problem is to say why HORSE represents horses, and not horses or cows on a dark-night, given that both can cause S's HORSE to trigger. Fodor's (1994) response to the disjunction problem involves noting that lookalikes causing HORSE to trigger asymmetrically

depends on horses triggering HORSE. That is, the only reason lookalikes trigger M is because the genuine X's trigger M. This is the basis of the asymmetric dependence theory of wide content. In order to explain the Twin Earth cases Fodor introduces an additional constraint. On Fodor's view we hold that Oscar's CAT representation represents cats and Twin Oscar's CAT represents twin cats, because in general M represents those X's which have actually caused M to trigger, thus Oscar and Twin Oscar's CAT relate to different kinds because different kinds have actually caused their CAT representations to trigger. In appealing to what X's actually trigger M, this rules out disjunctive representation of, say, H_2O and XYZ, because Oscar's WATER is triggered by H_2O whereas Twin Oscar's TWIN WATER (or his use of 'water') is triggered by XYZ.

2.3.1 Relational representation of appearances

While Fodor intends to identify a notion of wide content, the informational view of representation does not preclude the notion of narrow content, since it is possible to identify a form of narrow representation based on similar principles. There are two ways that an internalist might make this point, although, as we shall see, there are substantive differences between the views which put them at odds with each other.

Assume for now that Oscar's representation WATER relates to the natural kind property H_2O. On Fodor's account, this follows because it is H_2O that causes Oscar's WATER to trigger. Prinz (2000) notes that using the same form of representation we can identify a property that Oscar and Twin Oscar are both related to, and hence that there is a kind of narrow content that relates to their WATER representations. According to Prinz, mental representations are individuated by the superficial properties they function to detect[7]. By hypothesis, H_2O and XYZ share their superficial properties, for instance they have the same appearance properties (being clear, translucent, etc). On the informational model of representation, M's represent what actually cause them to trigger. Thus, supposing that the appearance properties of H_2O and XYZ are what cause Oscar and Twin Oscar's WATER to

---

[7] I will not consider the details of this account of concepts and concept individuation in detail here, leaving the discussion to Chapter 2 where I consider this particular view in some detail.

trigger, then their representations will represent the same appearance properties. For instance, it is plausible that being clear is a property common to H_2O and XYZ, and this property, along with the other appearance properties, will cause WATER to trigger in both Oscar and Twin Oscar. Consequently, there is a form of narrow content common to intrinsic duplicates that derives from the relational view of representation. Note that we can also accept something like Fodor's account of wide content whilst at the same time acknowledging narrow content. Thus, while Fodor's account may be suitable for representing natural kind properties that differ across possible worlds, it seems that narrow content might be required for representations that represent other kinds of properties, particularly those properties that are common in the Twin Earth scenarios.

The first worry for this view is that it is committed to a specific theory of mental representations which can be resisted by externalists and internalists alike. In Chapter 2 I outline why we should reject the appeal to just one kind of mental representation, although I show that Prinz' approach can provide an important part of the story. A more important worry for present purposes is that while narrow content is taken to relate to appearance properties, we aren't told what these are, other than that they are common to distinct natural kinds. As I will discuss in Section 4 below, externalists can raise the objection that the superficial properties, like being clear, could be further natural kind properties, and hence that Oscar and Twin Oscar in fact represent different appearance properties. However, one way of supporting the idea that appearance properties are not natural kinds is to understand appearance properties as a variety of heterogeneous kind property. A version of this approach is developed by Kriegel (2008) who construes appearances as response-dependent dispositional properties shared by different natural kinds. The relevant dispositions are to cause certain cognitive effects in a perceiver. This raises the question of whether this account of appearances is coherent, and whether an externalist might raise the Twin Earth argument again. In Section 3 I will say more about these properties.

Before considering appearance properties in more detail, it is important to note a third objection to this view of narrow content. In restricting narrow content to appearance properties, this doesn't allow narrow content to relate to natural kind properties, thus narrow content is restricted to a special kind of property which significantly restricts the internalist's position. Both Prinz and Kriegel contend that it is wide content that relates to the representation of natural kinds, which is a position at odds with the descriptive views outlined above where it seemed plausible that descriptions could provide a representational notion of narrow content that relates to natural kinds and heterogeneous kinds that subsume different natural kinds. Before developing an alternative position, I want to consider a way of showing how the relational view of representation may relate to more fundamental properties.

2.3.2 Relational representation of heterogeneous kinds

While Prinz (2000) appears to accept that the representation of natural kind properties relates to the natural kinds that actually triggered the relevant representation, Fodor's (1994) concessions elsewhere suggest that his account of representation is not strictly based on what has actually triggered a subject's representations. This provides the basis for rejecting the restriction of narrow content to a special kind of property. In particular, where Prinz aims to show how narrow content can relate to relations between S and superficial properties common to Earth and Twin Earth, Segal (2000; 2009b) has shown that Fodor's account unjustifiably restricts representation to S's actual world, even for the representation of kinds that don't involve relations between S and those kinds.

Consider the case of Swampman, who is an intrinsic duplicate of Fodor who has spontaneously appeared on Twin Earth. As yet, Swampman has no causal contact with twin water (XYZ), however, according to Fodor, Swampman's WATER representation represents XYZ because we are to consider what would cause his WATER to trigger at nearby possible worlds. The reason H_2O isn't represented by WATER is because there is no H_2O in any of these worlds, hence Swampman's WATER represents XYZ. As Segal (2000) rightly points out, Fodor's position relies on an unwarranted restriction of the relevant counterfactuals (what would cause

WATER to trigger) to nearby possible worlds where the watery stuff is XYZ. Instead, we should consider what would cause WATER to trigger in nearby worlds where the watery stuff is H_2O.[8] The worry with Fodor's stance is that the relevant set of possible worlds he considers can be extended to include worlds with H_2O. What is important in questions about what intrinsic duplicates represent is that the possible worlds are nomologically possible worlds. This requires that the relevant worlds subjectively appear the same and are functionally identical while the microphysical properties of the environments may be different, (I say more on the relevant set of possible worlds in Section 3 below). Thus, we should consider what Swampman would represent in worlds that are superficially identical to Twin Earth, but where there is H_2O, not XYZ. In which case, we should extend the range of relevant counterfactuals to the *actual and possible* causes of WATER. Consequently, the correct understanding of the relational view of representation is not what has actually triggered S's representation, but what has or would cause S's representation to trigger at relevant possible worlds. Therefore, since both H_2O and XYZ would trigger both Oscar and Twin Oscar's WATER, they represent the same heterogeneous kind.

Using the informational view as a basis for heterogeneous kind representation provides a useful contrast with the two-dimensional account, as the views rely on different mechanisms of representation and understand the possible worlds framework in different ways. Recall that on the two-dimensionalist's account of narrow content we consider what a relevant description (*the actual watery stuff of our acquaintance*) picks out at possible worlds considered as S's actual world. For wide content we consider what S represents at worlds considered counterfactually given that her actual world is known. The relational view holds that the relevant counterfactuals relate to the actual or possible triggers of S's representation (e.g. WATER) at a relevant set of possible worlds. On this view, S's narrow content is understood as a heterogeneous property that subsumes any natural kind that would have the relevant causal effect on S, whereas the descriptive view understands

---

[8] Segal (2009b) has recently noted that Fodor is perhaps entitled to exclude certain counterfactuals, yet maintains the general objection that this restriction is unwarranted.

narrow content as a reference-determining function that picks out a unique natural kind property at the subject's actual world.

As noted above, Segal's view is intended to be a radically internalist account of representation, dispensing with any notion of wide content, and does not rely on descriptive representation. However, just as he criticises Fodor's restriction of relevant counterfactuals, a similar point can be raised against Segal. If we are only to consider counterfactuals of the form *what did or would cause S's representation to trigger*, then this rules-out wide content that relates to counterfactuals in which we fix S's actual relations to natural kinds. All that the internalist should commit to is that there is a notion of narrow content that can be extracted based on the descriptive and relational frameworks, and that this can relate to the representation of natural kinds by representing a more generic heterogeneous kind.

I want to forestall one kind of objection here, although I will consider a specific version of this kind of response in Chapter 3. Note that in relating WATER to the actual or possible triggers of S's mental representation WATER, the view does not entail that S's representation takes as its content a disjunction or conjunction of different natural kinds (for instance, the set containing all the natural kinds with the watery properties across various possible worlds). The property view of representation relates representations to the properties they represent. Hence, on this view, since WATER fails to represent a unique natural kind, it takes as a property a heterogeneous kind. Such a property is of course an abstraction, and does not align with the kinds found at any one particular possible world, nor would they be discoverable via scientific methods. However, as with the case of many non-natural kind properties, such as algorithms and squares, representing properties that are abstractions from particular instances of the property is not intrinsically problematic. For instance, S's ALGORITHM representation may be triggered by a particular algorithm on Earth, however, there need be no microphysical properties in common among the various instances of algorithms on Earth and Twin Earth. Hence, this is not a serious worry for the internalist. What we should be willing to recognise is that

some representations fail to represent properties that align with the actual natural kinds at a subject's environment.

2.4 Taking stock of the debate

It's time to take stock of the wide/narrow content debate. I have outlined some of the background to these debates, and crucially I have highlighted two basic forms of narrow representation together with some objections to existing versions of these accounts. It is important to note that, as they stand, they involve incompatible views about what narrow content is, such as the relevant basis of narrow content in mental representations, the form of representation, and the properties represented. In Section 3, I will outline and develop a conciliatory approach which provides a way to accommodate both descriptive and relational representation. The resulting view can be independently motivated by considering the key issues and widening the notion of narrow representation, so that we don't have to hold that only a single representational relation provides all the varieties of narrow content available. My proposal is that we can outline a systematic basis for this difference by paying attention to the different properties represented by different kinds of mental representations. Thus, for some mental representations and the properties represented, descriptive representation may be more suitable, and for others non-descriptive relational representation may be more suitable. In anticipation of the objection that this approach is ad hoc, or that it begs the question against an alternative theory of mental representations, I consider in more detail the nature of mental representations in Chapter 2, and show, in particular, that the forms of narrow content to be outlined in Section 3 are consistent with recent work on concepts in psychology and philosophy.

3. Narrow representation and truth conditions

Section 2 provided a critical overview of two of the main forms of wide and narrow content. This review provides the basis for developing an alternative approach to narrow content which incorporates both forms of representation, but which is independently motivated by considerations relating to the nature of supervenience and mental representation. In Section 3 I will outline the case that one lesson from

the extensive literature on wide and narrow content is that a form of representational pluralism may be the best approach in questions of mental representation and content, which not only endorses both wide and narrow content, but also a variety of narrow contents. I begin Section 3 by outlining the general framework. In Section 3.2 I clarify the supervenience claim that supports the internalist's account of representation. In Section 3.3 I outline the range of mental representations and forms of narrow representation. In Section 3.4 I show that narrow content relates to a variety of truth conditions.

3.1 The framework

As noted in Section 2, externalism and internalism are characterised by different supervenience claims. Internalists contend that some content supervenes on intrinsic properties such that intrinsic duplicates will share a kind of content. In order to accommodate two kinds of narrow content, which relate to different conceptions of what narrow content is, we need to be clear about the notion of supervenience. This is especially important given that many objections to internalism revolve around conceptions of supervenience that internalist's needn't accept. Just as important as supervenience is the nature of the intrinsic properties which narrow content is said to supervene on. Most proponents of narrow content don't offer very specific accounts of which properties are important. One exception is Prinz (2000) who appeals to concrete conceptions of concepts, in particular, prototypes, and more recently proxytypes (Prinz, 2002)[9]. In Section 2 I noted that one worry for existing descriptive accounts is that we are not given much reason to endorse the specific descriptive component. An advantage of Prinz' approach is that his view is supported by empirical work in psychology. As I will develop more fully in Chapter 2, recent work on concepts supports a view of mental representations as being comprised of a variety of conceptual structures. In anticipation of that more detailed discussion, in this section I will show that permitting a range of mental representations associated with terms like 'water' provides support for endorsing this approach to narrow content, since a plurality of mental representations can support the representation of different kinds of properties. Finally, by paying attention to the

---

[9] I say more about these in Chapter 2.

nature of the properties represented, and the means by which they are represented, my approach shows how the different forms of narrow representation can be incorporated within a unified account of narrow content. In the rest of Section 3 I will develop these components in greater detail.

3.2 Supervenience

Internalism is the thesis that there is a kind of content that supervenes on the internal states or structure of a subject, such that a subject S and an internal duplicate of S will be in states with the same narrow content. Where narrow content is taken to be genuinely representational yet requires appeal to the external environment for its characterisation, an internalist should hold that narrow contents relate to the external environment. A proponent of narrow content may also accept that some mental states can be individuated by their wide contents, allowing for wide belief states like believing *that there is water in the cup*. Hence, the internalist and externalist can both allow that we individuate some beliefs in part by reference to factors outside of the subject. This understanding of narrow content in terms of supervenience and relations to the external world defies a simplistic, and incorrect, construal of internalism as the view that narrow content bears no relation to the external world, and that any intrinsic duplicates will be mental state duplicates. In the rest of this section I will clarify the notion of supervenience and the nature of narrow content.

3.2.1 Internalism, supervenience, and representation

The most common way of understanding the internalist's supervenience claim is to say that narrow content supervenes on internal states from the skin in. Supervenience is a relation that holds between two properties such that property B supervenes on property A iff any objects that share property A also share property B. On this view, internalism is the thesis that there is a kind of content that supervenes on the internal states of individuals, such that internal duplicates will be in states with the same narrow content.[10] A worry for this way of putting the thesis is that it is

---

[10] Farkas (2003) objects to this way of defining internalism, noting that an externalist argument can be run on properties that are internal to a subject, such as meningitis, and holds the view that all of a subject's mental properties are determined by how things seem to the subject. One line of response would be that only those internal states that relate to a subject's narrow contents are needed to be

implausible that all internal duplicates of S will, necessarily, share a mental representation with the same content. For instance, consider a duplicate of S on a world which is radically unlike Earth. Where the content of a mental representation is what it represents, if this world contains few or no properties or kinds found on Earth, then the internalist should accept the view that S and her duplicate may fail to represent alike. Consequently, it is crucial to be clear about the representational nature of narrow content when explaining the supervenience claim.

To recap the views introduced in Section 2, on the two-dimensional view of narrow content duplicates may represent differently when we consider Twin Earth scenarios such as where S and T are on Earth and Twin Earth, and their identical narrow contents function to determine different extensions at their respective environments. Narrow content reflects S's way of representing the world as being from her perspective. We consider this way (given descriptively or by S's Epistemic Space on Chalmers' view), at a given world and see what X would be represented at that world. For instance, on Earth, S's narrow content functions to pick out the unique local natural kind with the relevant descriptive properties S associates with 'water', i.e. H_2O, while T's narrow content will pick out the local watery kind, XYZ. The representational role of narrow content is identical on this view, but it can mean that Twins' concepts have different extensions and that different things make their thoughts true. S's thought expressed as 'water is wet' is true if and only if H_2O is wet, while T's similarly expressed thought is true if and only if twin water is wet. On the informational view of narrow content, internal duplicates in the Twin Earth scenarios can represent alike, such as where their narrow contents function to pick out an identical set of superficial, dispositional, or other property, at relevant possible worlds. S and T's concepts have the same extension, namely the set of properties P.

---

duplicated to ensure duplicates share narrow content. For instance, it seems reasonable to hold that S's narrow contents supervene on certain of her brain states, such that any duplicate with the same brain states must be in mental states with the same narrow content. However, I set aside Farkas' point here as her main concern relates to the individuation of mental properties, whereas I am concerned with supervenience. It's also worth pointing out that I will set aside issues about whether the mind may extend beyond the borders of the body, such as with the extended mind thesis (Clark & Chalmers 1998), as this is a question about the individuation of mental states/mental properties, whereas my concern is the question of whether there are mental states that bear narrow content.

We consider what would trigger S's concept at a range of possible worlds, and thereby identify the extension of S's concept.

The essence of the internalist thesis defended here is that there is a kind of genuinely representational content that supervenes on internal states of the subject. However, the supervenience claim alone is not sufficient to understand the representational nature of narrow content. While it sometimes appears to be held in the literature that externalism must be true because it is necessary for genuinely representational content to make reference to the environment, as the two basic forms of representation suggest here, internalists may also appeal to the environment when articulating narrow content. This requires the internalist to be careful when articulating the nature of narrow content.

3.2.2 Inter-world and intra-world narrowness

Following Jackson & Pettit (1993) and Segal (2000), narrow content is perhaps best likened to the property of being water-soluble. It should be noted, however, that despite agreement on the analogy with this kind of property, care needs to be taken with regard to the implications that the analogy has for the nature of narrow content. In particular, Jackson & Pettit (1993) and Segal (2000) appear to draw different conclusions about narrow content from its basic nature. As such, I will show below how the internalist who accepts the view outlined in Section 2 should understand the nature of narrow content.

To see why the property of being water-soluble is an apt comparison for narrow content, we can draw a distinction between inter-world narrow content and intra-world narrow content (Jackson & Pettit 1993). A rough characterisation of this distinction may be put as follows (subject to clarification below). A property of S is inter-world narrow iff any duplicate of S will share that property no matter which worlds they are in. A property of S is intra-world narrow iff S and a duplicate of S share that property within the same possible world. As noted above, internalists should not be committed to a view in which duplicates represent the same properties

at worlds that are radically different, where there are few or no properties in common at these worlds. Consequently, narrow content should not be inter-world narrow.

In order to see why narrow content is best modelled on intra-world narrow properties, we can consider three kinds of properties and whether they are shared by internal duplicates. Take the property of being square (Kallestrup 2012). The property of being square supervenes on the internal structure of a given object, such that an intrinsic duplicate of a square object is square at any possible world. Therefore, being square is an inter-world narrow property since a duplicate of a square object will share that property at every possible world. Contrast this with the property of being water-soluble. While an object's disposition to dissolve when placed in water appears to supervene on the internal structure of the object, Jackson & Pettit (1993) point out that to know if something is water-soluble you must also know about the causal connections between being in a certain internal state and what happens upon being placed in water. An important aspect of water-solubility is that it takes more than the intrinsic properties for the property to be realized, such that the environment must facilitate the realization of the property. In this case, the laws of nature must be such that the object will dissolve when placed in water. With regard to the causal connections, what matters is how the object relates to water, give the relevant laws of nature in place. Hence, it is the objects being placed within water which will result in the realization (or not) of the property of being water-soluble. Thus, while an object may dissolve when placed in water on Earth, if we consider an internal duplicate of the object at a possible world where the laws of nature are such that it would not dissolve when placed in water, then that property will fail to be realized at that world. However, the object will still possess the disposition to dissolve when placed in water at worlds sufficiently like Earth[11]. Consequently, being water-soluble is intra-world narrow, since a duplicate of such an object will share this property when considered at the same possible world.

---

[11] I will expand on the notion of worlds that are sufficiently like Earth below. For now, the idea is that the world will be such that the laws of nature facilitate the object's dissolution when placed in water.

Before explaining the implications for narrow content it's worth considering a third kind of property introduced by Stalnaker (1989) designed to put pressure on the notion of narrow content. Being a footprint fails to supervene on internal structure. Only something that is a foot-shaped imprint caused by a foot is a footprint, hence something that is merely an internal duplicate of a footprint, such as an imprint in the sand caused by the movement of the tide, will fail to be a footprint. In terms of the point above, if all you know is confined to the internal structure of the imprint, then you don't know if the imprint is a footprint. To know this, you must also know that the imprint was caused by a foot. Hence, unlike being water-soluble, being a footprint fails to supervene on internal states of an object since the causal history is crucial to the relevant property. An implication for narrow content is that in order to get intentional or representational properties from internal states, we must consider the relevant internal state in an environment where it bears causal connections to that environment, and what the state represents will involve more than just its internal state. Hence, internalists should appeal to causal connections to the world, however, this doesn't make content wide, since this content supervenes on internal structure.

One way of understanding the difference between wide and narrow properties is in terms of world-dependence, that is, the extent to which a given property depends on the specific environmental features at the relevant environment. Footprints have a stronger world-dependence than being imprints in the sand with a certain internal structure, since they depend on how feet are shaped at that world. Thus, where feet are shaped differently on Twin Earth, then an internal duplicate to an Earth footprint will not be a footprint. Note, however, that these imprints will still share a range of properties such as being imprints in the sand, having a certain internal structure, and appearing identical to observers. These properties, we might say, have a weaker world-dependence than the property of being a footprint, relying as they do on facts about the world which are in some sense more general than being a footprint.[12]

---

[12] This distinction between weak and strong world-dependence has an extensive literature, which I will not consider in detail here. For current purposes I will assume that the basic distinction has at least prima facie plausibility.

This captures an important point about narrow content. While it supervenes on internal structure, it nonetheless depends upon the external environment for its representational nature, though having a weaker world-dependence than wide content. S's mental representation WATER may have a narrow content which represents superficial properties, such as being clear, odourless, liquid, and these properties have a weaker world-dependence than the property of being H_2O, which may relate to the wide content of S's mental representation. Thus, while S's wide contents may be fixed by highly world-specific facts about her actual environment, her narrow contents may depend only on a more general way on the facts about her environment.

This way of understanding narrow content applies to both kinds of representation outlined in Section 2. On the descriptive view, S's narrow content is a function from worlds to extensions. We consider what would be represented at a given world, W, by considering what X at W meets the relevant description. On the narrow informational view, we consider what X at $W_1…W_n$ would trigger S's concept. On both views, S's C may represent relevant properties, hence Twins can share narrow contents at worlds that are suitable alike (again, I will say more about these worlds below). While the wide contents of S's concepts are fixed to her actual environment, the narrow contents can relate to a broader range of extensions by virtue of the weaker world-dependence of narrow content.

Narrow content therefore relates to the external environment, albeit in a more general way than wide content. In order to get genuine representation, narrow content thereby requires a subject to bear causal relations to a world and the properties represented at that world. On the descriptive view, we consider S at any of the worlds $W_1…W_n$ and what X at W S's narrow content would pick out. We can understand S's relations to w as underpinning the representational nature of her narrow contents such that at W, given her relations to things at w, her narrow content will pick out the set of X's which fit the description associated with her narrow content. On the informational view, causal relations to things at a world plays a central role in determining what falls within the extension of S's concepts. It is by considering what X's would trigger C at $W_1…W_n$ that we determine the extent of S's

narrow content. In practice, this means that we consider the sum total of causal relations S bears to things at W, and see what sets of things trigger relevant concepts.

The two accounts may then be seen to be compatible to some extent. For instance, when considering what narrow content descriptively represents, we may consider what X's at the relevant world would trigger S's C. However, the accounts of representation are importantly different. On the descriptive view, the extension associated with narrow content is determined on a world by world basis, fitting with narrow contents' role as a function from possible worlds to extensions at those worlds. On the informational view, narrow content relates to the representation of a property or set of properties, which needn't be considered on a world by world basis. The descriptive views tend to focus on representing natural kind properties, which differ at $W_1…W_n$, hence different extensions depending on the world considered. The informational view, however, tends to focus on superficial properties which are the same at $W_1…W_n$, resulting in the same extension at each of these worlds.

Throughout the above discussion, I have appealed to sets of possible worlds and made use of the notion of possible worlds that are sufficiently similar with respect to the laws of nature to characterise the notion of narrow content. In the next Section I conclude the discussion of intra-world narrowness and the nature of narrow content by considering which relevant set of worlds the internalist requires in order to facilitate the representational nature of narrow content.

3.2.3 Cognitive content and nomologically similar possible worlds

Above, I noted that internal duplicates of S share an intra-world narrow property within the same possible world as S. The main point here is to avoid a commitment to the view that any internal duplicate of S will represent the same properties as S, such that where S on Earth may represent H_2O, there is a sense in which T's narrow content may relate to the representation of XYZ. For instance, on the descriptive view, S and T can share narrow content while their concepts may have different extensions and their thoughts different truth conditions. This means that narrow content is not inter-world narrow, since where S bears causal relations to H_2O and T

to XYZ, they represent differently. However, when we consider S and T within the same world, they will represent alike since narrow content is intra-world narrow. This entails that if both S and T are on Earth, then S and T would share states that relate to the same extension and thoughts with the same truth conditions.

One worry with this way of characterising narrow representation is that it fails to allow that in Twin Earth cases there are a range of properties that S and T can represent alike, and hence that their thoughts can have identical truth conditions. This seems true of thoughts of superficial and abstract/functional properties like being clear or being an algorithm (more on this below). Despite S's being causally related to H_2O and T's being causally related to XYZ, this doesn't seem to prohibit their concepts representing the same superficial and functional properties (for instance the watery properties common to H_2O and XYZ). Where narrow contents relate to the representation of these non-natural kind properties, and Twins represent the same properties, then their mental states have the same narrow representational properties. Furthermore, on the informational view of narrow content noted in Section 2, being at the same possible world was not required in order for S and T to represent the same properties.

It is important to note that we can distinguish between kinds of Twin Earth scenario. Putnam's original Twin Earth case was such that Twin Earth was a different planet within the same possible world. An alternative is where Twin Earth is another possible world. On the descriptive view of narrow content, Twins may be in states with the same narrow content, but this content functions to pick out different properties at the different worlds $W_1...W_n$. On the informational view, however, since narrow content relates to superficial and other non-natural kind properties, we can allow that the property represented may be the same, even where Twin Earth is another possible world.

In relation to the characterisation of narrow content as intra-world narrow, Jackson & Pettit (1993), appear to hold that intra-world narrowness entails that narrow contents relate to the same representational contents only when considered at the same

possible world, such as worlds where the same underlying natural kinds would be represented by S's WATER concept. Kallestrup (2012, 120) characterises the view such that "Intra-world narrow properties are only shared by doppelgangers across nomologically identical possible worlds, i.e. worlds governed by the same laws of nature.". However, on the informational approach to narrow representation, we do not need to restrict the relevant set of worlds to nomologically identical worlds, since something less than identical worlds, i.e. nomologically similar possible worlds, will be worlds where Twins may still represent the same set of, say, superficial properties. Importantly for present purposes, this is consistent with the comparison with the property of being water-soluble, since an object could plausibly dissolve at possible worlds where the different laws of nature have no impact on this property being realized. Consequently, the internalist should allow that duplicates may represent the same properties or kinds at nomologically possible worlds that are similar enough to sustain representation of the same properties. This view requires an account of what it takes for these worlds to be similar enough.

On a first approximation, similar possible worlds are those where things seem the same to any intrinsic duplicate of S at those worlds, irrespective of differences in the underlying laws of nature or natural kinds at those worlds. Put another way, things seem qualitatively the same to S and T at all of these worlds, i.e. despite differences in the natural kinds on Earth and Twin Earth, and differences in the underlying laws of nature, from S's and T's perspective things seem to be qualitatively identical

If this is the case, then where narrow content relates to how things seem to the subject, then these worlds will be such that things seem the same to S and her duplicates, and a duplicate of S at any of these similar worlds will represent things as being the same as S on Earth. Where superficial properties relate to appearances, then these worlds will likely share superficial properties, and duplicates' narrow contents will represent the same properties. While some of these worlds will have

41

different natural kind properties, they may have a range of other non-natural kind properties in common such as functional, artefact, and computational properties. [13]

This is slightly stronger than the notion of intra-world narrowness, since appealing to nomologically similar worlds can allow that Twins at different possible worlds, not just at different planets within the same possible world, can represent alike with their identical narrow contents. However, this view can incorporate the idea that narrow content is not inter-world narrow, since there are clearly possible worlds where duplicates can fail to represent alike, especially the worlds where things would fail to seem the same to subjects considered.

At these nomologically similar worlds, many properties including superficial, artefactual, and computational, will be identical, while some natural kind properties and the laws of nature will differ. Yet narrow content may relate to these properties via the two forms of narrow representation identified here. So long as things seem the same at these worlds, then S's narrow content will be triggered by objects with the same superficial properties, and/or S's narrow content will still function to pick out the local X with relevant descriptive properties. The internalist's claim is therefore that intrinsic duplicates may share narrow content at nomologically similar worlds, and that this is compatible with duplicates' narrow contents relating to both the same properties and different properties, depending on which form of representation is in question.

Summarising the discussion so far, internalism is the view that narrow content supervenes on the internal states of subjects and is an intra-world narrow property of subjects, such that an internal duplicate of a subject S will be in a state with the same content at nomologically possible worlds that are sufficiently similar to S's world. While narrow content supervenes on internal states, in order to understand the representational nature of this content, we must appeal to the external world, and the causal connections subjects bear to that world. Depending on which form of narrow

---

[13] It is worth noting that some externalists suggest running the Twin Earth argument on these types of properties, or Burge's (1979) arthritis argument. I consider both of these positions in the thesis.

representation we consider, duplicates may or may not represent the same properties, but on both views they will share narrow contents at this relevant set of worlds.

In the rest of Section 3, I consider the two basic forms of narrow representation in more detail, paying particular attention to the kinds of properties represented and the relevant truth conditions associated with such narrow content.

3.3 The varieties of narrow representation

Internalists often construe the relevant intrinsic properties in terms of mental representations[14]. It may therefore be asked which intrinsic properties in particular relate to narrow contents, and whether suitable accounts of mental representations can sustain the view. As shown in Section 2 above the descriptive and relational accounts of narrow content make substantive claims about the relevant mental representations that are held to bear these contents. For instance, the descriptive account of Jackson (2003a) proposes that there is a kind of mental representation that encodes a descriptive condition that picks out a unique natural kind at the thinker's actual environment. Similarly, Prinz (2000; 2002) contends that there are specific mental representations that relate to the representation of superficial appearance properties, and appeals to the theory of concepts to give substance to this account of mental representation. The worry for these views was in their respective commitments and whether any such mental representations can be defended. Without going into too much detail here, the view I will develop holds that mental representations come in a variety of structures. Corresponding to these structures are different representational contents, and these can in turn relate to different forms of representation. As already noted, in Chapter 2 I will show that a recent approach to concepts, which draws on work in psychology, is consistent with individuating concepts by their narrow contents. In this chapter I will presume much of this framework without drawing on it explicitly. To this end, I will work with a rough notion of pluralistic mental representations where there is no single mental representation that accounts for all of a subject's representations. On the pluralist

---

[14] Segal (2000) notes that the answer will likely come from work in a future psychology, and so avoids speculation on which particular microphysical properties are relevant.

view, different mental representations relate to the representation of different kinds of property. In the rest of Section 3 I will focus on the remaining issues of the internalist's representational framework: the properties represented and the different representational mechanisms.

In Section 2 I noted that existing accounts of narrow content focus on representation of two main kinds of properties, and that this representation goes by way of different representational mechanisms. What I want to show is that the internalist should endorse both kinds of view, and that this is plausible once we consider the nature of the properties and mechanisms in more detail and relate these to different kinds of mental representations. I consider each of these in turn.

3.3.1 Descriptive representation of natural and heterogeneous kinds

Recall that one objection to a two-dimensional account of narrow content is that it appears to assume that subjects like Oscar will represent the unique local natural kind at his environment, rather than a more heterogeneous kind. I have suggested that the latter view could be associated with a version of descriptive representation, where the description *the clear liquid that fills the lakes and rivers* would pick out the same heterogeneous kind across $W_1 \ldots W_n$. Against proponents of the heterogeneous view, I noted that the exclusion of wide content was not well-motivated, and that internalists needn't choose between allowing subjects to represent natural kinds or heterogeneous kinds. The consequence of this critique leads to a version of representational pluralism, which allows for variation in the mental representations of different subjects, as well as individuals having a variety of mental representations associated with the same term.

Consider the case where Oscar's WATER represents a heterogeneous kind that subsumes H_2O and XYZ. This is consistent with allowing that someone in the know will represent the kind H_2O with the mental representation WATER. On the descriptive view this difference can be explained by understanding the differences in the ways of conceptualizing different properties. Oscar's WATER conceptualizes the world at a relatively superficial level in terms of properties common to different

natural kinds, whereas others may have a more sophisticated conception in terms of a natural kind individuated by more fundamental properties. However, a case can be imagined where an individual, S, has at least two different ways of conceptualizing the world with her WATER representation. Combining the two representations outlined, S could represent things in a superficial way with WATER, such as when she thinks only in terms of superficial properties, yet S may also represent things at a more fundamental way in terms of the fundamental natures of different kinds. Consequently, S's WATER may be associated with at least two ways of representing things. The descriptive view of narrow content can allow that S's WATER is associated with two kinds of narrow content, one which represents a natural kind, and another that represents a heterogeneous kind individuated by its superficial properties.

A different way of motivating the same view is to note that the accounts of content from Putnam (1975) and Jackson (2003a) make assumptions about the intentions of subjects and the associated descriptions that relate to the representation of natural kinds. Hence there is a precedent for deciding what Oscar represents by reflecting on the likely intentions of subjects in these cases. Thus, the pluralist view can be motivated by the prima facie plausible claim that Oscar's representation differs from those who know about the microphysical properties of water. Similarly, it is prima facie plausible that S could think of objects in various ways. Consider an alternative to the example above. Suppose S has a theory of colours relating to the microphysical properties of surfaces of objects and the different wavelengths of light. S may associate such a theory with her representation RED, yet S may also have a more superficial understanding relating to the redness of S's perceptual experiences, and thereby associate with RED a more superficial representation relating to the experience of redness [15]. Such a distinction between ways of conceptualizing different properties that are associated with the same linguistic expression ('water', 'red', etc) is all that is needed to motivate the pluralist view of narrow representation.

---

[15] Chalmers (2006a) has identified a notion of narrow perceptual content associated with primitive redness, understood as the redness of experience.

Two important issues at this juncture are whether the non-natural kind properties associated with WATER and RED are suitable kinds to be represented, and, importantly, how these properties should be understood. Below, I will consider in more detail the nature of appearance properties like redness and suggest that the relevant heterogeneous properties, such as that associated with WATER, can be understood in a similar way to other non-natural kind properties, including appearance and artefactual properties. Thus, in so far as we can make sense of mental representations representing these other non-natural kind properties, the heterogeneous view should be accepted. To this end, in Sections 3.3.3 and 3.3.4 I outline both appearances and other non-natural kind properties like artefacts. Before this, however, I outline in the next section why internalists should also endorse non-descriptive relational representation of heterogeneous kinds.

3.3.2 Relational representation of heterogeneous kinds

In Section 3.3.1 I suggested that the descriptive view of narrow representation can relate to the representation of natural and heterogeneous kinds, and that a pluralist view of mental representations can support this view. There is another approach to narrow representation which appeals to informational relations between representations and the actual or possible triggers of those representations. The representational mechanism and associated counterfactuals are different on this approach. This suggests that, in addition to pluralism about what is represented, the internalist should be willing to endorse pluralism about the form of narrow representation. On the relational view we consider what did or would cause a representation to trigger at relevant worlds in normal circumstances. Consider Oscar and Twin Oscar once more. Both H_2O and XYZ would cause their WATER to trigger, thus, on this view, they represent the same heterogeneous kind. The associated mental representation could be understood as relating to the same combination of functional and superficial properties that they each associate with WATER.

In 3.3.1 and 3.3.2 I have outlined two different accounts of narrow representation. The differences boil down to two things: First, it matters whether we think of S's

mental representation as picking out a unique natural kind or a more heterogeneous kind. I contend that we can't know this a priori since what is needed is an account of the relevant mental representations, such as what lay subject's theories of various natural objects are, and whether these pick out natural or heterogeneous kinds. Second, the representational frameworks are importantly different: one appeals to the actual or possible triggers of mental representations, the other stipulates that S represents whatever satisfies a certain description at S's actual world. The difference is subtle, but is the difference between a version of informational representation understood as a kind of narrow representation, and a version of descriptive representation which appeals to descriptions that function to pick out what satisfies the description.

Returning to the issue of the nature of heterogeneous properties, the next two sections motivate the representation of kinds other than natural kinds. While this doesn't say precisely what property is represented in the case of a heterogeneous kind common to H_2O and XYZ, it does suggest a way of supporting the idea that subjects' mental representations are not restricted to representing kinds that neatly align with the fundamental natures of natural kinds in the environment.

3.3.3 Relational representation of appearances

As noted above, Prinz (2000) and Kriegel (2008) defend views on which narrow representation goes by way of relations between S and superficial properties instantiated at the set of nomologically possible worlds $W_1 \ldots W_n$. We need to be very clear on what these superficial properties are so as to support the idea that narrow content relates to genuine properties. A basic approach can be extracted from Shoemaker's (2006) account of appearances. Take an object O and subjects A and B with different perceptual systems. Suppose that A and B are such that when A perceives O she has an experience associated with our experience of redness, and when B perceives O she has an experience associated with our experience of greenness. O's fundamental properties are fixed, but the appearance properties depend in part on how O appears to perceivers like A and B. O has the disposition to cause in A and B perceptual experiences with different phenomenal characters.

47

Phenomenal character is the specific quality of what it is like to undergo a given perceptual experience. A quirk of Shoemaker's view is that phenomenal characters are understood as properties of objects, thus it is O that has the relevant phenomenal character. Shoemaker contends that the same object O can have multiple phenomenal characters in so far as O has the relevant dispositional property to cause the different perceptual experiences in subjects like A and B. Setting aside this particular feature of Shoemaker's view, we can hold instead that O has a disposition to cause states with different phenomenal characters. Hence, the property is relational, since it partly depends on the states it causes in perceivers. When we consider intrinsic duplicates in the Twin Earth scenario, we can hold that objects X and Y on Earth and Twin Earth could be such that they have different natural kind properties but cause in subjects S and T states with the same phenomenal character[16], such as that associated with our experiences of redness. On this view, we can say that objects X and Y share a dispositional property to cause states with the relevant phenomenal character in S and T, and identify this dispositional property as the relevant appearance property shared by X and Y, represented by both S and T.

Such a view could be rejected by denying that appearance properties are dispositional properties in Shoemaker's sense. Here I defend the view that these properties, or something like them, are viable, although a complete analysis of the metaphysics of properties is outside the scope of this thesis. Shoemaker's account could be rejected on the basis of holding that the property of being red is identical to the relevant natural kind property. Thus, S would represent the local natural kind property on Earth, and T would represent the relevant natural kind property on Twin Earth, hence the content would be wide. However, the view of narrow content here is not tied to Shoemaker's view of appearance properties. What the view developed so far requires is that there is a kind of mental representation that can represent superficial properties primarily based on how things seem to the subject. Since objects X and Y at least have in common the property of appearing to S and T to be the same, then something less than Shoemaker's theory of appearances would be

---

[16] Phenomenal externalists (such as Tye 2000) deny that intrinsic duplicates will be in states with the same phenomenal character. I consider this view in more detail in Chapter 2, and assume for now that phenomenal character supervenes on intrinsic properties.

enough. For instance, an alternative account is to identify a more basic notion of a disposition to cause states with a given phenomenal character. Something has the dispositional property if they trigger in S the relevant state. Thus, S's representation RED could plausibly represent a dispositional property to cause experiences of redness, and whether or not it is held that these dispositions are appearance properties or whether appearance properties are some other kind of property, the appeal to dispositional properties would stand.

3.3.4 The representation of other non-natural kinds

The narrow content associated with mental representations of other non-natural kinds like algorithms, computers, and sofas can be understood in a similar way. However, there are two issues that need to be considered for functional and artefact properties: the appropriate form of representation, and how to understand the properties such that intrinsic duplicates represent the same property. With regard to the first point, when considering the range of other non-natural kind properties, it is not obvious what form of narrow representation is appropriate. Much of the debate has focused on natural kind terms because of Putnam's example, and this has lead some internalists to the descriptive account of narrow content. However, as a general model of narrow content one worry is that descriptive representation results in intrinsic duplicates not representing any properties in common. In addition, we don't want to hold that the properties are descriptively determined at $W_1 \ldots W_n$, since this is unnecessary. A better approach would be to allow either descriptive representation or relational representation. However, I will not consider the details of the form of representation that is most suitable here. What is more important for present purposes is the second issue, relating to the properties represented, and to indicate how these non-natural kind properties support the representation of heterogeneous kinds in general.

For non-natural kind properties, we want to say that S and T represent the same properties at possible worlds, not because they are related to a dispositional property, but rather because these properties (whatever they are) are also the same at $W_1 \ldots W_n$. Consider artefacts. While particular instances of artefacts like sofas might be

composed of different natural kind properties, the property of being a sofa is not associated with any particular set of natural kind properties that instantiate the artefact type. Being a sofa is a matter of being *a long upholstered seat with a back and arms, for two or more people*. Something is a sofa as long as it has these properties. Consequently, SOFA represents the artefact type, which can be instantiated by a wide range of objects with any number of different fundamental properties. The same principle applies to functional properties like being an algorithm, which may be defined as *a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer*. The same point can be made for COMPUTER and any of many other non-natural kind terms. Returning to the case of the heterogeneous kind common to H_2O and XYZ, the point can now be made as follows. Oscar and Twin Oscar's WATER represents a common heterogeneous property because the description each associates with WATER, *the clear drinkable liquid that falls from the sky and fills the lakes and oceans* (i.e. watery stuff), picks out the same set of non-natural kind properties. As is the case with artefacts and functional kinds, being the heterogeneous kind *watery stuff* is a matter of being a kind individuated by a certain set of properties, and it is these properties that are common across $W_1 \ldots W_n$. So long as we can countenance properties like being a sofa or being an algorithm, there should be no serious concern with the property of being watery stuff.

3.3.5 The pluralist view

Before considering the variety of truth conditions associated with narrow content, it is worth outlining the pluralist view developed so far. There are three central components relating to the nature of mental representations, what they represent, and the form of representation (or the representational relation).

The claim about pluralism concerning what is represented relates to the idea that different properties may be represented by the same concept. Where some internalists focus on one kind of property only, for instance Prinz (2002) contends that narrow contents relate to the representation of superficial properties only, the case can be made that narrow contents can relate to different kinds of properties,

including superficial and natural kind properties. On the descriptive view, natural kind properties are of primary concern, since narrow contents are functions from worlds to extensions, effectively representing the X at world W which satisfies the description associated with S's narrow content. On the informational view, superficial appearance properties are dominant, since these are what tend to trigger S's concepts when she encounters instances of a relevant kind, for instance, it is the superficial watery properties common to H_2O and XYZ which would both trigger S's WATER concept. According to the pluralist view developed here, S's concept WATER can represent different properties in so far as it is comprised of a number of conceptual structures which represent different properties. For instance, prototype structure may relate to superficial properties while theory structure may relate to more fundamental properties or to a range of non-natural kind properties.

This account of how a concept may represent different properties involves the other two aspects of the view, relating to pluralism about mental representations and the representational relation. Pluralism about mental representations relates to the concept pluralist thesis, according to which concepts are comprised of a variety of concrete conceptual structures, such as prototypes and theories. This contrasts with other internalists who restrict mental representations to, say, just prototypes, or don't explore the view that narrow content is suitable for individuating a wider range of mental representations than is typically the case. While narrow content is often related to idiosyncratic concepts which can differ from the standard public concept, thereby allowing individuals to have different concepts, the extent to which these concepts relate to a specific theory of concepts has not been developed in much detail (with Prinz (1999) being the exception). For reasons identified above, narrow content needn't be restricted to one kind of conceptual structure, hence the internalist should accept pluralism. Thus, the pluralist account is not simply the standard internalist view that different subjects possess different concepts, such as Alf's arthritis concept and the expert's arthritis concept. Pluralism is the thesis that concepts themselves are pluralistic, coming in a variety of structures, and indeed some concepts being comprised of a variety of structures. This is similar to a standard internalist position, but adds an additional dimension where subjects may

have different concepts in so far as they come in a variety of structures. My claim is that internalism is a natural fit with concept pluralism, and that this has implications for the nature of narrow content when we consider the narrow content of the different kinds of conceptual structures. For instance, we therefore get pluralism about the properties represented as noted above. Hence the first and second components of my view are closely connected.

The third aspect of my account, pluralism about the representational relation, relates to the view developed in Chapter 1 that the internalist's representational relation may include a kind of descriptive representation along the lines of the Jackson-Chalmers two-dimensional account (and similar accounts from Searle (1983), Mendola (2008), and Farkas (2008)), as well as a form of informational representation along the lines of an internalist version of Fodor's informational theory of content (versions of which have been developed by Segal (2000) and Prinz (1999)). What I defend is a hybrid pluralist account of the representational relation, which incorporates at least these two basic internalist accounts of representation. The differences in these accounts of representation lend some support to pluralism about what is represented in so far as the descriptive view is perhaps best suited for representing natural kind properties, while the informational view may relate to superficial and other non-natural kind properties.

3.4 Narrow content and truth conditions

I have shown that narrow content is genuinely representational content and comes in a variety of forms. In this section I show that narrow content is semantically evaluable for truth and falsity. The worry for internalism is that truth conditional content is standardly taken to be wide. The view developed so far holds that there is only a difference in the truth conditions of intrinsic duplicates' thoughts if they represent different properties. Where the properties represented are the same, then the truth conditions will also be the same. This allows the internalist to accommodate two apparently conflicting views about the truth conditions associated with narrow content, as I will now show.

Thoughts typically reflect how the world is taken to be, and the truth of a thought depends on whether the world is that way. Consequently, it matters how a subject's thoughts, and mental representations, represent the world as being, since this is crucial for determining what the truth conditions of the thought are. It is standard to take the truth conditions of a thought, such as the thought expressed by 'water is wet', to be comprised of the contents of the representations expressed by the relevant thought, i.e. the content of WATER and WET, together with the semantic significance of ascribing the latter to the former. In the possible worlds framework we can identify the relevant truth conditions with the state of affairs or scenario in which the thought is true or false. The issue with regard to wide and narrow contents is that proponents of wide content would hold that S and her intrinsic duplicates at $W_1...W_n$ would have thoughts with different truth conditions because they represent different natural kinds. Consequently, the truth conditions would relate to the different state of affairs at $W_1...W_n$, specifically relating to scenarios involving the relevant natural kind properties represented by WATER and WET at each world. This conforms to a standard account of the truth conditions associated with thoughts. For instance, the thought expressed by 'water is wet' is taken to have the following truth conditions:

(1) 'water is wet' is true if and only if water is wet

Typically the terms occurring on the right hand side are construed as natural kind terms, so that 'water' on the right hand side represents H_2O and could not be used to properly characterize the truth condition of S's intrinsic duplicate on Twin Earth whose thought has the truth condition:

(2) 'water is wet' is true if and only if twin water is wet

On the view developed above, intrinsic duplicates may share representations that represent the same heterogeneous, appearance, and other non-natural kind properties at $W_1...W_n$, thus the same scenarios at each of $W_1...W_n$ would make the thought expressed by 'water is wet' true when thought by each intrinsic duplicate. Put

another way, the thoughts expressed by each subject have the same truth conditions. These truth conditions can be expressed as follows:

(3) 'water is wet' is true if and only if watery stuff is wet

Here 'watery stuff' is used as shorthand for the set of functional and appearance properties that S associates with the term 'water'. Thus, while S might use the term 'water' to express her representation WATER, 'watery stuff' reflects that S's WATER represents a heterogeneous kind individuated by these non-natural kind properties. Since the properties represented by S's WATER includes the property of being wet, itself a non-natural kind property, then the thought would be true at all of $W_1 \ldots W_n$ where this is the case. The truth conditions will depend on what mental representations are being considered, and hence what the narrow content of the mental representation is. For instance, on Jackson's descriptive view of narrow representation the truth conditions of S's thought expressed as 'water is wet' may be given as:

(4) 'water is wet' is true if and only if the actual watery stuff of our acquaintance is wet

Note that S's theory of water might not be as specific as to pick out a natural kind, thus, unlike Jackson, et al, it might be better to allow that while some may have representations that relate to natural kinds, others have more generic heterogeneous kind mental representations with truth conditions similar to (3) (or even the same individual may have a representation that relates to both). For heterogeneous kinds, both H_2O and XYZ can make S's thoughts true, since it is not only the local natural kinds that would satisfy this truth condition as both H_2O and XYZ have the property represented by WET. In so far as the natural kind and heterogeneous kind views result in different truth conditions, S and her intrinsic duplicates may then have mental representations with different narrow contents, which represent different properties, and have different truth conditions. The same principle applies when considering a mental representation that relates to a scientific theory of a property

like H_2O. The truth conditions of S's thought expressed as 'H_2O is not combustible' would be:

(5) 'H_2O is not combustible' is true if and only if H_2O is not combustible

In contrast to the standard account of the truth conditions of thought, the view developed here contends that we must pay attention to the specific representations expressed by the subject, and to the properties represented. A useful way of characterizing some of the truth conditions associated with narrow contents is to appeal to the notion of relativized truth conditions[17]. A paradigmatic instance of relativized truth conditions involves the use of terms which generate contents that may differ depending on who the thinker is. For instance, if you and I both think a thought expressed as 'my pants are on fire', then our thoughts have different truth conditions because the person represented by 'my' differs when we each think the relevant thought. To get the truth condition of each thought associated with the sentence 'my pants are on fire', it is necessary to relativize the sentence to the relevant thinker. In a similar manner, when identifying the truth conditions of the thought expressed as 'water is wet', it is necessary to relativize the expression to the relevant thinker's mental representations associated with the thinker's use of 'water' and 'wet'. If S thinks of the property represented by WATER as a generic watery kind with certain superficial features then the truth conditions will be less specific than if S thought of the kind as a specific chemical kind composed of H_2O. For natural kind representation, we would also relativize the truth conditions to S's actual world, thus relativizing to both the mental representation and the actual world of the thinker. For heterogeneous, superficial, and other non-natural kind properties we don't need to relativize to a specific world, only to $W_1 \ldots W_n$, and to the subject's specific mental representations.

It should be noted that recognising truth conditions associated with narrow contents does not preclude recognition of the standard externalist truth conditions. Externalism can provide alternative truth conditions allowing for different ways of

---

[17] See for instance Recanati (2007) and Kallestrup (2011a) for accounts of relativized truth conditions.

semantically evaluating thoughts and utterances. For instance, in the Twin Earth cases we may allow Oscar to have two kinds of truth conditions associated with wide and narrow contents. The narrow content could reflect how Oscar's representation represents a heterogeneous watery substance, while the wide content reflects the wide content determined by the local natural kind, H_2O. Both contents are semantically relevant, since we may semantically assess Oscar's thought from the perspective of what 'water' means in the local linguistic community (or what Oscar is actually acquainted with) and from what properties Oscar's mental representations represent understood in terms of how his representation represents the world as being. It might be objected that a subject like Oscar could not then fail to have true thoughts, since whatever Oscar is disposed to think of as falling under WATER would thereby make his thought 'that is water' true. However, the view is not at risk of making a subject's thoughts trivially true in the sense that S could not have false thoughts. A thought could be false on this view if, for instance, a subject predicated a property of an object which it did not possess. For instance, where S thinks the thought expressed as 'water is combustible' is true. In addition, we can hold S semantically responsible to linguistic conventions in her linguistic community, or to the actual kinds at her environment, thus there would be a way to show that her thoughts could be shown to be false[18].

Hence, the internalist can allow wide content to serve as a constraint on the truth of a subject's utterances and perhaps on a subject's manner of conceptualizing and representing the world. Consider Burge's arthritis case, where Alf incorrectly believes that the term 'arthritis' refers to the condition in his thigh. On this view, we can hold that wide content relates to the conventional public meaning of the term which can serve as a constraint on the correctness of a speaker's usage. Hence, with respect to linguistic conventions, Alf's utterance 'I have arthritis in my thigh' is false. Norms of correct usage therefore relate to the conventional meaning of the term, and

---

[18] Loar (1988) makes a similar point noting that wide content may relate to truth conditions that pertain to linguistic facts in S's linguistic community, and which serve as norms from which to semantically evaluate S's thoughts and utterances.

thereby to wide content. [19] The same point applies to Oscar and his use of 'water', since the wide content of the concept he expresses relates to H_2O, say, and this serves as a basis from which to semantically evaluate his thoughts and utterances.

4. The Twin Earth argument revisited

One of the most common objections to narrow content is based on the view that any attempt to express or state what narrow content is will be thwarted, since the terms we use to express narrow content, and which are also used to attribute mental representations to subjects, may always represent different natural kind properties at intrinsic duplicates' actual worlds. As a result, any attempt to say what narrow content is will result in our expressing a form of wide content that relates to different properties represented by intrinsic duplicates. This objection relies on the point, disputed throughout this chapter, that there are no terms that are not subject to the Twin Earth argument. The idea is that there is no mechanism in natural language to express any sort of content common to the representations of intrinsic duplicates. The main response considered here revolves around denying that any such problem affects the view of narrow content developed in Section 3. [20]

4.1 The Twin Earth argument and the expressibility of narrow content

Versions of this objection are abundant in the philosophical literature, having been deployed by Lepore & Loewer (1986), Block & Stalnaker (1999), and Sawyer (2007), among others. The essence of the objection is that the terms used in expressing the narrow content of representations like WATER, such as 'clear', 'thirst-quenching', and 'liquid', are all themselves natural kind terms that pick out different natural kind properties on Earth and Twin Earth. Suppose an internalist were to try to define narrow content in such a way that the narrow content of WATER was given by the relevant set of descriptive terms, and held that this captured a content common to intrinsic duplicates. This approach is common among internalists, for instance on Jackson's view 'water' represents *the actual watery stuff of our acquaintance*, where

---

[19] I assume here that conventional linguistic meaning relates to wide content, however, elsewhere in the thesis I note that on some Internalist views there is a kind of semantic content of utterances which is narrow.

[20] As noted, I will consider Burge's (1979) arthritis argument, which makes a similar claim, in Chapter 3.

'watery stuff' is shorthand for the relevant set of watery properties (clear, thirst-quenching, liquid, etc). The objection is that terms like 'liquid' would not pick out properties common to intrinsic duplicates' respective worlds because what each term picks out at that world may in fact be a different natural kind property. For instance, Block & Stalnaker (1999) consider a case where on Twin Earth 'liquid' refers to a kind most of which is a slippery granular solid, which shares the superficial features of Earth liquids but the underlying microphysical properties are radically different. The problem for a view like Jackson's is therefore that we can't use terms like 'liquid' or 'watery stuff' to identify a shared narrow content, because in using these terms we may in fact just be expressing, and attributing, different wide contents. So, for Oscar, 'liquid' represents liquids, while for Twin Oscar it represents twin liquids. Hence, where 'watery stuff' is understood as a shorthand term, we cannot rely on this neologism to express a shared narrow content, since its meaning derives from that of the relevant descriptive terms, which, the objection contends, express wide contents. A similar objection is made by Lepore & Loewer (1986) who contend that 'red' in English is a natural kind term that picks out the relevant physical property on Earth. They consider a version of the Twin Earth scenario where 'red' on Twin Earth picks out an object with the same microphysical properties of the green objects on Earth. However, due to a peculiar feature of the environment, when Twin Earthers encounter these objects they have experiences that are the same as our experience of red objects. Thus, in twin English, 'red' represents a different natural kind from that on Earth, despite intrinsic duplicates having subjectively identical colour experiences of these different natural kinds. Consequently, colour terms, and by extension appearance terms in general, would not be suitable for expressing narrow content, since they, like 'liquid', are further natural kind terms.

4.2 Responses to the argument

The response I will make holds that if mental representations (and hence the terms in natural language we use to express these representations) represent the range of properties noted in Sections 2 and 3, then the internalist can deny that 'liquid', and even 'red', necessarily function to exclusively pick out natural kind properties. In response to Block & Stalnaker (1999), an alternative account of 'liquid' is that the

term expresses a mental representation which represents a mixture of functional and appearance properties, and not the underlying natural kind properties. Thus, intrinsic duplicates' LIQUID representation may pick out the same functional and appearance properties despite different natural kind properties at $W_1...W_n$ instantiating these properties. For instance, liquids and twin liquids, by hypothesis, will appear and function in superficially the same way. Thus, the point the internalist should make is that we can appeal to non-natural kind properties, and hold that many terms which externalists hold are natural kind terms, may in fact function, in part, like typical non-natural kind terms that pick out properties common to $W_1...W_n$. It is possible that such terms have both wide and narrow contents, thus the internalist needn't be committed to the denial that LIQUID also has wide content that relates to a relevant natural kind property.

Similarly, in response to Lepore & Loewer (1986), an alternative view would be that RED may be associated with several mental representations that represent different properties. For instance, S' mental representation RED may represent a dispositional appearance property common to objects across $W_1...W_n$ that have the same disposition to cause experiences with the same phenomenal character of redness in S and her duplicates. S may also have a further mental representation associated with RED which may function to represent a relevant natural kind property, and here the content of intrinsic duplicates' representations could differ. For instance, on the descriptive model of narrow representation, it could be held that S's RED represents the local natural kind with the relevant appearance properties. The appeal to the meaning of terms in English appears to beg the question against the internalist. If we understand the content of terms like 'red' through the mental representations they express, then we can reject the apparent assumption that in English terms like 'red' are natural kind terms, since there is a plausible case that they might represent a mixture of natural and non-natural kind properties possessed by objects with different underlying natural kind properties. Consequently, we can express narrow content using terms like 'liquid' and 'red' because there is no good reason to hold that English, and the mental representations expressed by linguistic expressions, relate fundamentally to natural kinds.

I will conclude this chapter by noting how this response avoids a worry with Chalmers' account of narrow content. Recall that on Chalmers' (2003) account we understand the narrow content of WATER in terms of epistemically possible scenarios which we express by means of an idealized language that reflects the subject's a priori knowledge. A central requirement of this language is that the terms are semantically neutral, that is, they do not pick out different natural kinds at $W_1…W_n$. If this requirement can be met, then we can express the relevant epistemically possible scenarios. Sawyer (2007) and Stalnaker (2008) contend that there is no such semantically neutral language available, and hence that there is no way to express these scenarios. However, contrary to Chalmers claim, the internalist doesn't need to appeal to an idealized, semantically neutral, language. Instead, if we pay attention to the properties represented by non-natural kind terms, many of the terms plausibly pick out kinds that are identical across $W_1…W_n$. Consequently, terms like 'red', 'liquid', and so on, are immune to the Twin Earth argument, hence there is no need for the internalist to accept that we must introduce a novel language to express narrow content, since much of ordinary language appears to be semantically neutral in the required sense.

5. Conclusion

The main aim of this chapter was to show that narrow content is genuinely representational and to make the case for a pluralistic account of the nature of narrow content. The pluralistic account was motivated by reflecting on the externalist accounts of Putnam (1975) and Fodor (1994) and the descriptive and informational accounts of narrow content that are consistent with these approaches. Given the objections and limitations of each approach to narrow content considered individually, I suggested that a conciliatory view can be motivated by paying attention to the supervenience claim, the relevant mental representations, and the range of properties represented. This resulted in a range of truth conditions associated with narrow contents in so far as different properties are represented by, for example, the term 'water' or representation WATER. I also showed that a common objection to narrow content that involves redeploying the Twin Earth argument can be resisted. As noted throughout the chapter, my account relies on the

viability of pluralism about mental representations, since this provided the basis for allowing a variety of narrow contents that relate to the representation of different properties. In the next chapter I defend this view of mental representations.

Chapter 2

Concept Individuation and Cognitive Content

1. Introduction

In Chapter 1 I argued that some content is narrow and genuinely representational on the basis of a correct understanding of the internalist's supervenience claim together with an account of the form of representation, the properties represented, and the nature of the underlying mental representations. I proposed that narrow content relates to a range of natural kind and non-natural kind properties, allowing for two main forms of narrow representation. However, it was assumed that mental representations come in a variety of forms, thus in Chapter 2 I develop and defend the account of pluralistic mental representations in more detail.

Concept pluralism has recently been defended as the most plausible account of mental representations given work in empirical psychology which proponents argue gives rise to a pluralist account of conceptual structure and an account of the ontology of concepts as concrete mental representations. Concept pluralists reject traditional monistic theories of concepts which treat concepts as having just a single structure, as well as accounts of concepts construed as abstract entities. This development has an important impact on the psychological externalism/internalism debate, since many of the most prominent accounts in the literature defend either an abstract ontology of concepts and/or a monistic theory of conceptual structure. The possibility of an internalist construal of concept pluralism would thereby provide a solid basis for defending the psychological internalist thesis if it could be shown that the range of structures could be individuated by their narrow contents. One potential barrier to an internalist account of concept pluralism comes from a preference for externalism by leading proponents of pluralism. Laurence & Margolis (1999)

propose that Fodor's (1994) theory of content could serve as an account of the relevant content responsible for the individuation of concepts. In addition, Weiskopf (2007) raises two objections to internalism: the first is that narrow content is always subject to the Twin Earth argument (as noted in Chapter 1); the second relates to the inability of narrow content to play the role of cognitive content. Against these externalist construals of pluralism, in this chapter I will develop and defend an internalist account of concept pluralism.

The structure of the chapter is as follows. In Section 2, I motivate pluralism by raising two lines of objection to existing externalist and internalist accounts of concepts. First, concepts are typically treated as concrete mental entities in the psychological literature, and abstract accounts appear to be incapable of explaining the relevant data that a theory of concepts must explain. Second, given the data to be explained, monistic theories that treat concepts as a single type of structure are similarly unable to handle all the relevant data. This leaves pluralism, and the remaining question is whether it should be understood as fundamentally internalist or externalist. To decide this question we need to know how concepts are individuated. I begin the positive case for internalism about general concepts in Section 3 by drawing on the account of narrow content from Chapter 1. In particular, I argue that two of the main conceptual structures, prototype and theory structure, may be individuated by narrow content. In support of this account of individuation in psychology I consider and reject an argument designed to show that internalist individuation is incompatible with psychological generalisations. I then show, in Section 4, why we should take the cognitive content of concepts to be narrow. First, I respond to Weiskopf's argument that cognitive content is identical to a kind of wide content. I then show that the cognitive content of prototypes should be understood in terms of the perceptual representations encoded in perceptual templates. I argue that the relevant cognitive content of these perceptual templates derives from the phenomenal character of relevant perceptual representations, which in turn supervenes on internal states. The basis of cognitive content therefore relates to the thesis of phenomenal internalism. I defend this account of cognitive content against two objections. First, I consider an objection that might be raised in relation to Tye's

case for phenomenal externalism, and second I consider whether the resulting notion of narrow cognitive content is too conceptually sophisticated to play the role of cognitive content.

2. Concepts and the psychological externalism/internalism debate

In Section 2 I provide the basis for an internalist account of pluralism by highlighting problems with existing externalist and internalist accounts of concepts, and rejecting two externalist conceptions of pluralism. I begin, in Section 2.1, with an overview of the issues relating to the individuation of concepts and the externalism/internalism debate. In Section 2.2 I develop two basic arguments against existing externalist and internalist accounts of concepts. The first argument makes a case based on the ontology of concepts. The idea is that concepts should at least be considered as concrete mental entities, since this is how they are treated in empirical psychology. Therefore, if we adopt an abstract conception in addition to the concrete conception, we must account for the concrete conception. This raises the question of why we should also endorse the abstract conception. The second argument relates to the structure of concepts as required to explain all the data that an adequate theory of concepts should be able to account for. I contend that accounts from externalists such as Fodor, as well as internalist's like Prinz, should be rejected on the basis of being inadequate, as they can't handle all the data a theory of concepts should account for. Finally, in Section 2.3, I consider and reject two bases for an externalist account of pluralism from Laurence & Margolis (1999) and Weiskopf (2007; 2009).

2.1 Concepts, content and the externalism/internalism debate

Concepts are often construed as the basic units of thought. For instance, the thought expressed by 'the cat caught the bird' is partly composed of the concepts CAT and BIRD. What concepts are and how they relate to what they represent is a vexed question, with a wide range of accounts in the philosophical and psychological literature. We can identify two basic questions that a theory of concepts should explain. First, there is an ontological question which asks what kind of entities concepts are. On one view, concepts are abstract entities of some kind, with the same concept being shared or grasped by a number of different individuals. On

64

another view, they are concrete entities, with each individual having their own concrete token concept. Second, there is the question of what kind of structure concepts have. For instance, on one view, the concept CAT can be understood as a complex concept comprised of further concepts, such as TAIL, FURRY, FELINE, and so on. On another view, they lack constituent structure and are atomic, so the concept CAT is primitive in the sense of not being composed of further concepts. I will set these two issues regarding ontology and structure aside for now.

A different set of questions relate to the content of concepts. The representational content of a concept can be distinguished from its cognitive content. Representational content is associated with what concepts are about or represent while cognitive content is associated with how concepts represent what they are about. Consider the concept CAT. The representational content is plausibly the property of *being a cat*, while the cognitive content reflects the manner in which that property is represented, such as how an individual thinks of cats as being. The question of content raises an important question about when two concepts are the same or different, that is, how to individuate concepts. For instance, on the concrete view of concepts, two subjects, S and T, may have different token CAT concepts. We then need some criteria for individuating their concepts. Concepts are typically taken to be individuated by what they represent, thus, we may individuate the concept CAT by its representational content, i.e. by the property represented.

A final distinction can be made between wide content and narrow content, which is intimately related to the issue of individuation and the representational content/cognitive content distinction. As was noted in Chapter 1, wide content supervenes in part on factors specific to the actual environment of an individual, thus intrinsic duplicates on Earth and Twin Earth will have concepts with different wide contents. Narrow content, on the other hand, supervenes on internal states. With these distinctions in place, it is possible to understand the psychological externalism/internalism debate as arising in part by how representational content/cognitive content and wide content/narrow content intersect. Representational content is often held to be wide content, and externalists contend

that intrinsic duplicates considered at worlds $W_1 \ldots W_n$ will have different wide contents, and therefore will be ascribed different concepts. However, as I argued in Chapter 1, wide content does not have sole claim to being genuinely representational content, thus an internalist will contend that some representational content is narrow content. On this view, where intrinsic duplicates have the same narrow content, they could thereby be ascribed the same concepts. Psychological externalism and internalism can therefore be seen to come into conflict with regard to the individuation of concepts by virtue of their different treatments of content. Thus, an account of the content of concepts is of central importance for resolving this debate.

While externalists reject the idea that narrow content is genuinely representational, there is less consensus among externalists as to whether cognitive content is wide or narrow. Recall that cognitive content reflects how a property is represented as being. Some externalists hold that while concepts have, and are individuated by, representational content which is wide content, they may also have cognitive content that is a kind of narrow content. For instance, Laurence & Margolis (2007) allow for the possibility that while concepts are individuated by their wide content, they may have narrow content which plays the role of cognitive content, where narrow content is construed in terms of the narrow conceptual role of a concept. Conceptual role semantics has been proposed as a fully-fledged internalist theory of conceptual content by Block (1986) and Loar (1988). On this view, a concept's content is defined in part by its conceptual relations to other concepts. For instance, the content of CAT would be given by the concept's relations to other concepts, such as TAIL, FURRY, and FELINE. Furthermore, some externalists have expressed acceptance of the idea that there is a kind of narrow content that relates to the phenomenal character of perceptual experience (for instance Williamson, 2000). Internalists are divided about what to say on the issue of representational content and cognitive content. For instance, psychological internalists who take narrow content to be a form of representational content defend a more significant role for narrow content than just accounting for cognitive content. On one view, concepts may be individuated by narrow content that is representational content, and the cognitive content of concepts is also a variety of narrow content. Such views do not

necessarily reject wide content. What the psychological internalist denies is that wide content individuates the relevant class of concepts or plays the role of cognitive content.

As the range of views outlined so far suggests, there is considerable disagreement about how to understand concepts and the notion of content. Consequently, in so far as both sides of the externalism/internalism debate disagree about the fundamental nature of concepts, there is no simple way to decide whether externalism or internalism is true, given the different accounts of the basic units of thought. To this end, in the next section I outline two lines of argument that raise problems for existing externalist and internalist accounts of concepts. The purpose of raising these objections is to narrow the scope of the subsequent discussion and case for psychological internalism by showing that pluralism is the best account of concepts. With this foundation, I will then focus, in Sections 3 and 4, on showing that pluralism is consistent with individuation by narrow content, and that narrow content is best suited to playing the role of cognitive content.

2.2 The ontology and structure of concepts

It is possible to identify two dimensions along which a theory of concepts must be adequate. The first is that the theory must get the ontology correct, that is, the theory must tell us whether concepts are abstract or concrete entities of some kind. The second dimension relates to the structure of concepts, which tells us whether concepts have one kind of structure (monism), no structure at all (atomism), or a variety of structures (pluralism or hybridism[21]). From these considerations about ontology and structure it is possible to identify two basic lines of objection to several prominent accounts of concepts in the externalism/internalism literature.

---

[21] Vicente & Martinez-Manrique (2014) defend an alternative to pluralism, hybridism, according to which concepts are identified as a single entity, whereas pluralists prefer treating concepts as a range of entities. The difference between pluralism and hybridism is not important for my purposes, and my view can allow that concepts are either pluralistic or hybridistic. My aim is to show that on either view which treats concepts in terms of the variety of structures from psychology, we can individuate general concepts in terms of narrow contents.

2.2.1 Abstract Vs concrete ontology of concepts

Typically, theories of concepts are of two kinds: on one view, concepts are abstract entities where the same concept is grasped or ascribed to all members of a population. On another view they are concrete entities, normally taken to be realized in the neural architecture of the brain. The basic case for construing concepts as concrete entities is that empirical work on concepts treats concepts in this way, and does so in order to account for data stemming from experiments concerning concepts and categorisation. Both externalists (e.g. Fodor 1998) and internalists (e.g. Prinz 2000) defend views of concepts as concrete entities, and are motivated by issues that they understand as being of concern to psychologists.

As an example of the concrete view, concepts have been identified with prototypes by psychologists (Barsalou 1987) and philosophers (Prinz 2000). Prototype concepts are sets of representations of features or properties that are stored in memory. Prototypes were introduced in part to account for data in categorisation tasks (the Typicality Effect experiments) where subjects classified various kinds of objects based on their judgements of the typical superficial properties of a given category such as the category *bird* (Laurence & Margolis 1999). Depending on which properties are judged to be typical of a category, different members of a category can be judged to be more or less typical of the focal category. For instance, some birds will have more of the typical properties associated with an individual's concept BIRD than others. Suppose being able to fly is a typical property of birds. In this case, robins may be more likely to be classified as birds than penguins are. Prototypes are often taken to provide a basic or default concept of a category used in everyday conceptual activities, by virtue of the way in which quick judgements about category membership prevail over more measured judgements about fundamental properties. Thus, the construal of concepts as concrete arises from considering how subjects make classificatory judgements, which involves drawing on information from memory to classify objects on the basis of their typical superficial properties. On this view, two subjects, S and T, may have similar dispositions to classify robins, but differ in the extent that they would classify penguins as birds. This could provisionally be accounted for by differences in the information stored in memory.

Where concepts are construed as abstract entities of some kind, there is a worry about what to say about empirical work and the concrete view of concepts. For instance, Burge (1979) defends an abstract account of concepts, and his view faces this question[22]. On Burge's view, concepts are identified with Fregean senses. Fregeans make a distinction between the sense of an expression which is objective, public, and grasped by different subjects, and ideas which are subjective, private, and specific to each subject. Burge (1979) specifically rejects the view that we can identify a subject's concepts with the subjective conception of a category as determined by dispositions to classify objects[23]. Therefore, the approach to concepts in psychology creates a problem for philosophers like Burge who treat concepts as abstract entities. In particular, the connection between the subject's concepts and their judgements of category membership based on retrieving information stored in memory is lost. The main worry then is that in treating concepts as abstract, empirical work in psychology must in some sense be considered as not really revealing conceptual nature. The problem here is that the data provides very good reason to at least treat concepts as in one important sense concrete entities of some kind.

One possible line of defence for proponents of the abstract view is noted by Laurence & Margolis (1999; 2007) who raise the possibility of construing concepts as both abstract and concrete entities (call this the 'dual view'). Concrete entities would be identified with concepts as construed in psychology, whilst abstract entities resemble something akin to Fregean senses. The dual view is initially attractive, as it holds out hope of accommodating empirical work on concepts as well as Fregean intuitions about the distinction between public senses and an individual's ideas. However, there is a simple line of reasoning which can be used to urge the rejection of the dual view. On the one hand, a proponent of the abstract view could deny the concrete conception relates at all to concepts, as Burge seems to do. However, this has the unfavourable implication that they would thereby deny that empirical psychology has any bearing on the fundamental nature of concepts, which is not an attractive

---

[22] See also Peacocke (1992) for a version of the abstract view.
[23] I highlight problems for Burge's (1979) argument due to this claim in Chapter 3.

position. On the other hand, they could endorse the dual view as outlined above, where there are both abstract and concrete concepts, and provide an account of how they would combine. The main problem with giving preference to the abstract conception is that there needs to be some convincing reason why we should allow two kinds of mental entities rather than one, and why the abstract conception is required at all if, as proponents of the concrete view will contend, the concrete view is sufficient for a theory of concepts. Proponents of the dual view would then need to explain why we can't treat concrete structures as the relevant mental entity, and perhaps deflate or reject Fregean senses altogether when discussing issues about concepts. This places an explanatory burden on the proponent of the dual view to explain why we should allow the abstract view to take central place within a theory of concepts. While this doesn't provide a knock-down objection to abstract accounts of concepts, the treatment of concepts in psychology puts proponents of the abstract view under pressure to explain why we shouldn't just look to psychology for the theory of concepts.

I return to this issue of abstract and concrete views of concepts below, where I suggest that the two-component theorist who endorses wide and narrow contents may associate these with the abstract and concrete construal of concepts, and that we can identify two distinct roles for each account.

2.2.2 Monistic Vs pluralistic conceptual structures

As noted above, theories of concepts can also be distinguished in terms of their structure. On one view, concepts have no significant structure at all (atomism), while the traditional approach is to treat concepts as having a single structure (monism). Monism has historically been the standard view of conceptual structure, with the various proposals about concepts revolving around a debate as to which structure is the correct structure of concepts. A third view, concept pluralism, rejects atomism and monism and contends that concepts should be understood in terms of a variety of structures, each of which can play different roles[24].

---

[24] It should be noted that some philosophers who appeal to the structures from psychology conclude that we should give up on the notion of a concept. For instance, Machery (2005) contends that, given

Pluralists (for instance Laurence & Margolis 1999; Weiskopf 2009) make two main points about the data arising from the empirical literature. The first point is that each of the monistic theories developed in the literature are individually inadequate because they cannot explain all the data. The second point is that if we hold that a theory of concepts should be able to explain all the data, then we should construe concepts as being comprised of a range of conceptual structures which collectively can explain the data. The basic argument is therefore that a theory of concepts must be able to account for all the data, however, since monistic theories are unable to do this, concepts should be identified with a variety of structures which are jointly necessary and sufficient for a theory of concepts. Some of the structures typically included within pluralism relate to theories which psychologists and philosophers have previously proposed as monistic theories, for instance Weiskopf notes that:

> …concepts have been identified with prototypes, bundles of exemplars, theory-like structures of some sort, perceptual 'proxytypes', and unstructured atomic symbols. (Weiskopf 2009, 1)

As noted above, prototypes were introduced to account for judgements of category membership based on superficial features. However, prototypes alone are unable to account for different ways in which categories can be conceptualized, and psychologists have introduced a range of other structures. The second structure worth considering relates to more complex conceptualizations of categories in terms of fundamental properties not represented by prototypes. Theory-theory (Laurence & Margolis 1999) construes concepts in terms of theory-like structures which allow for a far richer conception of a relevant kind, being able to encode information about more fundamental, perhaps hidden, properties of a category. What pluralists like Laurence & Margolis (1999) and Weiskopf (2007; 2009) contend is that in construing concepts as having pluralistic structure, we can combine the benefits of each of the proposed theories in the psychological literature, and avoid their individual limitations. Taking this basic line of reasoning in favour of pluralism, we can use this to show that several existing externalist and internalist accounts of

---

the different roles played by these structures, there is no useful notion of a concept. Pluralists and hybridists, however, contend that we may still make use of the notion of a concept. I set this debate aside here as it is outside the scope of this thesis.

concepts as concrete entities fail to be adequate in so far as they are committed to atomic or monistic structure. I will briefly consider three prominent accounts in the literature to demonstrate the advantages of pluralism.

On Fodor's (1998) theory, concepts are atomic entities in a Language of Thought, and lack constituent structure. On this view, concepts are individuated by their representational content, which for Fodor (1994) is a variety of wide content. A similar objection as was raised against the abstract view can be raised against atomic concepts. While psychologists have treated concepts in terms of a variety of structures to account for different phenomena, this places the onus on Fodor to either explain the data via atomic concepts, or show how atomism is compatible with explaining the data in some other way. Considering each of these issues is outside the scope of this thesis, however, the inability of atomic concepts to account for typicality effects suggests that Fodor's account is at best incomplete[25]. As noted above, Loar (1988) has held that concepts can be understood in terms of their conceptual role. The conceptual role of a concept consists in its conceptual relations to other concepts, so that the conceptual role of the concept BIRD would be given by its conceptual relations to WING, BEAK, FEATHER, and so on. One worry here is in determining what the conceptual role of a given concept is, and hence with how to individuate concepts on this approach (Prinz 2000). Whereas Fodor's theory individuates concepts by their representational content, which allows for individuation by what the concept represents, according to conceptual role theories we must individuate by a concept's relations to other concepts. However, specifying this is a notoriously complex issue, and may lead to a highly abstract specification of a concept. Finally, as was also noted above, Prinz (2000) has defended an internalist account of prototypes. To recap, prototypes represent superficial properties and are motivated by the typicality effect experiments. In endorsing a monistic account of concepts, however, Prinz is open to the pluralist's objection that we need other kinds of structure to account for all the data. In particular, in restricting concepts only to prototype structure, Prinz's account seems unable to account for the representation of

---

[25] Laurence & Margolis (1999) provide a detailed treatment of the various objections to atomism and other monistic accounts of concepts.

fundamental properties, as was noted in Chapter 1. I have not intended to provide a knock-down argument against these accounts of concepts. Instead, I aim to highlight the basic motivation for concept pluralism, and to draw attention to its benefits over other accounts, as well as to the advantages in being able to draw on a wide range of conceptual structures. It is this diversity in conceptual structure which provides the basis for a novel account of psychological internalism as I will show in Section 3.

2.3 Pluralism and externalism

The objections to existing externalist and internalist accounts relating to ontology and structure provide prima facie support for concept pluralism. Assuming for now that pluralism provides the best basis for a theory of concepts, the issue at stake is whether pluralism is fundamentally an externalist or internalist theory of concepts. The pluralists discussed so far take pluralism to be externalist. However, their reasons for endorsing externalism can be resisted, as I will now show. The first case rests on a rejection of narrow content. Weiskopf (2007) contends that pluralism is fundamentally externalist by providing two objections to narrow content. The first objection is the Twin Earth argument, which was rejected in Chapter 1. I will return to this argument below. The second argument relates to the individuation of primitive concepts, and I will consider this objection in Section 4 below after I have developed the internalist account of pluralism in more detail. A positive case for an externalist construal comes from Laurence & Margolis (1999) who take concepts to be individuated by Fodor's (1994) theory of wide content. However, in Chapter 1 I noted that an account of narrow representation can be developed in keeping with Fodor's general approach to content. On Fodor's account, wide content relates to the representation of properties at possible worlds. What S's concept C represents on Earth is determined by what would trigger C, in line with the asymmetric dependency constraint which rules-out mere lookalikes that trigger C from being represented by C. What purportedly makes such content wide is that we are restricted to considering a certain proscribed set of counterfactuals, namely what has actually caused C to trigger, or what would cause C to trigger at worlds where the relevant natural kind property is the same as S's actual world. However, on the relational view of narrow representation, we consider what did or would cause C to

trigger at worlds $W_1...W_n$, i.e. the set of nomologically possible worlds that are superficially, though perhaps fundamentally, different. Another problem for Fodor's account is that it seems to entail that intrinsic duplicates never represent the same property. Against this, it is plausible that the property of being able to fly is a functional property and that two creatures on Earth and Twin Earth with different natural kind properties may share this functional property. Thus, when we consider a subject S on Earth and her intrinsic duplicate T on Twin Earth, S's concept FLIES would represent a property that is also instantiated on Twin Earth, and vice versa for T. This is all that is required for there to be a kind of content shared by S and her intrinsic duplicate T, which meets the definition of narrow content. Consequently, where we individuate concepts by representational content, S and T may share the concept FLIES, since the concepts have the same representational content, namely, the property *flies*.[26]

3. Concept pluralism and internalist individuation

In Section 2 I outlined two objections to externalist and internalist accounts of concepts. I also suggested how to respond to the externalist accounts of pluralism from Laurence & Margolis (1999) and Weiskopf (2007). In Section 3 I provide a positive case for an internalist account of concept pluralism. In particular, I show how prototypes and theory structures may be individuated by two kinds of narrow content identified in Chapter 1. I then consider how the account provides a response to an objection which questions the consistency of internalist individuation with psychological generalisations in psychology.

3.1 Pluralism and internalist individuation

As noted in Section 2, some concept pluralists treat pluralism as fundamentally externalist. Where concepts are individuated by their representational content, the externalist equates representational content with wide content. Consequently, since wide content supervenes in part on external factors specific to the actual environment of subjects, intrinsic duplicates S and T on Earth and Twin Earth would be ascribed

---

[26] Although there are other accounts of wide content available, a complete review of these would require more space than is available here.

different concepts. However, in Section 2 I noted that Laurence & Margolis (1999) appeal to Fodor's (1994) account of wide content, and a problem with Fodor's theory was that it doesn't seem to rule-out narrow content. Rather than consider other accounts of wide content, in this section I will provide a positive case for construing pluralism as being consistent with internalist individuation with respect to general concepts. . If it can be shown that general concepts comprised of prototype and theory structures can be individuated by narrow content consistently with the approach to individuation of these structures, then the internalist does not need to rely on rejecting each possible account of wide content.

Before showing that prototypes and theory structures have, and are individuated by, narrow content, it will be useful to revisit the accounts of narrow representation developed in Chapter 1.

In Chapter 1 I identified two main forms of narrow representation: informational (or relational) and descriptive. It was noted that appearance properties relate to the relational view of narrow representation, such that S and T represent the same property by virtue of information-like relations to the relevant property. The relational view was also taken to be suitable for representing heterogeneous kind properties such as a relevant property that subsumed H_2O and XYZ. For these properties, attention was paid to the relevant counterfactuals. In particular it was held that we need to consider the actual or possible triggers of S's concept C at worlds $W_1 \ldots W_n$, since the restriction to what would cause C to trigger at S's actual world was not sufficiently motivated. In contrast, the descriptive view of narrow representation appealed to the idea of what a description would pick out at $W_1 \ldots W_n$. The descriptive view was found to be suitable for representing natural kinds, since we consider what natural kind is represented at S's actual world, and a version of the descriptive view can allow for the representation of non-natural kind properties where the relevant properties are the same at $W_1 \ldots W_n$. For psychological internalism to be true, therefore, it must be the case that the individuation of prototype and theory structures is consistent with supervenience* and the accounts of

narrow representation. To this end, in the next two sections I show how the accounts of narrow content from Chapter 1 sit comfortably with prototype and theory structure.

3.2 The individuation of prototype structure

Turning first to prototypes, recall that prototypes encode information about superficial properties, which are the properties that subjects base their judgements of typicality on. Prototypes can therefore be construed as representing a category, such as *bird*, by virtue of the representation of a set of superficial properties. We can then say that prototypes are individuated by the superficial properties they represent, and which subjects' judgements are based on:

(Prototype structure)   Prototype structures are individuated by sets of superficial properties

What the internalist needs to show is that there is a notion of narrow content which can relate to the representation of superficial properties so that prototypes are individuated by such narrow contents. At this point an externalist may appeal to the Twin Earth argument against narrow content, as Weiskopf (2007) does. The Twin Earth argument is that the properties at issue are potentially natural kind properties, and hence that the superficial properties like being wet, red, and so on, at Earth and Twin Earth, relate to the different underlying natural kind (i.e. microphysical) properties at each world. Thus, while prototypes represent superficial properties, the relevant superficial properties would be construed as being further natural kind properties, so that the prototypes of intrinsic duplicates would represent different properties, and the relevant representational content would be wide content. Consequently, an internalist needs to be very clear about what superficial properties are in order to show why they relate to narrow content. This issue is not addressed by Prinz (2000) who has argued that prototypes are individuated by the appearances they function to detect. Prinz uses this as a basis for the claim that intrinsic duplicates share prototypes because they detect the same appearances, however, we need more from such an account in order to avoid the externalist's appeal to the Twin Earth argument.

Focusing for now on appearance properties represented by prototypes, in Chapter 1 I suggested that appearances are plausibly non-natural kind dispositional properties, and that we can identify a notion of narrow content that relates to the representation of such properties. In particular, I suggested that following Shoemaker (2006) we can understand appearance properties in the following way[27]. Take an object O and subjects A and B with different perceptual systems. Suppose that A and B are such that when A perceives O she has an experience associated with our experience of redness, and when B perceives O she has an experience associated with our experience of greenness. O's fundamental properties are fixed, but the appearance properties depend in part on how O appears to perceivers like A and B. O has the disposition to cause in A and B experiential states with different phenomenal characters, that is, the way things appear perceptually to A and B is different. Shoemaker advances a particular view of phenomenal character, holding that the same object O can have multiple phenomenal characters in so far as O has the relevant dispositional property to cause relevant experiential states in subjects like A and B. However, it is not obligatory to accept this account of phenomenal character, and here I assume a view of phenomenal character in which it relates to the relevant property of a subject's perceptual experience that relates to what the experience is like for the subject. I say more about phenomenal character in Section 4 below.

Applying this view of appearance properties to the Twin Earth scenario, when we consider intrinsic duplicates S and T we can hold that objects X and Y on Earth and Twin Earth could be such that they have different natural kind properties but cause in S and T states with the same phenomenal character. On this view, we can say that objects X and Y share a dispositional property to cause states with the relevant phenomenal character in S and T, and identify this dispositional property as the relevant appearance property shared by X and Y. This would provide a suitable account of an appearance property that can avoid the Twin Earth argument, since the

---

[27] In Chapter 1 I noted that Kriegel (2008) develops a view of narrow content in which it relates to response-dependent properties, and these are similar to the dispositional properties I developed from Shoemaker (2006). There are two main differences between my view and Kriegel's. First, we characterise the dispositional properties differently. Second, and most importantly, while Kriegel restricts narrow content to these properties, I develop a pluralistic view which allows narrow content to relate to the representation of natural and non-natural kind properties.

same dispositional property is represented on Earth and Twin Earth, so the externalist cannot say that the superficial properties represented by prototypes could just be further natural kind properties and that S and T will represent different properties.

What remains to be seen is whether these properties are consistent with how prototypes are individuated. The following example serves to highlight the central issue at hand. Suppose a given object O belonging to category C has the superficial property of appearing red. Where a subject S judges that O belongs to C on the basis of O appearing to be red, the question is whether the theorist should treat this property as being a natural kind or a non-natural kind dispositional property. Answering this question provides a plausible basis for an account of what property S's prototype represents, since it is that property which S bases her judgements on.

The basic case in favour of S making her judgement based on a dispositional property can be motivated as follows. When S encounters an object O that she is to classify, she makes her classificatory judgement based on the appearance properties of O since, in the first instance, it is the appearance of O that would be of relevance to S when she makes her judgement. Consider the alternative position that S bases her judgements on O's fundamental properties. Here, S would have to be aware of what these properties were, which may require further investigation. However, prototype structures are taken to relate to quick judgements of category membership, and hence S is not in a position to base her judgements on fundamental properties.

Following the view outlined above, appearances are dispositional properties, so it is the dispositional property which is relevant to the individuation of S's prototype. This suggests that a theorist would be able, in principle, to attribute the same prototype to S and T despite their being at worlds with different natural kind properties, since they base their judgements of category membership on the same superficial properties. So long as we can construe superficial properties along the lines of these dispositional appearance properties, then the approach to individuation from the typicality experiments is consistent with internalism.

Further examples support this account. Judgements relating to the category *bird* may plausibly involve the superficial properties such as having wings, feathers, and a beak. These could in principle relate to further superficial properties relating to appearances. Where S makes her judgements of category membership on the basis of these appearance and functional properties, we can identify narrow content with this set of properties, and so S's prototype may be individuated by the associated narrow content. This view is unaffected by the Twin Earth argument. One version of the argument might run as follows. Take feathers: the externalist may contend that the concept FEATHER is a natural kind concept in so far as this relates to the representation of a property specific to a species of bird native to Earth. The bird-like creatures on Twin Earth, that look and act the same, are not birds, and they don't have feathers. Instead, they have *twin feathers*, which look and feel just like feathers, but have different fundamental properties. It is important to note that the internalist needn't deny that the creatures are different species and have different fundamental properties. What the internalist holds is that feathers and twin feathers share superficial properties that relate to their appearance properties, and that S and T on Earth and Twin Earth plausibly have prototype concepts that represent the same set of superficial properties. S and T may also have wide contents that relate to these different creatures, such as when they deploy singular concepts like THAT BIRD or THAT FEATHER. Consequently, since prototypes relate to superficial properties, and these do not seem to be subject to the Twin Earth argument, prototypes are consistent with internalist individuation.

3.3 The individuation of theory structure

There are several problems for an internalist who commits only to prototype structure. First, concept monism comes with the general problem of being unable to account for all the data a theory of concepts should ideally handle. Second, where prototypes only represent superficial properties, it would seem that the internalist would thereby deny that narrow content can relate to the representation of natural kind properties. As noted in Chapter 1, some internalists, such as Prinz (2000) and Kriegel (2008), appear to restrict narrow content only to superficial properties. On this view, intrinsic duplicates can represent the same superficial properties, but their

narrow contents cannot relate to the representation of fundamental properties like H_2O or XYZ. This is a large concession to externalism. An advantage of pluralism is that by introducing further conceptual structures like theory structure, which relate to the representation of fundamental properties, there is scope for a more extensive internalist position that would allow narrow content to relate to the representation of a variety of fundamental properties.

In Section 3.2, I outlined the informational and descriptive views of narrow representation. These were taken to be suitable for representing both natural kind and heterogeneous kind properties. The problem identified in Chapter 1 for existing views was two-fold. First, I suggested that we should permit both forms of representation, allowing different mental representations to be suitable for representing natural kinds and heterogeneous kinds, and hence that the internalist should not restrict narrow content to one or the other. Second, I noted that existing accounts fail to provide adequate support for the psychological basis for this content and the relevant mental representations. In particular, I noted problems for Chalmers' (2003) accounts that relies on the notion of epistemic space, and Jackson's (2003a) which holds that subjects like Oscar in the Twin Earth scenario possess natural kind concepts of some form that represent the local natural kind. Thus, in the rest of Section 3.3 I show how the two forms of narrow representation are consistent with the notion of theory structure, and hence that this provides a plausible basis for the view developed in Chapter 1.

To recap the descriptive view, narrow content provides a rule or condition that picks out the property represented. On this view, narrow content operates as a determiner on what is represented at a context. For instance, on Jackson's (2003a) account, the narrow content of WATER can be expressed as 'the actual watery stuff of S and her colinguals' acquaintance'. The narrow content of intrinsic duplicates' concepts will be the same, however, considered at their relevant contexts, i.e. their respective actual worlds, the description functions to pick out different kinds which play the role of watery stuff at these worlds, i.e. H_2O on Earth and XYZ on Twin Earth. As noted in Chapter 1, what was lacking on descriptive views was a convincing

treatment of narrow content that would satisfy the externalist who denies that our concepts have such content. The worry relates to both the nature of the descriptions, and the two-dimensionalist's treatment of such content. On Jackson's (2003a) view, the narrow content of S's WATER concept picks out H_2O by virtue of the narrow content functioning to pick out *the actual watery stuff of S and her colinguals' acquaintance*. The main worry is with regard to where these descriptions come from and why we should believe that concepts encode this sort of descriptive information.

An important question is why we should think that a subject's concepts have a kind of content that relates to descriptive representations at all. In particular, linking a subject's WATER concept to content that relates to the notion of one's actual world, seems to require a high level of conceptual sophistication on the part of the subject, or reliance on such a subject having appropriate representational intentions to pick out a unique natural kind. It is a substantive thesis what kind of descriptive information, if any, our concepts encode, and the mere fact that we can identify a description that would work in this way is not the same thing as identifying a genuine kind of content of normal subjects' concepts. There is a worry then that the view can appear to be ad hoc since the descriptive components are introduced specifically to save the internalist's view that there is a kind of content that is narrow and relates to natural kind representation. What would support the proponent of a descriptive account of narrow content would be some reason to think that concepts either explicitly or tacitly encode descriptive information that can function in this manner to pick out natural kind properties. To this end, in the rest of Section 3.3 I will outline how theory structure may relate to two kinds of narrow content in a way which sits comfortably with the construal of theory structures within concept pluralism.

In Section 2, theory structures were described as theory-like entities that encode information about essential or fundamental properties of categories. The representational content of theory structures should therefore be understood in terms of the relevant fundamental properties that they function to represent. Consequently,

theories should be individuated by these fundamental properties. We can gloss the individuation claim as:

(Theory structure)   Theory structures are individuated by fundamental properties

I will focus on natural kind and heterogeneous kind properties which are plausibly a variety of fundamental property. It is possible to identify two ways in which a plausible internalist account of theory structure can facilitate the representation of these properties:

(1) Descriptions that pick out natural kind and heterogeneous kind properties
(2) Relational representation of heterogeneous kind properties

One way to motivate the descriptive account is to understand how theory structures are taken to represent. A common approach is to understand theory structures in terms of bodies of knowledge, and to liken them to theoretical terms (Laurence & Margolis 1999). A second approach is to see them as relating to subjects' conceptualization of categories in terms of essential or fundamental properties, which are typically hidden or unknown. For instance, Medin & Ortony (1989) introduced a version of theory structure to account for the tendency to conceptualize certain kinds on the basis of such essential properties, even where knowledge of the specific essences is unknown. Examples are easy to find, such as where knowledge that the watery stuff on Earth is composed of H_2O has not been acquired, or where the specific essential properties of tigers is unknown, yet it is believed that both water and tigers have some kind of essential properties.

With this background in place, a first pass at S's theory associated with WATER, or BIRD, might be *the X with hidden properties that explain the superficial properties*. Here, S links the hidden properties to the superficial properties of which she is aware, and relates the former to the latter. Put this way, the account does not obviously support the representation of a natural kind. To get this, we would treat 'the X' in terms of *the unique local X,* or *the actual X of my acquaintance*. The motivation here

might be that we can understand S as tacitly intending to represent the unique X's around her, at her world (as suggested by Jackson 2003a). Thus, duplicates will have the same intention to pick out the X around them, and this will then relate to different X's in so far as 'the actual' functions as a reference-determiner at a context. However, this account requires support for the claim that subjects typically intend to represent a unique X. To believe that there is a unique hidden essence is to already attribute tacit beliefs that correspond to the representation of a local natural kind. This is partly why the relevant reference-determining component was incorporated within the narrow content, since the motivation was to explain the externalist's intuition that a unique natural kind was represented at each of $W_1...W_n$.

Suppose we ignore the relevant reference-determining component. It would then be plausible that what is directly encoded within a subject's theory is merely that there are essential properties that explain the superficial properties of which the subject is acquainted. If this is the case, then the description would fail to pick out a unique natural kind. This is because the same superficial properties considered across worlds $W_1...W_n$ relate to kinds with different hidden essences, for instance H_2O, XYZ, or birds and twin birds, and so on. This would lend support to the view that certain subjects have token WATER theory concepts that represent a heterogeneous kind, which would subsume H_2O, XYZ, and any other kind with relevant superficial properties. Where S's theory structure represents a heterogeneous kind, both forms of narrow representation will get the desired result. For instance, where S's theory structure for WATER relates to certain unknown essential properties, the property represented can be determined either descriptively, so that whatever X at $W_1...W_n$ satisfies the description is represented, or relationally, so that whatever triggers WATER at $W_1...W_n$ is represented by WATER.

This potential proliferation of concepts, and forms of representation, should not be a concern to the concept pluralist. Where prototypes and theory structures relate to the information encoded in memory, we should allow that there could be conceptual and representational variation between subjects and kinds of conceptual structures. Only once there has been adequate empirical investigation of the range of theories

possessed by subjects will many of these issues be resolved. Thus, I tentatively conclude that pluralism is consistent with theory structures that represent both natural kind and heterogeneous kind properties. Subject to empirical investigation regarding the specific essence beliefs of different subjects, pluralism leaves it open as to whether a unique natural kind or a more heterogeneous kind is represented by a particular subject's theory structure, hence both views are compatible with concept pluralism. This suggests that internalists should be willing to countenance a range of narrow contents that relate to different kinds of mental representations as well as different representational mechanisms, rather than rely on just one kind of property, and one form of narrow representation.

3.4 Concept pluralism and pluralistic representation

The view I am defending is that a given subject's concept, such as WATER, which is comprised of a variety of conceptual structures, may lend itself to representing a range of different properties. This arises because each of the different conceptual structures may relate to the representation of different properties, while each of the conceptual structures mutually comprise the subject's single concept WATER.

This contrasts with a view accord to which S's concept WATER would represent H_2O, while the other pieces of information she associates with the concept may represent other properties. On this view, the concept would not itself represent plurally, and it would be the diverse pieces of information, perhaps stored in a mental file, that related to a plurality of properties being represented by S. This is one contrast that the pluralist view makes with, say, a Burgean distinction between concepts and conceptual explication, where S and T may share their concept but may also have different theories/conceptual explications. The concept pluralist takes the different theories/explications to relate to distinct conceptual structures that partly comprise the subject's concept itself.

My suggestion is that if we understand mental representations in the way concept pluralists do, as concrete mental entities comprised of a range of conceptual structures, then where each conceptual structure associated with that concept may

relate to a different property being represented (such as deeper or more superficial properties), then we get the idea of mental representations representing plurally. Now, the view is not that a token concept such as a specific conceptual structure, like prototype structure, will represent plurally. The claim is rather that a wide range of concepts are plausibly comprised of a range of conceptual structures, and thus the mental representation associated with, say, the expression 'WATER', represents plurally in so far as the concept WATER (belonging to a relevant subject) is comprised of a range of conceptual structures each of which represent different properties. For example, S's theory structure may represent H_2O, while her prototype may represent watery stuff (comprised of a set of superficial properties like being clear, thirst-quenching, etc). It is the composite mental entity WATER that represents plurally, but each constituent conceptual structure represents individually.

It is worth noting that many concepts are complexes comprised of primitive concepts. For instance, WATER may be comprised in part by various primitive appearance-based concepts. Where each of the constituent primitives may individually represent different properties, WATER may then represent each of these individual properties, as well relating to the representation of, say, watery stuff or H_2O.

Above, I noted that the linguistic token 'water' may express the concept WATER. Where S's concept WATER is associated with a range of structures, S's term 'water' may then be used on different occasions to express any of these conceptual structures. In this way, concept pluralism supports a kind of linguistic pluralism, where terms in natural language may systematically represent different properties depending on which conceptual structure the speaker uses the term to express, and what property that structure represents.

3.5 Internalist individuation and psychological generalisations

One important objection internalist accounts of concept individuation face is that it is incompatible with assumptions about the ascription of concepts in psychology. In particular, so the objection goes, if we individuate concepts by narrow content that supervenes on intrinsic properties, then there is a risk that the differences in intrinsic

properties of different subjects will be such that there will be hardly any concepts shared at all. The worry is that this runs against a preference for generalisations in psychology, where the goal is to identify the concept of an X and generalise across the population, thereby ascribing that concept to all members of the population.[28] The objection then is that empirical psychology should be construed as identifying, say, the prototype BIRD, or a theory structure for WATER, which is ascribed to all subjects of the relevant population under study. Thus, the objection to my account would be that it is too liberal in permitting subjects to possess different token BIRD prototypes in so far as the superficial properties they associate with the category are different. However, I am not forced into holding a view at odds with psychology, as I will now explain.

The first point to note is that it is not clear that psychology is committed to identifying a single prototype for each category that is ascribed to each individual. The approach to the individuation of prototypes is compatible with holding that subjects possess distinct prototypes in so far as they make different judgements of typicality. For instance, applying the procedure for different individuals, we could get different results about what their concept BIRD represents. In further support of this point, we can understand psychology as being concerned with the basic concrete conceptual structure and hence with how category information is represented by subjects. This would facilitate the identification of conceptual structures, which would then allow significant generalisation across a population in terms of what kind of conceptual structure to attribute. We should therefore understand pluralism as individuating concepts construed as concrete tokens of structural types like prototypes and theories. When we talk about S's prototype, it is a concrete token and it can differ from T's by virtue of differences in the associated narrow content. S and T's prototypes are different tokens of the same structural type, namely prototype concepts. Prototype concepts as a type encode information about typical features. This is perfectly consistent with some variation in content (the superficial properties

---

[28] Sawyer (2007) provides a version of this objection, and contends that taking concepts to relate to narrow content would be too subjective, leading to ascriptions of highly idiosyncratic concepts which would fail to allow significant generalisation in psychology. Segal (2007) provides a response to this kind of objection which is different from the one I develop here.

represented) among subjects who may otherwise share a type of prototype concept. This may fall short of attributing the exact same conceptual structure, however, there is no prima facie good reason to think that this is mandatory, and even externalists about concept individuation like Weiskopf (2007) appear to permit significant variation between the concepts of subjects.

In Section 2.2.1 above, I introduced the dual view which contends that concepts may be both abstract and concrete entities in some sense. While the response to the generalisations objection above could be pursued, the two-component theorist can appeal to wide and narrow contents and the dual view of concepts. The advantage of this response is that it incorporates much of the externalist's view, whilst preserving the centrality of narrow content.

On the pluralist view, S and T's WATER concepts are comprised of a variety of conceptual structures, and the constituents of their concepts may represent the same superficial properties. What makes these constituents relate to an individual's concept WATER is that they relate to the same thing. Thus, where S's WATER represents watery stuff, the constituent conceptual structures comprising her WATER concept are constituents of that concept because they relate to this higher-level heterogeneous kind. The kind *watery stuff* is individuated in part by its superficial properties. Hence S's token concept WATER is understood on the pluralist view as a complex comprised of constituents which relate to the kind *watery stuff*.

With this in place, we can now say that S and T could have WATER concepts which represent the same kinds, and hence both possess tokens of a watery stuff concept. With regard to the wide contents of S and T's WATER concepts, we might say that where S and T's wide contents are the same, then they can be ascribed the same abstract concept, even where their concepts have different narrow contents. Thus, while S and T may possess distinct token concrete watery stuff concepts, they may also both be ascribed the same abstract water concepts (such as where their wide contents both relate to H_2O, or where both their narrow contents relate to the same

kind). This can be understood as appealing to an abstract concept as a way of making convenient psychological generalisations, for instance, despite variation in the concrete tokens, we can say that S and T both possess the concept of water, or of watery stuff, in so far as their wide and/or narrow contents relate to the same thing. However, it is important to distinguish this informal way of comparing and grouping individuals by way of an abstract view of concepts, from an account of the specific token concepts these subjects possess which will relate to each of their particular concrete conceptual structures.

This abstract conception can serve as the public concept, which in turn relates to the public language meaning of the term 'bird'. While each subject has their own mental representation, we can talk meaningfully of speakers of English who intend to express the same public concept. But we should not confuse this informal way of talking with subjects' concepts construed as concrete mental representations. This is reserved for the concrete view, which allows that talk of shared public concepts being grasped may belie substantive differences in the conceptual structures of S and T, even where their different structures ultimately relate to the representation of the same property.

4. Cognitive content, conceptual role and perceptual templates

In Section 3 I suggested that some of the conceptual structures that comprise a pluralist theory of concepts can be individuated by representational content that is a kind of narrow content. As noted in Section 2, concepts have both representational content and cognitive content. In Section 4 I develop an approach to narrow cognitive content suitable for the concept pluralist framework. However, such a view faces a number of problems. In particular, while some externalists about concept individuation allow that cognitive content may be a form of narrow content, Weiskopf (2007) provides an argument against equating cognitive content with narrow content based on the cognitive content of primitive concepts, the basic units which combine to form complex concepts. If Weiskopf is right, then narrow content can't play the role of cognitive content. This would be a significant consequence, since many internalists and externalists tend to restrict the role of narrow content

specifically to the role of cognitive content. Thus, before developing the notion of narrow cognitive content, I consider Weiskopf's argument. I will show that it can be rejected, and that cognitive content is most plausibly a kind of narrow content. I then develop the notion of narrow cognitive content in more detail and consider two possible objections to the view.

4.1 Weiskopf against narrow cognitive content

As noted in Section 2, Weiskopf's (2007) rejection of narrow content on the basis of the Twin Earth argument can be rejected by highlighting how some narrow content relates to dispositional properties that are the same at possible worlds $W_1...W_n$. However, Weiskopf provides an alternative argument intended to show that cognitive content can't be narrow content. Weiskopf's argument relates to the nature of primitive concepts.[29] The basic notion of a primitive concept can be understood as follows. Primitive concepts are what complex concepts are comprised of. Primitiveness relates to the idea that the concept cannot be further decomposed into more basic concepts. Where a prototype concept like CAT represents a set of superficial properties, we may construe certain members of that set of superficial properties as relating to primitive concepts. One such concept might be FURRY.

Weiskopf's argument is as follows. The cognitive content (CC) of primitive concepts is given by their representational content (RC), that is, how they represent is given entirely by what they represent. Now, according to Weiskopf, RC is typically a kind of wide content. So, since CC = RC for primitives, and RC = wide content, then CC = wide content. It follows that the CC of complexes composed of primitive concepts will also be wide content, since the CC of complexes derives from the CC of primitives, which is wide content. The first response is to draw attention to Weiskopf's (2007, p20) commitment to the assumption that narrow content can't be RC, where he notes that "R-content is a form of wide content, as reference is usually taken to be". I have shown in Chapter 1 and Section 3 above that narrow content can be RC. The RC of both prototypes and theory structures can be a kind of narrow

---

[29] Just what primitive concepts are is a debated issue, and a detailed discussion of this issue is outside the scope of this thesis.

content, thus we should reject this assumption. Given the account of narrow content developed above, it seems natural to hold that the RC of certain primitives will be narrow content. Consider colour primitives. The concept RED itself may not be a primitive concept, as it plausibly relates to the representation of a specific microphysical property. REDNESS, on the other hand, relates to the primitive colour experience of redness. Suppose then that the CC of REDNESS is given by its RC. On the dispositional view of appearances, the RC is given by the dispositional property represented that can be understood as the disposition to cause experiences with a certain phenomenal character. In order to generate Weiskopf's conclusion, it would have to be the case that these properties relate to wide content and not narrow content, however, I have shown in Chapter 1 and Section 3 above that there are good reasons to think that dispositional properties relate to narrow content.

The second point to make is that it is not clear why the CC of primitives must be given by RC. CC tells us how a property is represented as being by a subject's concept, thus it relates to something like Frege's notion of a mode of presentation of the referent. In Chapter 1 I noted that the mode of presentation can be understood in terms of a rule for determining reference or a condition that something must satisfy to be the referent. Another feature of modes of presentation is to reflect the cognitive value or significance of an expression, thus modes of presentation also play a cognitive role in reflecting the way a subject thinks of the referent of an expression. While the Fregean view makes the mode of presentation play a role in determining what is represented, this semantic role does not itself tell us about the cognitive role of providing a way in which the referent is represented as being. Thus, if we treat the CC of primitive concepts as relating only to RC, there is a worry that we have not been told anything about how the relevant property is represented as being, which is what we are mostly interested in when considering the CC of a primitive concept like REDNESS. We should therefore distinguish what is represented (RC) from the cognitive role of the primitive concept (CC). We can then deny that RC=CC for primitives (or that RC provides the CC of all primitives). Considering REDNESS again, we can take the RC to be narrow content (the dispositional appearance property), without holding that the CC must strictly relate to what is represented.

The RC relates to a dispositional property, whereas the CC (understood in terms of the mode of presentation of the relevant property) may relate only to *how* that property is represented subjectively as being. This latter point would require only an account of the mode of presentation itself, i.e. the phenomenal character associated with experiences of red things, and not the property represented by the concept. An internalist who holds that the mode of presentation of a property supervenes only on intrinsic properties could therefore deny that CC = RC for primitives and complexes.

4.2 Conceptual role and perceptual templates

Having shown that we can reject Weiskopf's argument against narrow cognitive content, I turn now to how we should understand the cognitive content of prototype concepts. I suggested that we should not focus on representational content in order to understand cognitive content. This seems to cohere with the view of Laurence & Margolis (2007) who have suggested that cognitive content could be narrow content understood in terms of conceptual role semantics (CRS). However, CRS has been rejected by internalists for being too abstract as an account of concepts (Prinz 2000), and being unable to account for the cognitive content of perceptually-based concepts (Loar 2003). The first worry is that this leads to a highly abstract specification of a concept in terms of relations to other concepts. In addition, CRS fails to account for representational content, since we are not told how a concept like CAT can be about cats if the concept is individuated by its conceptual role. The problem I want to focus on here is Loar's (2003) worry with CRS as a basis for cognitive content. The problem here can be understood by considering what it would mean to account for cognitive content by conceptual role. Conceptual role relates to the relations a concept has with other concepts, so that we are to understand the cognitive content of BIRD in terms of its relations to relevant concepts, such as WING, FEATHER, FLIES, and so on. When S deploys her BIRD concept, the cognitive content of the concept is given by that concept's relations to these other concepts. According to Loar (2003) the worry with this approach is that this leaves something important out of cognitive content, in particular it doesn't seem to tell us how S is thinking of the relevant property. For instance, in citing the typical properties of the category bird, we are not told how these contribute to the way S is thinking of birds. As already

noted, it is useful to understand cognitive content in terms of Frege's notion of the mode of presentation of the referent. On a Fregean view, an expression like 'bird' has both sense and reference, where the sense relates to the mode of presentation of the referent, reflecting the cognitive value of 'bird'. The mode of presentation is how the referent is presented as being, thus we can understand cognitive content in terms of the way a subject is thinking of the referent of a concept. What is therefore missing on CRS as an account of cognitive content is how the relations to other concepts contribute to the mode of presentation of the relevant kind represented. The worry is that conceptual role alone is unable to tell us much about S's perspective on the world.

Loar's (2003) suggestion was to appeal to phenomenal intentionality. Phenomenal intentionality is the purported phenomenal character associated with intentionality, that is, a distinctive phenomenal character of a state's being about or representing what it is about. On Loar's version of phenomenal intentionality, it relates to the phenomenal character of the intentionality of occurent perceptual experiences. [30] However, this doesn't seem to give us what we want in terms of the mode of presentation or cognitive content of concepts.

An alternative approach to cognitive content that provides a more suitable notion of the mode of presentation of the kind represented can be developed by considering the relevant conceptual structures available to concept pluralists. Returning to prototypes and the typicality experiments, it is held that subjects make their judgements about category membership on the basis of superficial properties such as appearance properties. We can understand this in the following way. S is presented with an object O, and forms relevant perceptual representations of O. When S perceives or detects an O and classifies O as a C, this is based in part on the information encoded in memory which furnishes S with the relevant disposition to classify O's under C. Such information thereby relates directly to her prototype of C's.

---

[30] Note that there are other versions of phenomenal intentionality including proposals from Horgan & Tienson (2002) and Farkas (2008a).

This model of prototypes provides an alternative basis for cognitive content to that of conceptual role. Where conceptual role would account for cognitive content purely in terms of citing the relevant concepts that the focal concept was related to, with prototype structures we can appeal to the perceptual information stored in memory. Consider again the concept BIRD. When S classifies an instance of the category, this judgement is based on the perceptual data from her perceptual representations matching the information encoded in memory. S's prototype encodes information about the various appearance properties of typical birds, such as those relating to the superficial properties of wings, beaks, feathers, and so on. We can group this information together using the notion of a perceptual template. [31] The perceptual template for S's BIRD prototype will thus encode all the typical appearance properties S associates with the category, which derive from her perceptual representations of those properties. This perceptual information can then directly serve as the basis for the cognitive content of S's prototype concept. Perceptual templates are created when subjects form perceptual representations of various category features. For instance, on encountering a robin, a perceptual template is created in memory which encodes information relating to the appearance properties detected. On this view, the cognitive content of prototype concepts is understood in terms of the contents of perceptual templates. Such content is at base perceptual or experiential, which coheres with prototypes being representations of appearance properties.

In order for perceptual templates to provide a basis for a notion of narrow cognitive content, it needs to be shown that we can individuate the relevant templates narrowly, such that the resulting cognitive content supervenes on intrinsic properties. Given the architectural relation between perceptual representations and the cognitive content of concepts through the mechanism of a perceptual template, there is a constitutive connection between how things appear in perceptual representations of a category and the cognitive content of a concept of that category. The case for the

---

[31] This use of the notion of a perceptual template should not be confused with the use in relation to accounts of perceptual demonstrative concepts discussed by Papineau (2006). It is perfectly consistent to hold that the perceptual demonstrative concept THAT BIRD has a kind of wide content, relating to specific bird indicated, while the general concept BIRD has narrow content, relating to the general appearance properties represented.

narrowness of cognitive content can be based on the claim that the cognitive content of perceptual templates and the relevant perceptual representations comes from the phenomenal character of those representations. If phenomenal character is narrow, and this provides the mode of presentation associated with the perceptual representation, which then provides the basis for the mode of presentation associated with the perceptual templates, then at base the cognitive content of concepts will be narrow. The argument for narrow cognitive content therefore depends on the truth of phenomenal internalism: the view that phenomenal character supervenes on intrinsic properties. To this end, in Section 4.3 I motivate the case for phenomenal internalism.

4.3 Perceptual templates and phenomenal internalism

In Section 4.3 I do two things. First, I make the case that the mode of presentation that derives from perceptual representations can be understood in terms of the mode of presentation related to the underlying perceptual experiences which plausibly supervene only on intrinsic properties, specifically in terms of phenomenal character. Second, I argue that nothing external fixes phenomenal character (or, hence, mode of presentation) other than causally, so mode of presentation (or cognitive content) cannot be identified with wide content.

In Section 4.2 I outlined how to relate modes of presentation to prototypes through the mechanism of perceptual templates. What needs to be shown is how the perceptual base can play the required role of modes of presentation, i.e. reflecting how things are represented as being. The mode of presentation associated with perceptual templates can be associated with perceptual phenomenology because the phenomenal character of a given perceptual representation reflects what it is like to be in that state, i.e. how things seem to the subject subjectively. This includes the kinds of experiences associated with vision, audition, and the other perceptual modalities. Internalism about modes of presentation will then depend on whether phenomenal internalism is true, which is the thesis that the phenomenal character of perceptual representations supervenes on intrinsic properties. That is, fixing intrinsic properties fixes the phenomenology of a given perceptual state.

Phenomenal internalism enjoys considerable support in the literature on perception and representation. Different forms of the thesis are traceable to proponents of qualia (the purportedly non-representational aspects of experience directly responsible for phenomenal character) who see qualia as intrinsic properties (e.g. Block 1990), as well as proponents of phenomenal intentionality who contend that there is a kind of representation that is based on phenomenology which is seen as a genuinely internalist kind of representation (Horgan & Tienson 2002; Loar 2003; Chalmers 2006a; Kriegel 2008; Farkas 2008a). There are, of course, critics of the view such as Tye (2000) who deny that phenomenal character is intrinsic. I consider Tye's view in Section 4.4 below, but first I make the positive case for thinking that phenomenology supervenes on intrinsic properties of individuals and is a suitable basis for narrow cognitive content.

Phenomenal internalism entails that any intrinsic duplicate of a subject S, who is currently in a perceptual state with a given phenomenal character E, will also be in a relevant state with E. According to this view, nothing external, including differences in distal causes or embedding environments, can lead to any differences in S and her duplicate's phenomenal character E. Several cases appear to support this view. Consider a variation of the Twin Earth scenario involving Earth and Inverted Earth. On Earth, S perceives a red ball with the relevant objective microphysical properties that cause experiences of redness in S. On Inverted Earth, T perceives a green ball with the relevant objective physical properties that would ordinarily cause experiences of greenness in S and T on Earth. However, a peculiar feature of Inverted Earth is that there are atmospheric irregularities that when T perceives these objectively green objects, T is caused to be put into a perceptual state normally associated with experiences of red things. Thus, S and T can be intrinsic duplicates, and have experiences as of a red ball in front of them, despite T's state being caused by a ball that has the relevant objective microphysical properties of green objects. It seems that given their status as intrinsic duplicates, it is natural to see S and T as being in states with identical phenomenal characters, since both are plausibly in states with a phenomenal character pertaining to a reddish ball. The upshot is that what matters for phenomenal character is not what is causally responsible for putting

a subject into a given state, but how that state translates to subjective experiences with a distinctive phenomenology.

Suppose that S's brain is duplicated molecule for molecule by a mad scientist who creates an intrinsic replica of S's brain and stores it in a vat. If S and the brain-in-a-vat (BIV) are in identical states at time t, then when S enjoys a perceptual experience of an evening sunset, from the BIV's perspective things will appear to be exactly the same as from S's perspective. When S focuses her eyes on the waning light, the BIV is caused by the mad scientist to have relevant experiences as of focusing its eyes in an identical way. The upshot is that despite the BIV not having normal causal relations to an environment and having no perceptual system, it can still be in brain states that have the exact same phenomenal character as that of S's perceptual states.

What these cases show is that there is good reason to think that phenomenal character supervenes on intrinsic properties, and hence that phenomenal internalism is true, so that when intrinsic properties are fixed, the phenomenal character is also fixed. What underlies support for this view is the intuition that what accounts for how things seem in perceptual experience are primarily the internal structures of the brain and perceptual system. Thus, BIVs can be in states with a given phenomenal character despite having no perceptual contact with anything. What matters for things seeming a certain way is only that the BIV's brain is in a certain state. This is true also of Inverted Earthers who are in the same internal state as Earthers, and enjoy similar subjective experiences, despite being put into that state by causal relations to objects with different objective microphysical properties. On this view, what ultimately matters for things appearing a certain way in perception is how S's head is at a certain time. If this is right, then modes of presentation are narrow as they are provided by the phenomenal character of perceptual representations that supervene on intrinsic properties that are shared by intrinsic duplicates.

What drives this position is the idea that intrinsic states are both necessary and sufficient for being in a state with a given phenomenal character. Such states are necessary since having intrinsic states is required to facilitate phenomenology. They

are sufficient, since nothing external appears to be required to be in a state with a certain phenomenal character, other than causally. This comprises an argument for the narrowness of the mode of presentation of perceptual representations:

(1) Intrinsic properties are necessary to be in a state with phenomenal character

(2) Intrinsic properties are sufficient, since nothing external is required to be in a state with phenomenal character, other than causally

(3) Therefore, intrinsic states are both necessary and sufficient for phenomenal character, so phenomenology is narrow

If the argument succeeds then externalism about modes of presentation based on perceptual phenomenology would be false. One possible objection might be that a genuine mode of presentation must be a mode of presentation of something, so a BIV has no mode of presentation, and we should distinguish Inverted Earthers' mode of presentation from Earthers' mode of presentation. The sufficiency claim would then be false because concepts only have modes of presentation if they are genuine representations, and hence the mode of presentation (cognitive content) of the relevant concept cannot be separated from the representational content of the relevant concept. If this was true, then a BIV could not have a genuine mode of presentation, and the mode of presentation of an Inverted Earther would relate to the different representational content of her perceptual representation, namely the objectively green object. However, the point to note in response is that given the account of narrow content developed in Chapter 1 and Section 3, an internalist can hold that Earthers and Inverted Earthers represent the same appearance properties. The treatment of the BIV would be slightly different, since an internalist could say that the BIV fails to represent anything at all (or perhaps only features of the relevant simulation), so its states lack representational content, but can still have cognitive content. Consequently, the argument can be accepted without serious difficulties arising.

Before considering objections to the view developed in Section 3 and 4, I want to briefly explain how the defence of phenomenal internalism supports the view in

Section 4.1 that the cognitive content of concepts is narrow. The cognitive content of prototypes derives from the stored information in memory, which relates to the perceptual representations of the relevant properties. Since the modes of presentation associated with the perceptual states relate to the phenomenology of these states, and the cognitive content associated with concepts derives directly from this information stored in memory, the cognitive content likewise supervenes on these intrinsic properties. Consider the prototype BIRD. When S encounters an instance of the category, S goes into a state with phenomenal character E, relating to the relevant appearance properties her perceptual system detects. The mode of presentation of S's perceptual state is given by E, and E is encoded into memory. When S deploys her prototype concept BIRD, thinking of typical appearance properties, she retrieves the relevant perceptual data, and this provides the mechanism which facilitates S thinking of the category in a certain way, her cognitive content.

As noted in Section 2 and 3, the internalist account developed here is consistent with a two-component theory of content which allows that there is also wide content. The proponent of narrow content can allow that perceptual experiences have wide content and narrow content, while holding that narrow content best reflects how things are represented as being to the subject and is therefore most suitable for playing the role of cognitive content. Despite the plausibility of phenomenal internalism, some philosophers have tried to reject the supervenience claim and hold a view according to which phenomenal character partly supervenes on extrinsic properties. Tye (2000) defends a version of this view and if correct, then the case for internalism about modes of presentation would be undermined, thus it is crucial to consider Tye's position which I will do in Section 4.4.

4.4 Phenomenal externalism

A possible objection to phenomenal internalism is to contend that in some sense the phenomenal character of perceptual states is individuated broadly. Tye (2000) has defended an externalist view of phenomenal character, and in this section I consider whether that view serves as a convincing objection to phenomenal internalism.

According to Tye (2000), the phenomenal character of perceptual experiences is identical to the content of the relevant experience. On this view, where content is wide, intrinsic duplicates will be in states with different phenomenal characters because the phenomenal characters of their identical internal states are individuated by the different wide content of their states. Despite things being subjectively indistinguishable to such subjects, the fact of their being in identical internal states does not guarantee that the phenomenal characters of their states are identical. If this view is correct, then phenomenal internalism is false because differences in the environment, here reflected in the different wide content, result in differences in the phenomenal character of duplicates' states. With regard to the argument above, Tye's approach would involve rejecting point (2) on the sufficiency of intrinsic states, since if phenomenal character is identical to a kind of wide content, then intrinsic states are not sufficient because taking away the wide content would result, apparently, in there being no significant phenomenal character.

The problem with this consequence is that it seems highly counterintuitive. The view is too strong as it would rule out BIV's enjoying states with phenomenal character as, it seems, they cannot be in states with wide content because they lack the required relations to the world. Now, it might be possible to deny that the BIV enjoys the same kind of phenomenal character as a subject in a state with an associated wide content. For instance, we could say that the BIV merely enjoys states that involve phenomenal goings on, which differ in some substantive way from what Tye means by phenomenal character. The problem with this is that it highlights that Tye has developed a very specific theory of phenomenal character which needn't be accepted. As noted above in Section 4.2, some proponents of phenomenal internalism endorse the notion of qualia, which are non-representational components of perceptual states that provide the relevant phenomenal character of these states. Thus, where Tye contends that phenomenal character is fundamentally a content-involving notion, qualia theorists deny this and take phenomenal character to be independent of content. This discussion is outside the scope of this thesis. For present purposes all I need to show is that the kind of objection that externalism

about phenomenal character poses to my view has highly counterintuitive consequences and is based on a contentious framework.[32]

4.5 The complexity of narrow cognitive content

A final objection to the account developed here might focus on the account of cognitive content defended in Section 4 above, upon which part of my internalist thesis rests. The objection contends that this kind of account of narrow cognitive content renders cognitive content too complex. The problem about complexity arises from Chapter 1 and Section 3 above where I held that narrow content relates to the representation of dispositional properties. When we specify this content, so the objection might go, it would seem to be too conceptually sophisticated to reflect how subjects think of things as being, thus, it couldn't play the role of cognitive content. In addition, the view of modes of presentation developed in Section 4 above is based on the phenomenology of perceptual representations, which requires that cognitive content reflects how things seem to the subject. The problem is that when we state the narrow content in such a way that it represents a dispositional property, this might seem overly sophisticated or complex, and doesn't reflect how the subject is thinking of the property represented, furthermore, it would seem implausible if such properties were reflected in the phenomenology of perceptual experiences of such properties. For instance, it might be objected that S doesn't think of the property *redness* in terms of its being a dispositional property. As Vaughan (1989) has argued, such a view of narrow content seems problematic since if in stating the content of a concept we appeal to sophisticated notions like dispositional properties, then it is not clear that we are accurately reflecting how subjects represent things as being.[33] It may also be claimed that there is a tension with my account of mode of presentation in Section 4, where it rests only on the phenomenology of perceptual representations. Whether phenomenology could reflect that a dispositional property was being represented is not clear. Thus, on the one hand, my account of the representational

---

[32] A proponent of phenomenal internalism does not need to deny that Tye's positive account of phenomenal content as wide content could be correct to an extent. That there is a kind of wide content that relates to perceptual experience is compatible with phenomenal character supervening on intrinsic properties. A two-component theorist can accept that there are a range of contents associated with perceptual experience. Again, these issues are outside the scope of this thesis.
[33] Vaughan's objection was directed at Searle's (1983) account of narrow content, but the objection applies generally.

nature of narrow content seems too complex, since in representing dispositional properties it doesn't seem to capture how S would think of such a property, and on the other, the account I gave of the mode of presentation associated with narrow content doesn't seem consistent with the account of representational content.

The solution is to make a distinction between how we specify what is represented, i.e. the representational content, and how such content relates to the mode of presentation or way in which S thinks of what is represented, the cognitive content. What this means here is that we need to distinguish how we specify the relevant properties represented, from how we understand the mode of presentation of these properties. When we undertake the former project, we need to make reference to complex dispositional properties, which S may in fact be unaware that she is representing. For instance, how things seem to S when she deploys her general concept RED, need not alone determine the representational content of her concept, especially if the account of individuation in Section 3 is right. There, it was argued that S's prototypes are individuated in part by S's judgements about which superficial properties fall within a given category, and that S would likely make judgements based on dispositional appearance properties. Now, S needn't know she is doing this, and hence may be unaware that her responses relate to dispositional properties. Nonetheless, her prototypes can plausibly represent such properties. Furthermore, S's mode of presentation will be determined, as noted in Section 4, by the phenomenology of the relevant states, and this may mask which properties are represented by such states. Consequently, there is no tension between the narrow representational and cognitive content associated with the relevant concepts.

5. Conclusion

The main aim of this chapter was to show that concept pluralism is consistent with individuating concepts by their narrow contents. I suggested that the accounts of narrow content identified in Chapter 1 provide suitable representational contents for individuating prototype and theory structures. In support of the internalist construal of pluralism I showed that internalist individuation of these structures is not inconsistent with psychological generalisations about concepts. A second aim was to

show that the cognitive content of concepts is narrow. In particular, I showed that Weiskopf's (2007) argument against narrow cognitive content can be rejected, and I outlined a suitable account of cognitive content for prototypes using the notion of perceptual templates. Perceptual templates support narrow cognitive content by encoding perceptual information relating to the phenomenal character of perceptual representations, which I showed relates to phenomenal internalism. The view developed here provides support to the account of narrow content in Chapter 1, where I assumed that mental representations come in a variety of structures. There are two main ways that the internalist account of concepts and content might be rejected, and I consider these in Chapter 3.

Chapter 3

Social Externalism and Perceptual Psychology

1. Introduction

In this chapter I consider two approaches to individuating mental states which purport to have anti-internalist implications. These represent two of the main ways externalists have tried to show that internalism and narrow content do not relate fundamentally to the individuation of mind. I show that each view ultimately fails to have such implications, and hence that the view developed in Chapter 1 and 2 withstands objections arising from these externalist positions. The first part of the chapter considers a version of social externalism associated with Burge's (1979) arguments for an externalist account of meaning and mind, as well as an argument from Goldberg (2002) regarding social externalism as a basis for a subject's cognitive content. The central thread of my critique of social externalism is that the claims about concepts and cognitive content can be rejected by considering the relevant arguments, and, importantly, showing that the social externalist proposals are at odds with the work in psychology cited in Chapter 2 and hence do not have the intended anti-internalist consequences for mental nature. The second part considers accounts of perceptual psychology from Burge (1986a; 2009b; 2010) and Egan (2009; 2010). The main objection to these views is that the proposed anti-internalist conclusions can be rejected and so the basis for denying any role for narrow content in psychology may also be rejected.

2. Social externalism and the ascription of propositional attitudes

In Chapter 1 I considered Putnam's (1975) view that certain natural kind terms have an externalist semantics. On this view, externalism applies only to terms or concepts that represent natural kinds, which leaves open the possibility that concepts that

represent other kinds or properties are not externalist. Burge (1979) introduced an example that was intended to demonstrate that externalism may extend to most, if not all, terms/concepts. One further intended consequence of Burge's thesis is that externalism may have radical consequences for the individuation and nature of mental states. Such a view poses a threat to the claims of Chapters 1 and 2, and in the rest of this section I consider Burge's view in detail and raise several objections to it.

2.1 Burge's arthritis argument

Social externalism is the view that social facts, most notably those relating to conventional linguistic meaning, play a constitutive role in the individuation and specification of mental states and mental content. Burge generates radical social externalist consequences by means of his arthritis example. The arthritis case involves a subject called 'Alf' who incorrectly believes that he has arthritis in his thigh. One feature of the scenario is that Alf apparently has many true beliefs such as that his father had arthritis and that he has arthritis in his wrists and ankles. According to Burge, Alf meets minimal conditions on competence with the term since he can use it in these cases to talk about the condition and to express his true beliefs. Furthermore, we can use 'arthritis' to express Alf's true beliefs where 'arthritis' occurs in the expression of Alf's propositional attitudes, for instance we can report that 'Alf believes that he has arthritis in his ankles'.

The step to motivating social externalism involves considering a counterfactual scenario in which 'arthritis' is now used to talk about a broader condition which would apply to Alf's thigh. Suppose in this scenario Alf remains in the exact same physical state as in the former scenario. The social externalist claims that Alf now has beliefs involving this other condition, which we may call 'tharthritis'. For instance Alf now believes that his father had tharthritis, and that he has tharthritis in his ankles. A consequence of this scenario is that Alf now has a true belief that his tharthritis has spread to his thigh. Taken literally, the relevant attitude ascriptions ascribe to Alf different beliefs in the two situations. In the first scenario, Alf has beliefs concerning arthritis, and in the second has beliefs about tharthritis. We may

also say that different concepts are ascribed, since in the first situation Alf will grasp and be attributed the concept ARTHRITIS, whereas in the counterfactual scenario Alf grasps and is ascribed the concept THARTHRITIS.

The key social externalist claim is that all of this follows from the fact that in each scenario the relevant linguistic communities apply the term 'arthritis' to slightly different conditions. The resulting beliefs and concepts are therefore different by virtue of the different meaning of the term 'arthritis' in each linguistic community, since we get the appropriate belief/concept from the literal meaning of the term 'arthritis' when ascribing the relevant beliefs/concepts. Social externalists may therefore conclude that concepts are partly individuated by facts about the social or public meaning of words (as used in sentences we use to ascribe those concepts/beliefs), since this is the only difference between the two situations. Consequently, psychological internalism is false, since intrinsic duplicates need not be psychological duplicates, that is, propositional attitude content fails to supervene on intrinsic properties. The radical nature of this view lies not only in its externalism, but in the way in which mentality appears to be intertwined with social facts about linguistic meaning.[34] The main stages of Burge's argument can be summarised as follows:

(1) Alf has many true beliefs about arthritis, but falsely believes that he has arthritis in his thigh, as shown by his disposition to say things like 'I have arthritis in my thigh'

(2) In a counterfactual scenario, 'arthritis' refers to a condition that also affects tissue outside of joints. In this scenario, Alf's belief expressed as 'I have arthritis in my thigh' is true

(3) Thus, in the counterfactual scenario Alf does not believe that he has arthritis in his thigh, instead, he believes he has tharthritis in his thigh, since 'arthritis' has a different meaning in this scenario

---

[34] Burge (2003b) has more recently down-played this approach to externalism and emphasised the kind of approach that I will consider in Section 4, however, others (for instance Goldberg 2002; Sawyer 2007) continue to defend a version of social externalism based on the considerations that arise in the arthritis case. Thus, it is worth considering the view in some detail.

If correct, Burge's view could provide a problem for the account of narrow content developed in Chapter 1 as it would show that intrinsic duplicates S and T don't share a kind of narrow content since what their words or concepts represent is fixed by facts about the linguistic environment. This is sometimes cited as a way of showing that no terms are immune to the Twin Earth argument. In Chapter 1, I suggested that non-natural kind terms like 'clear' and so on are not subject to the argument and thus there is no problem with expressing narrow contents. If Burge's argument succeeds, then this could put pressure on that response. My aim then is to highlight some fundamental problems with Burge's argument that show that the internalist is entitled to defend narrow content common to S and T which relates to a shared mental representation. The internalist can accept that there may be a form of wide content relating to social externalism, however, this does not preclude narrow content, as I will now show.

2.2 Responses to the argument

Burge's interpretation of the case is initially plausible, and Burge seems right that the example illustrates a pervasive phenomenon common to ordinary speakers' practices. However, it is not so clear that the example accords with the theory of meaning in linguistics or with ascriptions of concepts in psychology. In this section I raise three main responses to Burge's social externalism [35] which show that the example cannot motivate the radical social externalist consequences.

2.2.1 Assumption about natural language meaning

The first response relates to the assumed view of natural language meaning required to generate Burge's conclusions about the ascription of beliefs. Burge's view requires that natural language meaning is essentially social, and this idea drives the psychological externalist conclusions. If, on the other hand, we have reason to think that meaning is not social, or at least that there is an aspect of meaning that doesn't supervene on social facts, then this could block the externalist conclusions, or at least block the conclusion that there is no narrow content in this case.

---

[35] Others focus on different problems with the argument (for instance Wikforss 2001; Georgalis 2003). My approach is different from these other objections, and relates more closely to the view defended in Chapters 1 and 2.

The importance for Burge's social externalism of assuming that meaning is social relates to the way in which beliefs/concepts are ascribed on the basis of what the meaning of the relevant term is in the local linguistic community. We are told that in scenario 1 'arthritis' has its standard meaning, and in scenario 2 it has a broader meaning. The meaning of the term is fixed by the conventions operative in the linguistic community, and the conventions are themselves fixed by agreement among the experts. When we ascribe to Alf a belief/concept, this is done by taking the conventional meaning of the term, thereby identifying Alf's concept by way of the appropriate meaning of the term in the local community.

The worry I want to briefly raise here is that we are not given any independent reason to accept this account of linguistic meaning. The problem for Burge is that, although such a view of meaning is accepted by others (for instance Ludlow 2005; Higginbotham 2006), the issue of the meaning of natural language expressions is an unresolved problem. Helping himself to this view begs the question against a proponent of an alternative view of meaning. For instance, some linguists and philosophers defend an internalist view of natural language meaning, according to which the meaning of expressions is determined psychologically, and not socially. That is, meaning supervenes on intrinsic properties of speakers. On this view, what a speaker means with an expression is determined by their language faculty (for instance see Chomsky 2000) or concepts (e.g. Jackendoff 2002), and not by conventions fixed by relevant experts. Supposing that this internalist view of meaning is viable, then we could reject the idea of reading-off the beliefs or concepts to ascribe to Alf on the basis of the conventional meaning of 'arthritis'. I will not consider this debate in any detail here as this would require a chapter of its own. The point to draw from this debate is that Burge takes the example to independently demonstrate externalist principles about mentality, but this only follows if Burge is entitled to make assumptions about linguistic meaning. No such justification has been provided, therefore an internalist who has reason to defend the psychological view of meaning will not be too concerned by the arthritis example.

2.2.2 Reinterpretation of the thought experiment

The second response is that the view makes substantive claims about the individuation of concepts which have not been independently motivated. In particular, if Chapter 2 is right then the view is in conflict with concepts as treated in empirical psychology, and as a consequence we should reject the purported radical implications for the mind.

Several internalists (for instance Crane 1991; Segal 2000) have responded to Burge by showing how the scenario is consistent with attributing to Alf a concept that relates to the condition from scenario 2. On this view, we should ascribe to Alf the same concept in both cases because, irrespective of the meaning of 'arthritis' in the two scenarios, Alf understands the term to mean the same thing in both. Burge (1979) considers the possibility of reinterpreting the example and attributing to Alf a concept that matches his understanding. Burge allows that the term 'tharthritis' could be used to capture something about what Alf understands. However, Burge urges us to resist such an interpretation:

> Suppose we are to reinterpret the attribution of his erroneous belief that he has arthritis in the thigh. We make up a term 'tharthritis' that covers arthritis and whatever it is he has in his thigh. The appropriate restrictions on the application of this term and of the patient's supposed notion are unclear. Is just any problem in the thigh that the patient wants to call 'arthritis' to count as tharthritis? Are other ailments covered? What would decide? The problem is that there are no recognized standards governing the application of the new term. In such cases, the method is patently ad hoc. (Burge 1979, 22)

The worry Burge highlights here is of genuine interest, since any kind of reinterpretation of Alf should be able to say something about what concept Alf possesses. If there is no way to decide on the boundaries of Alf's concept, then we would, presumably, be in a worse position than if we ascribe the standard concept of arthritis. However, it's far from obvious that a) there is no alternative way to state precisely what Alf's concept is, and b) even if there was some doubt about how Alf applies the term, then we should opt for ascribing the standard concept.

The first problem with Burge's response, relating to a), is that given a view of concepts such as that developed in Chapter 2, ascribing a non-standard concept to Alf is not ad hoc, but independently motivated from the nature of concept individuation. For instance, the account of prototypes provides a standard for determining what Alf would judge as falling within various categories. This account of concept individuation actively draws on what subjects like Alf are disposed to judge as falling within a category, thus we could carry out the kind of tests involved in the typicality effect experiments, where subjects make category judgements based on properties deemed to be typical of the category, and record what properties Alf is disposed to judge as being typical of the category represented by ARTHRITIS. With regard to b), suppose we were unsure about the data, so that Alf was so confused that we couldn't get any clear handle on what property or kind he represented with ARTHRITIS. Burge appears to urge us to accept the ascription of a standard public concept to Alf merely because there is no alternative. However, if Alf really is confused, then it might be better to ascribe no specific concept at all. This would seem to reflect Alf's psychological position better than ascribing the standard concept of arthritis, since Alf does not meet normal conditions for ascription of a concept.

We don't have to decide in exactly which cases Alf is deeply confused or Alf possesses a different concept from the standard. It is enough that in principle it is possible to identify, say, a prototype based on what Alf is disposed to classify under ARTHRITIS. This is consistent with there being a scenario where no coherent prototype can be identified. As Crane (1991) and Segal (2000) note, here we would just refrain from ascribing any concept, since Alf's error relates to a serious confusion, which, we may imagine, fails to provide a stable basis for identifying any meaningful prototype.

Burge considers the possibility of appealing to Alf's dispositions, and appears to concede that there are cases where we could attribute a concept that fits with the subject's behaviour and which captures his incomplete understanding. However, Burge does not make it clear when this would be appropriate:

None of the forgoing is meant to deny that frequently when a person incompletely understands an attitude content he has some other attitude content that more or less captures his understanding…There are also cases in which it is reasonable to say that, at least in a sense, a person has a notion that is expressed by his dispositions to classify things in a certain way-even if there is no conventional term in the person's repertoire that neatly corresponds to that "way."… Certain animals as well as people may have nonverbal notions of this sort. On the other hand, the fact that such attributions are justifiable per se yields no reason to deny that the subject (also) has object-level attitudes whose contents involve the relevant incompletely understood notion. (Burge 1979, 23-24)

Burge seems to accept that classifying behaviour could indicate the 'notion' (concept) grasped by the subject. Following the account of prototype individuation outlined in Chapter 2, it could be argued that the classificatory or categorizing behaviour of a subject is central to the individuation of concepts. These concessions are therefore enough for the internalist view to get off the ground. Now, Burge's basis for denying that the notion we identify in this way has the status of the subject's concept appears to rely on the idea that we lack terms for saying what this notion is, hence the comparison with the notions of other animals.

There is an important dialectical point to make regarding Burge's set-up of the issues. On Burge's view, we are limited to the social meaning of terms, fixed by the experts, to communicate and express the concepts of others. Given this, there is no way to say what other non-standard concepts a subject like Alf may possess. However, there are alternative stances on both of these issues. Suppose that the meaning of 'arthritis' is determined by an individual's psychology. While it may be true that most subjects associate the same, or a similar, meaning with the term 'arthritis', it would be wrong to take the view Burge does in assuming that how a subject thinks of things is not expressed by the terms they use. For instance, suppose that terms always express what speakers intend them to express. On this view, we might not be immediately aware of what Alf intends to express with his use of 'arthritis', but he expresses it nonetheless. Thus, Burge is wrong in saying that there is no term in the individual's repertoire that expresses the individual's specific understanding of that term. All we are entitled to say is that there is no simple reading-off from a term

expressed by such an individual what they intend to express, and hence their understanding of the term.

Burge's assumption, noted in Chapter 2, that concepts are abstract entities is relevant here. When we use 'arthritis' to talk about a concept of arthritis, we are really talking about the relevant abstract mental representation. I don't want to deny that we may ascribe a wide content along the lines indicated in the example. The two-component view allows such a position. However, as noted in Chapter 2, it is not clear that the proponent of the abstract view would be able to defend the claim that the ascription relates to Alf's concepts, since there is a commitment here to concepts as abstract rather than concrete entities, and it was noted that the dual view of concepts is not well motivated. Again, as noted in Chapter 2, Burge would need to say more about why we can't just opt for the concrete view of concepts and say that the wide content related to some other phenomena.

The problem for Burge is that the various conceptual structures discussed in Chapter 2 relate to the ways in which subjects categorize, which depends on the kinds of ways they think of categories, which are individual psychological capacities. We do not start with a public concept to be grasped and decide whether Alf possesses that. Instead, we assess what concepts are, as mental structures, and see whether Alf has the standard prototype/theory for ARTHRITIS. Since he doesn't, we might say he is conceptually confused and doesn't possess the concept, and Burge is wrong. However, if Alf's mistake is systematic enough, then reinterpretation may be appropriate. It all depends on the seriousness of the error. We assume that subjects will generally possess, say, the same CAT prototype, but for ARTHRITIS it isn't at all obvious that subjects will think alike, since it is less common and its features less well-known. In fact, it would be odd to say that most possess the public/expert concept, if, as seems possible, that it is not common knowledge that arthritis is only a condition of the joints (something that can only be confirmed through enquiry of normal subjects).

Concept pluralism is compatible with there being a variety of ways of conceptualizing arthritis, without there being a single standard concept. This applies to other concepts as well, especially where differences between a scientific understanding and common sense conception are concerned. In certain cases, it might be the case that some subjects have a complex theory structure about a certain property, whilst others only possess a prototype structure. Furthermore, we are not lead to always having to reinterpret subjects who don't completely master a relevant term since in many cases we individuate concepts by a set of properties, or by properties most central to the category. This allows for some flexibility, so that if S thinks incorrectly that X's have property P, then S still represents X's because she associates enough of the actual properties of X's with her concept. However, pluralism allows that in principle we could distinguish the concepts of lay and expert subjects by virtue of their encoding enough different properties so as to really have different concepts that should be distinguished.

2.2.3 The nature and use of attitude ascriptions

The third response questions the role which attitude ascriptions play in establishing social externalism about subjects' concepts. Burge holds that the manner of ascribing beliefs/concepts in the example is standard practice. Thus, ordinarily we ascribe to subjects the literal content of the terms that we use to talk about their beliefs, even when they lack complete mastery of the term. For this to have radical conclusions concerning the nature of mental content, it must be shown that even if this reflects standard practice regarding how we talk about subjects like Alf, then this has significant consequences for the nature of the concepts ascribed. Without this connection, then there may be something else going on in standard practice which does not have the radical implications.

Before considering this point in more detail, I want to point out that as far as psychology is concerned we should reject the claim that concepts are ascribed in this way. This is because, as noted in Chapter 2, theories of concepts like prototypes appeal to data generated from experiments, and so any intuition-based approach to concepts will not have consequences for how concepts are treated in psychology.

Thus, we cannot move from the way ordinary speakers ascribe beliefs to a substantive theory of the concepts actually possessed by the ascribee.

As already noted, Burge makes a plausible point in support of the social externalist view that the interpretation of Alf in the two scenarios is common place. Standard practice involves ascribing the literal meaning of a term to true propositional attitudes in order to say what beliefs a subject has. Since attributing literal content is standard practice, we say that Alf believes he has arthritis, not tharthritis, in his thigh. When Alf says that he thinks his arthritis has spread to his thigh, we might ordinarily say his belief is false. Burge's main point is that standard practice does not involve attributing something other than the standard concept associated with the relevant terms that occur in the propositional attitude ascribed, even when we discover, as in the case of Alf, that the subject only partially understands the meaning of the term. As Burge notes:

> An attribution to someone of a true belief that he [has arthritis], is not typically reformulated when it is learned that the subject had not fully understood what [arthritis] is. (Burge 1979, 22)

This point seems right as regards how we talk about belief. For instance, even if we had reservations about saying Alf possessed the standard concept of arthritis, it seems wrong to say that we would want to say that Alf only had true beliefs involving a non-standard concept. In Chapter 4 I will consider the nature and use of propositional attitude ascriptions in more detail. In particular, I will highlight that attitude ascriptions are consistent with the ascription of narrow contents. For now, the relevant issue is whether we must conclude from how lay speakers use attitude ascriptions that this has implications for the concepts subjects like Alf possess. The worry for Burge is that even if we often do talk about subjects like Alf in this way, this does not commit us to the view that we have thereby ascribed to Alf the standard concept. For instance, just because we say of Alf that he has true beliefs concerning arthritis, and do not say that he has true beliefs about tharthritis, this does not lend support to the social externalist thesis.

Given the point made above in Section 2.2.2 on noting whether to ascribe a non-standard concept or just say Alf is deeply confused, this suggests that standard practice may be more sensitive to subjects like Alf who reveal their confusion by saying things like 'I have arthritis in my thigh'. It seems reasonable then that an ordinary speaker might have reservations in continuing to ascribe Alf the concept of arthritis, even if they might not want to reformulate the propositional attitudes relating to Alf's belief about the condition in his thigh. All this suggests, however, is that ordinary speakers might be inconsistent in their ascriptions of belief and belief reporting practices. This does not seem like an unreasonable position, since to expect ordinary speakers to have well developed thoughts on how to handle such cases doesn't seem particularly plausible, especially given the wide range of intuitions elicited in the philosophical literature, which itself suggests that we don't tend to have a ready response to these cases.

2.2.4 Burge on reinterpreting the arthritis scenario

Above, I appealed to several reinterpretations of the arthritis scenario, and to ways in which these reinterpretations can be supported. However, Burge does consider several reinterpretations of the arthritis scenario and rejects all of them. Two approaches to reinterpretation are relevant to this discussion. On the first approach, we reinterpret Alf as possessing a true belief regarding tharthritis. On this approach, we do not ascribe to Alf an incorrect belief that he has arthritis in his thigh, despite Alf apparently having several true beliefs about arthritis. Instead, we reinterpret all such beliefs as relating to tharthritis, since this reflects more closely the concept actually possessed by Alf. However, Burge is surely correct that that we don't normally reinterpret subjects, especially since it is common for subjects to partially misunderstand many terms, and we do not normally reinterpret a speaker as possessing an idiosyncratic concept when they misapply a term. On the second approach, we attribute a false metalinguistic belief that 'arthritis' is the correct term for the condition in his thigh. Here, we do not say that Alf falsely believes that he has arthritis in his thigh, and instead note that Alf has a false belief that the correct term to apply to the condition in his thigh is 'arthritis'. In response, Burge notes that only ascribing a metalinguistic belief to Alf in this scenario is implausible, since his

belief is not simply about worlds but about the condition in his thigh. While Burge's responses to these reinterpretations are plausible, as noted above, some internalists have responded by adopting variations of the reinterpretation approach. However, there are weaknesses in these responses, and I shall suggest that the two-component view I have developed provides the basis for a stronger form of the reinterpretation of the scenarios which avoids Burge's objections.

Crane (1991) has suggested that we combine both approaches and that in doing so we can reject social externalism wholesale. On this view, we reinterpret Alf as having a true belief about tharthritis, as well as a false metalinguistic belief that the term 'arthritis' is the correct term for the condition in his thigh. While this may provide a coherent interpretation of the scenario, a problem for Crane is that Burge seems to have the upper hand regarding ordinary attitude ascriptions in so far as saying that Alf believes that he has arthritis in his thigh, and so Crane's view is not strictly able to accommodate the intuitive interpretation of the scenario. In contrast, on the view I have developed we can say that the wide content associated with social externalism relates to Alf's false belief that he has arthritis in his thigh, and perhaps to the public concept that Alf incompletely masters. We distinguish this from the narrow content relating to Alf's concrete conceptual structure which represents the category *tharthritis*, and indeed his false metalinguistic belief regarding the correct usage of 'arthritis'. Thus, the two-component view has the advantage of being able to incorporate the idea that ordinary attitude ascriptions follow Burge's interpretation, whilst also highlighting that narrow content can play a central role in such cases. It is worth noting that Chalmers (2002) also supports the metalinguisitc interpretation, although on the approach he advocates Alf's arthritis concept does not seem to have a narrow content that relates to tharthritis while Alf's thoughts do have a wide content derived in the social externalist way. However, this seems overly concessive given the view of concepts developed in Chapter 2, which lends support to the approach I have outlined.

Burge does consider the combination of both reinterpretations, but he notes that this requires two things: establishing that the subject holds these two attitudes, and

justification for the denial of ordinary attributions. The two-component theorist needn't commit to this latter point about denying ordinary attributions and I have suggested that we can attribute to Alf's concept a wide content that relates to arthritis, hence that the ordinary attribution can be applied. This is consistent with also holding that Alf's concept has a narrow content that relates to tharthritis, and that this better reflects the nature of concepts construed as concrete mental entities. What favours this two-component view is that it can incorporate Burge's plausible claim that his account captures ordinary uses of attitude attributions, whilst also incorporating the highly plausible metalinguistic account and the reinterpretation of Alf's conception as relating to tharthritis. Indeed, as noted, Burge accepts that we may be able to identify Alf's conception, thus it's not clear why we should opt for Burge's view over the two-component view, especially given the support for the concrete view of concepts in the psychological literature. Hence, the internalist should combine both approaches as Crane has suggested, but also allow that Alf's beliefs may relate to both arthritis and tharthritis via the wide and narrow contents of his concepts. Thus, while Burge's interpretation is plausible as an account of ordinary attitude ascribing practice in such cases, if we accept that ordinary ascriptions may relate to a kind of wide content, this does not preclude the combination of both approaches to reinterpretation and a central role for narrow content.

3. Social externalism and the subject's conceptions

Goldberg (2002) develops and defends a view according to which social externalism can adequately capture a subject's conceptions, that is, the subject's way of thinking of X's. How S thinks of X's is given by the mode of presentation of S's concepts, thus, Goldberg's claim can be understood as being that the mode of presentation of a subject's concept can be given in terms of social externalist foundations. A second facet of the view, to be expanded on below, is that the social externalist base provides a minimalist account of modes of presentation, which requires only that we cite the relevant term in natural language, which is sufficient to convey S's mode of presentation. In support of this minimalist view, Goldberg rejects a picture in which modes of presentation may often be (and perhaps always are) expressed descriptively,

a view which, as noted previously, is supported by externalists like Putnam (1975), as well as internalists, including my own account developed in Chapter 2. I will raise objections to all three points, and explain why modes of presentation may be both internalist and at least partially minimalist and descriptive.

3.1 The minimalist account of conceptions

Goldberg develops his case for social externalism about mode of presentation by appealing to the sofa example first introduced by Burge (1986b), which is itself a variation of Burge's (1979) arthritis scenario discussed in Section 2. The sofa example does not rest on a subject incompletely mastering a term or incompletely grasping a concept (as in the arthritis example), but instead involves a subject holding a non-standard theory about sofas whilst being aware of how others' think of sofas. We are to consider a subject, S, who is aware that sofas are thought of as pieces of furniture made for sitting, but comes to believe that they are in fact religious artefacts or works of art (for simplicity I will stick with the idea that they are taken by S to be religious artefacts). S's non-standard theory entails that S falsely believes that sofas are religious artefacts and that sofas are not pieces of furniture made for sitting. In addition, S truly believes that other people believe sofas are pieces of furniture made for sitting. We are then to consider a counterfactual situation where things are such that the objects called 'sofas' in S's community really are religious artefacts and are not made for sitting (we can call these objects 'safos'). However, in this scenario, S also believes that people in his community think that safos are made for sitting, yet, we can suppose, his reasons for this are based on situations where people have merely been joking, and in fact they really do believe that safos are religious artefacts. S then believes truly that safos are religious artefacts, but believes falsely that other people believe they are pieces of furniture made for sitting.

Before considering Goldberg's view of modes of presentation, it is useful to have in mind some background regarding the typical consequences externalists may take to follow from the scenario. To recap, in the first scenario, S is aware of how others think of sofas, but has an idiosyncratic theory of sofas according to which they are

fragile religious artefacts. Social externalists might argue that, despite the idiosyncratic theory, S possesses the standard concept SOFA, and no other concept. As Sawyer (2007) contends, S's idiosyncratic theory may be encoded into S's mental file (a file comprised of all the information S associates with the term 'sofa'), but the concept we should attribute to S is the public concept SOFA. Sawyer is partly motivated by the claim that psychology requires generalizations to hold across the population, which requires that we ascribe the same concept, otherwise, her objection runs, there would be too few concepts of interest for psychology to study. In Chapter 2, I suggested that the relevant notion of generalizations does not require the same concept to be ascribed, since generalizations can be in terms of the conceptual structure type, such as prototypes, theories, and so on, without this requiring that all members of the population must be ascribed the same concept. Consequently, variation in the properties represented by a prototype is compatible with concepts as conceived in psychology.

With this background in place, the social externalist view of modes of presentation can be understood as being motivated by similar thoughts regarding the public meaning of 'sofa'. Where modes of presentation capture how a subject represents things as being from their perspective, for instance how they think of sofas as being, the social externalist contends this is reflected in the public meaning of the term 'sofa'. Thus, where S believes sofas are religious artefacts and that others think that sofas are soft comfortable furnishings made for sitting, the social externalist contends that the way S thinks of sofas is nonetheless given by the conventional meaning of 'sofa' in the relevant linguistic community. Such a view conflicts with the view developed in Chapter 2 according to which the mode of presentation associated with concepts (or cognitive content) should be understood as deriving from perceptual templates of the associated concepts and underlying perceptual states associated with the properties represented by concepts. On this view, we cannot read-off the mode of presentation of a subject's concept by taking the public meaning of the relevant term that expresses the concept, since the subject's mode of presentation relates specifically to the information they have encoded in long-term memory.

An initial worry with Goldberg's view is that it seems plausible, given S's idiosyncratic theory, that S would think of sofas as being religious artefacts, hence Goldberg's proposal fails to reflect how S really thinks of sofas. This idea is compelling since if the information is encoded in S's mental file, then activating this information seems sufficient for S to think of sofas in a way specific to his deviant theory. I will consider below how this kind of approach provides a suitable response to Goldberg's argument and is consistent with my own account of modes of presentation from Chapter 2. Before that, I will consider Goldberg's reasons for resisting this sort of view, since Goldberg is aware of this response and develops a case in support of the social externalist view.

According to Goldberg, when S thinks of sofas, despite having a deviant theory, he thinks of them *as sofas*. In support of this view, Goldberg contends that S's way of thinking of sofas can be specified purely by using the term 'sofa', with its conventional meaning implied, so that when we report how S thinks of sofas as being by simply stating 'as sofas', we are stating that S thinks of sofas in the standard way. This comprises a minimalist account of the subject's conceptions, or way of thinking of X's. Minimalism is therefore a claim about what we need to do to express how a subject conceives of things as being from his perspective.

Goldberg's defence of minimalism is related to the social externalist view of concepts noted above. On that view, it was held that even though S believes sofas are religious artefacts, S possesses the standard concept of sofas. In support of minimalism, Goldberg appeals to data regarding S's judgements, and holds that it is plausible that S would be disposed to utter the following:

> I mean to be using 'sofa' literally, as expressing the concept English associates with 'sofa'. I take it that this is precisely how my colinguals are using 'sofa'. (Goldberg 2002, 604)

One implication is that if we ask S how he thinks of sofas as being, then he is likely to respond with 'as sofas' which would then relate to the conventional meaning. The obvious response is to ask whether S might have different dispositions, and question

Goldberg's suggestion. It is not clear why we must hold that S would only be disposed to utter the above, since it appears to be just as plausible that S could have different dispositions to utter something like the following: I mean to be using 'sofa' in two different ways, such that when talking about what sofas really are I mean religious artefacts, and when talking about how others think of sofas, I mean objects made for sitting.[36]

It seems possible that S may think of sofas in the standard way, as when thinking about how others think of sofas as being, and yet S may also think of sofas as religious artefacts when thinking about what he understands as the real nature of sofas. On this view, S's sofa concept would be associated with two distinct mode of presentation, relating to the two ways he can think of sofas as being. One reason to favour this response from S is that in the scenario S is aware that how he thinks of sofas is different from how others think of sofas, and therefore we have reason to think that S can entertain both ways of conceiving of sofas. The worry for resting minimalism on such issues is that intuitions can flow in both directions.

Again, Goldberg has a response to this sort of worry, and provides two main reasons to reject this proposal. First, Goldberg claims that all we need is a minimalistic account where citing 'sofa' is sufficient to reflect how S conceives of sofas. According to this proposal an account of modes of presentation can be trivial in the sense that it is not necessary to provide a specific description of the various properties that capture the way S thinks of things as being. It should be noted that minimalism has some plausibility. For instance, if we want to explain how S thinks of sofas, using the expression 'as sofas' tells us something that is likely to be true, since S is able to think of sofas as sofas, where the latter occurrence of 'sofas' has its conventional meaning, e.g. [soft comfortable furnishing]. However, holding that minimalism is always sufficient to capture how S thinks of sofas is not as obvious as Goldberg appears to think as I will show below.

---

[36] Segal (2007) discusses this case in response to Burge's view. He contends that we need to recognise two different ways in which a subject might think of an object. I return to this point below.

The second main reason Goldberg provides in support of minimalism relates specifically to the proposal just identified relating to providing descriptions of S's mode of presentation rather than trivially citing the relevant term. Goldberg notes that the view that we must find a description of S's mode of presentation comprises the traditional view of mode of presentation. On the traditional descriptive view, modes of presentation can be given in terms of a description which reflects the kinds of properties a subject associates with the thing represented. For instance, the mode of presentation of WATER might be given in terms of the set of properties [clear, drinkable liquid that fills the lakes and rivers]. As already remarked, a descriptive account of how S thinks of sofas as being looks plausible given the scenario. We are told that S believes that others wrongly think of sofas as soft comfortable furnishings and correctly believes that they are in fact fragile religious artefacts. Put this way, we might be tempted to say that S thinks of sofas as [fragile religious artefacts].

Goldberg's central contention is that the traditional view is committed to always providing a specific descriptive account of S's conceptions, such that we must always provide a non-trivial specification. Triviality here pertains to the fact that in using the term 'sofa' to reflect how S thinks of sofas, the minimalist account does not need to do any more work, thus a non-trivial account of S's conceptions would involve describing in richer detail S's non-standard theory of sofas. The problem, according to Goldberg, is that this commitment should not be accepted, since there is no requirement that in all cases we need to provide a description to capture how S is thinking. This point is worth dwelling on since it is initially a plausible view. For instance, suppose on an occasion S is talking about sofas, but is not thinking of them in any particular way. In response to the question 'how is S thinking of sofas?', simply stating 'as sofas' would seem to be sufficient. In these cases it would be unnecessary, and potentially misleading, to provide a description which captured how S was thinking of sofas, since S might not be thinking of sofas in any specific manner that warrants a descriptive account. Now, if the descriptive view is committed to always providing a description, then minimalism looks promising since the requirement to provide a description each time S thinks of X's doesn't seem sufficiently well motivated in the face of the plausibility that on certain occasions S

121

is not strictly thinking of sofas in any particular way. However, it's not clear that Goldberg has correctly captured the position that a non-minimalist is committed to, and there are ways to resist minimalism without endorsing the descriptive view as outlined by Goldberg.

3.2 Responses to minimalism

I have identified three main parts of Goldberg's thesis (dispositions to assent, the sufficiency of minimalism, and anti-descriptivism) which can be subjected to criticism. In this section I show how each part can be rejected, and raise additional objections to the view.[37]

3.2.1 Subjects' judgements about their conceptions

Recall that Goldberg suggests S might be disposed to say that he intends to use 'sofa' to express the same concept others express. It is possible to accept Goldberg's claim about what S might be disposed to say in the sofa case, while holding that S's concept SOFA, is associated with two modes of presentation. The step from S's dispositions to what his modes of presentation are like has not been made. For instance, by accepting that S intends to be using 'sofa' to express the concept others express, S might mean that the meaning of his term 'sofa' is intended to conform with normal usage. This does not entail that the mode of presentation is fixed by the public concept S intends to express. Furthermore, as was noted above, it seems compatible with the scenario that S is just as likely to report that he is using 'sofa' to mean what he now believes it to mean, i.e. that it refers to religious artefacts. S might even have a more sophisticated response, relating to his awareness that although he knows how standard English speakers use the term, he thinks that it correctly applies to religious artefacts. The main worry then with resting the case for social externalism about conceptions on what S might say is that short of conducting a much larger study on what subjects think of their conceptions, we can debate whether S would only have the dispositions Goldberg has identified.

---

[37] Bach & Elugardo (2003) develop an alternative critique of Goldberg's view.

One reason why an externalist might be tempted to restrict S's possible replies to what 'sofa' means in English is the assumption noted in Section 2 regarding the social basis of natural language meaning. If 'sofa' only has meaning by virtue of public conventions, then it would make little sense for S to say that his use of 'sofa' could possibly have another meaning. But if we resist that view, there is space to allow that S may be using the term differently, and so be disposed to say something different. Most importantly perhaps is that it only really makes sense to hold that this entails anything about minimalism because S would think of sofas only in terms of the conventional meaning of the term. But this has not been shown to be the case.

3.2.2 The cogency of the minimalist account

The second response is that a proponent of a descriptive view need not be committed to S always thinking non-trivially of X's. A descriptivist could hold that sometimes we need to use a description to capture how S thinks of X's, such as when S is thinking of X's in a particular manner. If minimalism is associated with the view that sometimes S does not think of X's in any particular way, then this is consistent with also holding that sometimes S does think of X's in a way which requires a descriptive reconstruction. However, this is not Goldberg's view, since he contends that minimalism is always sufficient to capture S's conceptions. A worry with this combination of minimal and descriptive conceptions is that minimalism may in fact presuppose a descriptive foundation. The central idea of minimalism is that we can capture S's mode of presentation by citing the relevant term, such as 'sofa'. However, any mode of presentation generated by citing 'sofa' seems to derive from the associated conventional meaning of the term in the linguistic community e.g. [soft comfortable furnishing made for sitting], and hence there is a mode of presentation provided by the minimal view only because of this descriptive conventional meaning. Unless we can make sense of how S thinks of sofas by glossing 'sofa' in terms of some sort of descriptive conception, there is no obvious way to get a handle on how S is thinking at all, since citing 'sofa' without the prior knowledge of the conventional meaning would fail to express any mode of presentation.

3.2.3 The architectural basis of conceptions

The final response relates to the social externalist foundations of the view. As noted above, the social externalist can grant that there is information encoded in S's mental file that reflects S's idiosyncratic theory of sofas. On Goldberg's account, S's mode of presentation relates to the conventional meaning of the term in the linguistic community. Now, in order for S's mode of presentation to relate to the conventional meaning of 'sofa', it is plausible that this must relate to S having also encoded the relevant information in the mental file for SOFA, otherwise, it might be argued, there is no connection between mode of presentation and S's concepts construed as psychological entities associated with mental files. The point here relates to the architectural basis of mode of presentation outlined in Chapter 2. There I developed the view that the mode of presentation of concepts, construed as concrete psychological entities, derives directly from encoded information in long-term memory. The capacity to think of X's as being a certain way requires having such information stored in memory.

If this view is correct, then there are two consequences for Goldberg's view. First, we should allow that S associates two modes of presentation with SOFA, relating to information encoded in long-term memory (or his mental file). Second, the only reason that S can think of sofas as sofas (i.e. associate the standard conception of sofas with 'sofa'), is because of this information in memory. It is possible that S may fail to think of any specific properties of a category when deploying a given concept. For instance, upon introspection, S does not focus on any particular properties of sofas. In this case, S may use the concept without thinking of the category in any particular way. This seems a reasonable assumption to make, and is compatible with pluralism which allows that some conceptual structures may lack distinctive cognitive content (for instance, as Weiskopf (2007) notes, prototypes are good candidates for having cognitive content, which suggests others might lack cognitive content). Here, a minimal account of S's conceptions would seem appropriate, and citing 'sofa' without implying any specific mode of presentation would perhaps be reasonable. However, the architecture supports the idea that in other cases S can think of a category in terms of, say, the prototypical properties S

associates with the category. At the very least, the scenario described suggests that S is likely to think of sofas in both ways.

The point to note here is that were Goldberg to challenge this approach to concepts and cognitive content, say, by holding that concepts are abstract entities, not concrete, then the same objections raised against Burge in Section 2 will apply. The problem with holding that social externalism, and the relevant social basis of linguistic meaning, plays a role in individuating either concepts or cognitive content, is that this view appears to be at odds with work in psychology which treats concepts and their cognitive content as concrete psychological entities. Without first showing that the social externalist is entitled to hold such a view, or that a version of the dual view of concepts can be happily defended, then neither the arthritis nor sofa examples should be taken to have externalist implications for concepts and cognitive content. This is not to deny that there may be some kind of wide content associated with both examples.

3.2.4 Concluding remarks

Before moving on, it is worth summarising the criticism of Goldberg. While Goldberg can accept that S may think of sofas as religious artefacts (such as where S's deviant theory of sofas allows S to entertain the thought that sofas are religious artefacts), Goldberg's view is that citing 'sofa' adequately captures how S thinks of, or conceives of, sofas. The obvious problem here is that from the scenario it seems that S can think of, or conceive of, sofas in at least two distinct ways. This makes the minimalist claim a bit odd, and it is why Goldberg brings in the critique of the descriptive view. Thus, combining the claim that 'sofa' is adequate with the claim that the descriptive view is too strong, Goldberg intends to support the minimalist claim. Following Segal (2007), we can introduce a distinction between the two ways S can think of sofas, as soft comfortable furnishings made for sitting ($sofa_1$) and religious artefacts ($sofa_2$). Nothing Goldberg says explains what is wrong with this, and the view developed in Chapter 2 provides a systematic basis for identifying these differences with difference conceptual structures. Pursuing this further, the minimalist view commits to an odd view of cognitive content. Cognitive content

reflects how the subject represents the world as being. If we were to say that the cognitive content of S's SOFA was just given by the standard meaning of 'sofa', or simply cited 'sofa', this would seem to be inadequate as an account of S's cognitive content since S can think of sofas in at least two distinct ways. When S thinks of sofas under $sofa_1$, then the standard conception is appropriate, but when she thinks of them under $sofa_2$, we need to draw on S's idiosyncratic theory of sofas to properly reflect the cognitive content associated with this way of thinking of sofas.

Now, I do not need to reject the minimalist view entirely. In fact, it seems plausible with regard to ordinary language and how we might talk about S's conceptions. However, this does not show that we can do without a richer account of cognitive content which reconstructs the subject's conceptions or way of thinking about the world in terms of a description which more accurately captures their specific cognitive content. Supporting a two-component picture avoids the problems with opting for either view exclusively, and appears to better account for the data in these cases, i.e. intuitions about ordinary attitude ascribing practice, and concepts as understood in the psychological literature. The worry for Goldberg, according to the view of concepts and cognitive content I have developed, is that we cannot neatly distinguish between S's (public) concept of sofas, and her theoretical beliefs relating to sofas, since on my view these are intimately related. Furthermore, S's cognitive content associated with SOFA is related to this information, which we can use descriptions to capture, and to distinguish between $sofa_1$ and $sofa_2$.

4. Externalist individuation in perceptual psychology

In the second part of Chapter 3 I will consider two versions of the view that perceptual psychology is fundamentally at odds with internalism. The challenge these views pose to internalism and narrow content is two-fold. First, on one view narrow content is not required in perceptual psychology. On another, the ascription of narrow content is deemed problematic. One possible consequence of these views is that they may imply, or motivate, general anti-internalism in psychology. This would put pressure on the position developed in Chapter 2 where I argued that the

conceptual structures from psychology have, and may be individuated by, narrow contents.

I begin, in Section 4, by considering Burge's view that psychology presupposes an externalist view of psychological individuation. Should this view be correct, it stands as a direct competitor to the internalist view I developed in Chapter 2, since if the relevant states are individuated broadly, then mental content will be wide, supposing that such states are individuated by their contents. Burge's case for externalism about psychology rests on his interpretation of certain theories and phenomena in perceptual psychology. In response, I will show that a proponent of narrow content can accept that some states and processes may have wide contents but that this does not show that there is no useful notion of narrow content for both perception and thought.

4.1 Burge on the individuation of perceptual representations

In several works, Burge (1986a; 2007; 2009b; 2010) makes the case for externalism about the individuation of perceptual representations based on claims about how such representations are individuated in psychology. The challenge this approach raises to proponents of narrow content is that it purports to demonstrate fundamental externalist principles operative in psychology which give no role to narrow content. Burge's central thesis is summarised as follows:

> …the natures of…perceptual states are what they are through a systematic network of causal relations between instances of the environmental attributes and processes that entered into the formation of the specific kinds of perceptual states that an individual is capable of being in and that are as of (and even specify) those environmental attributes. (Burge 2010, 85)

On this view, a given perceptual representation is what it is by virtue of its bearing patterns of causal relations to objects and properties at the subject's actual environment, and these relations are constitutive of the state's status as a particular type of perceptual representation. For instance, a perceptual representation of some object O is a representation of O by virtue of constitutive relations between S and O at S's actual world. If this claim is right, then the perceptual representations of

intrinsic duplicates S and T, whose perceptual states were formed on Earth and Twin Earth respectively, will be type-distinct because each subject's representations are individuated by constitutive relations to the relevant kinds at their respective environments.

At this juncture it is worth pointing out that perceptual representations are understood as accurate representations of specific distal objects/properties, and as such the relevant states are construed as essentially wide states because successful or accurate perception presupposes both relata. For instance, for the statement 'S perceives O' to be true, it is necessary for S to stand in the right kind of relation to O. If S were to merely hallucinate that there is an O, then *S perceives O* would be false. A proponent of narrow content can accept that perceptual states such as *S's perception of O* are wide mental states. In Chapters 1 and 2 I defended the view that the content of general concepts is narrow and that this is compatible with accepting that singular thoughts have wide content. In order for Burge's account of perceptual psychology to raise doubts about the relevance of the notion of narrow content in psychology, Burge would need to show that there is no equivalent notion of general narrow content in perceptual psychology, and/or that individuation in psychology is in principle opposed to individuation by narrow contents.

Now that we have Burge's view in focus, I will consider two examples Burge draws on to demonstrate that perceptual psychology is fundamentally externalist. In work that prompted an extensive debate between externalists and internalists, Burge (1986a) aimed to show that a contemporaneous computational theory of human visual perception developed by Marr (1982) could be interpreted as presupposing a version of externalism. More recently, Burge (2009b; 2010) has made the same claims about perceptual psychology in general, and one example he appeals to is the phenomenon of colour constancy. I will consider these two cases and assess whether Burge's case for externalism shows that there is no suitable role for narrow content in perceptual psychology.

4.1.1 Marr's theory of vision

The first example comes from Burge's (1986a) interpretation of Marr's (1982) computational theory of vision. Marr's project was to provide an account of visual representation that explained the construction of a three-dimensional representation of objects from two-dimensional retinal images. Burge quotes a passage from Marr (1982) which he takes to show that Marr's account of perceptual representations presupposes externalist individuation:

> The purpose of these representations is to provide useful descriptions of aspects of the real world. The structure of the real world therefore plays an important role in determining both the nature of the representations that are used and the nature of the processes that derive and maintain them. An important part of the theoretical analysis is to make explicit the physical constraints and assumptions that have been used in the design of the representations and processes...It is of critical importance that the tokens [representational particulars] one obtains [in the theoretical analysis] correspond to real physical changes on the viewed surface; the blobs, lines, edges, groups, and so forth that we shall use must not be artefacts of the imaging process, or else inferences made from their structure backwards to the structures of the surface will be meaningless. (Marr 1982, 43-44)

From this passage, Burge (1986a, 27) makes the following point:

> Marr's claim that the structure of the real world figures in determining the nature of the representations that are attributed in the theory is tantamount to the chief point about representation or reference that generates our non-individualist thought experiments… (Burge 1986a, 27)

The important point for Burge's interpretation is where the *real world* plays a determining role in the nature of the representations attributed. With this in mind, Burge (1986a) considers a Twin Earth scenario involving a subject on Earth, call her 'Visua'[38], and another intrinsically identical subject, Twin Visua, on Twin Earth. On Earth, Visua perceives certain shadows and sometimes misperceives visually indistinguishable cracks as shadows. On Twin Earth, there are no shadows, and so Visua only ever perceives cracks. Visua and Twin Visua's perceptual experiences are subjectively indistinguishable, that is, things seem the same from each subject's perspective. According to Burge, however, Visua would represent shadows as

---

[38] The name 'Visua' is applied by Segal (1988) in his discussion of Burge's example.

shadows as well as cracks as shadows, whilst Twin Visua would represent the cracks as cracks. This is taken to follow as a consequence of the reading of Marr which says that attributions of representations are made by reference to the normal distal objects detected, and which are causally responsible for the formation of those states.

The importance of this consequence for Burge is that despite Visua and Twin Visua being in identical internal states, and things being subjectively indistinguishable, their perceptual representations are type distinct. If Burge is correct, then the contents of perceptual representations will be wide, relating to the specific properties represented by internally identical, but type distinct, perceptual states.

4.1.2 Colour constancy

A second example that Burge uses to support this view is that of colour constancy (Burge, 2010). Colour constancy is the phenomenon that the perceptual system can represent the colour of objects as being the same even through changes in ambient lighting which affect the appearance of the object. For instance, consider the representation of a red wall where the perceptual system keeps track of the property *red*. If a different coloured light was projected onto the wall, then despite the change affecting the phenomenology of the perceiver's perceptual experience of the wall, the perceptual system keeps track of the objective colour of the wall and represents the same property, *red*. Consequently, the property represented is constant through changes in ambient lighting which may affect the phenomenology of the relevant state. According to Burge this shows that perceptual representations are fundamentally externalist in that they represent specific environmental properties.

A common theme can be identified in these two cases. In the Marr example, Visua and Twin Visua had subjectively identical experiences, and yet were held to represent different properties. In colour constancy, what is represented is likewise not explicitly tied to how things seem subjectively in experience, as perceptual representation relates to specific features of the local environment. What this suggests is that perceptual content does not relate to subjective features of experience, such as phenomenal character. Burge specifically notes that the relevant cases in

perceptual psychology make no reference to phenomenology, apparently holding that all mental content is wide. On the face of it, this would rule out the account of narrow content developed in Chapter 2 since this was based partly on the phenomenal properties of perceptual states.

4.2 Responses to the examples

In this section I will consider two responses to Burge's case for externalism, and show that narrow content is consistent with psychology.

4.2.1 Computational content and the contents of experience

The first point to note is that there is still debate about the consequences of Marr's theory for the notion of content. Segal (1988) and Prinz (2000) have outlined internalist construals of Marr's theory which aim to show that the theory is compatible with ascribing narrow representational content. Both Segal and Prinz contend that Visua and Twin Visua could plausibly represent the same heterogeneous property, aptly named 'crackdow' by Segal, to reflect that Visua cannot distinguish between the property of being a shadow or a crack. This property is shared by shadows and cracks, and relates to the way in which both properties appear subjectively indistinguishable to Visua. Such a view would stand as a counter to Burge's claim that Marr and other perceptual psychologists appeal to the specific properties at the relevant environment in order to individuate the relevant representations. A different approach is taken by Egan (1992; 1995; 2009; 2010) who has defended the view that computational processes are not individuated by their contents. If this view is right, then Marr's representations should not be individuated by their contents at all.[39]

I do not consider any of these positions in detail here, as this would require much more space than is available. Instead, I want to suggest a way that narrow content can be motivated even if we accept much of Burge's position. Suppose that Burge is right about Marr and that perceptual psychologists typically construe computational

---

[39] I consider Egan's (2009; 2010) position in more detail in Section 5, focusing on her objection to ascribing narrow contents to computational representations.

representations as fundamentally externalist. Rather than arguing that perceptual psychologists should attribute heterogeneous narrow contents to computational representations, a different kind of point stemming from the issues discussed by Segal and Prinz can be raised. What seems to drive their accounts of the Visua scenario is that an important feature of the scenario relates to the fact that Visua and Twin Visua enjoy states that are subjectively indistinguishable. By introducing a common narrow content that represents a more generic property, *crackdow*, shared by shadows and cracks, this approach aims to capture the sense in which things are subjectively indistinguishable to Visua and Twin Visua. However, what may cause a problem for this proposal in relation to the Marr debate is that it's not obvious that at the computational level there is a need to introduce narrow contents. In addition, the proponent of narrow content needn't aim to refute all bases for wide content, which has typically been the approach pursued in the literature. However, this is not to accept that there is no interesting notion of narrow perceptual content.

A different approach would be to distinguish computational perceptual contents, which might be wide, from other possible contents of perceptual states and representations.[40] In particular, suppose that perceptual experiences have their own contents, distinguishable from the content associated with computational states or processes. This is plausible in the colour constancy scenario, where a subject may keep track of the objective colour of the wall, whilst undergoing experiences with different phenomenal characters. In Chapter 2 I introduced Tye's (2000) account of phenomenal character according to which it is identical to a kind of representational content. It is then possible to distinguish two contents in the colour constancy case: the content associated with objective perceptual representations involved in colour constancy, and the content associated with the changing perceptual experiences undergone by the subject. Assuming that perceptual experiences have content, we can draw on the account of phenomenal internalism defended in Chapter 2 to show that colour constancy is consistent with narrow contents. Thus, even granting that some strand of perceptual psychology may interpret certain states as representations

---

[40] On the Marr debate, Sprevak (2010) has recently made a distinction between computational and intentional representational contents. Such a distinction could in principle lead to possibilities of perceptual representations having both wide and narrow contents. I set this distinction aside here.

of specific properties at the local environment, this does not rule out narrow contents entirely. What this could mean is that there are two kinds of perceptual content, relating to different aspects of perceptual representational states. Wide content reflects the representation of the specific environmental property, whilst narrow content could reflect the phenomenal character of the relevant perceptual experience.

In response to this bifurcation of perceptual content, Burge could hold that the supposed narrow content is not really content, as it merely reflects phenomenology and doesn't represent anything. This response is a version of a common objection raised against accounts of narrow content. The point to make in response is that the narrow content of experience can be representational. In Chapters 1 and 2 I drew on Shoemaker's (2006) account of appearances to show that there is a kind of property that relates to the narrow content of certain concepts. On Shoemaker's view, objects have multiple phenomenal characters, since phenomenal characters are held to be properties of objects, relating to dispositions to cause experiences in perceivers. It is possible to alter the view so that the contents associated with phenomenal character are the dispositional properties themselves while phenomenal character is a property of the subject's experience, relating only to how things seem subjectively to the subject. Thus, we can understand appearance properties in terms of the dispositional properties of objects to cause states with a given phenomenal character in the subject. Allowing perceptual experiences to represent appearances is a plausible account of the contents of experience. Returning to the colour constancy example, where S undergoes a perceptual experience relating to the projected colour on the red wall, S's experience may represent a dispositional property construed as a disposition of the wall together with the light source to produce the relevant phenomenal character of S's experience. Consequently, narrow content can be representational and of interest in psychology.

4.2.2 Concepts and content

The second main response is to reject the implication that psychology in general is fundamentally externalist. Burge's externalist construal of psychology seems to arise predominantly from reliance on the idea that perceptual representations must be

individuated by wide contents. If this view was generalized to psychology, then an externalist might hold that the view of concept individuation outlined in Chapter 2 is incorrect. For instance, where computational representations and colour constancy mechanisms represent properties at a subject's actual environment, the same point may be taken to hold of concepts. However, I showed above that it is consistent to hold that computational and other perceptual representations may have wide contents, whilst also holding that the contents of perceptual experiences are narrow. When we consider the treatment of concepts in psychology, externalism seems even less central.

The problem in applying Burge's strategy to the individuation of concepts is that when we consider the range of concepts discussed in Chapters 1 and 2, such as appearance, artefact, and functional concepts, then wide contents seem less suitable than narrow contents. Take a subject's ALGORITHM concept. It is intuitive to hold that whatever is represented is common to possible worlds $W_1 \ldots W_n$ since differences in the relevant natural kind properties at these worlds will make no difference to the content of the concept ALGORITHM, as it will represent the same functional property irrespective of the relevant world. Similarly, concepts that relate to appearance properties (such as those associated with colour perception) may have narrow contents that relate to the representation of dispositional properties, serving as suitable contents for both perceptual experience and general concepts of appearances. Consequently, Burge's construal of perceptual psychology can largely be accepted, without this posing a threat to narrow content, and more importantly to my claim that general concepts may be individuated by their narrow contents.

5. Computational psychology and the case against narrow content

In this section I evaluate Egan's case for rejecting narrow content in perceptual psychology.

5.1 Egan on computational individualism and mathematical content

In Section 4 I noted that Burge's construal of Marr's theory is challenged by Egan. Egan (1992; 1995; 2009; 2010) contends that computational perceptual

representations are specified without talking about what they represent, and that content ascriptions are based on pragmatic considerations to do with our interests in what a computational system functions to detect at its normal environment. Egan's central claim is as follows:

> It would be a mistake…to conclude that the structures posited in computational vision theories must (even in the gloss) represent their normal distal cause, and to find in these accounts support for a causal or information-theoretic theory of content…Even within a single content assignment (cognitive interpretation), no single, privileged relation is assumed to hold between posited structures and the elements to which they are mapped. The choice of a cognitive gloss is governed by explanatory considerations, which we can, following Chomsky, characterize as 'informal motivation'. (Egan 2010, 257)

Egan therefore contends that the notion of content plays no role in the theories of scientific psychology. The idea is that a theory like Marr's is informally motivated by considering how visual perception works and how it allows external objects to be represented three-dimensionally. The scientific theory, however, only deals with internal properties of the representational system, and talk about representational content amounts to a semantic interpretation of the device based on informal motivation and our parochial interests in what computational mechanisms may represent at their normal environments. Thus, according to Egan (2009; 2010), representational content does not play a part in the central aims of computational perceptual psychology, which are to explicate the computational processes involved in perception. Instead, computational psychology is individualistic, so that the kinds of perceptual states of interest in psychology should be individuated purely in terms of the computational processes themselves. The status of the computational structure as a representation is therefore different from the sense in which Burge and others take mental states to be representations of particular environmental kinds.[41]

While Egan places content outside of serious scientific theory, she holds that the informal contents ascribed will be wide, not narrow. The central reason for this is

---

[41] Sprevak (2010) has recently argued that computational processes are individuated by their contents and hence that Egan's view is incorrect. I will not be concerned with Egan's account of Marr here, as my interest is in her critique of narrow content.

that, as noted, pragmatically ascribing content to a given computational process relates to the informal motivations of the theory. The idea is that we consider what initial questions motivate cognitive scientists to seek a computational description of a given process. In the case of Marr's theory the informal motivation can be understood as how a three-dimensional representation of some particular distal stimuli, such as a ball, can be constructed from a two-dimensional image constructed from light impacting on the retina. Once the computational description is complete, we semantically interpret the relevant computational process by reference to what motivated the search for an account in the first place. Consequently, the semantic gloss relates to the actual distal objects, and the content ascribed is wide.

5.2 The case against narrow content

To show why we should prefer wide content ascriptions over narrow content ascriptions for computational representations, Egan considers a version of the Twin Earth scenario. The scenario aims to motivate the ascription of wide content by showing that ascriptions of narrow contents are problematic, and that our interests tend to relate to the actual local causes of perceptual representations. On Earth, Visua is a mechanism that functions to detect changes in depth (property C_1), and hence is a component in a larger perceptual system in an organism. Twin Visua is internally identical and is said to spontaneously appear on Twin Earth as part of a fundamentally different organism, and detects a different unspecified property (C_2), which Egan says is a function of local light, surfaces, and local optical laws (hence it is a perceptual property of some kind). By considering what an account of narrow content is committed to in this scenario, the preference for wide content follows from showing a) that narrow content is problematic, and b) that wide content is preferable.

Egan considers and rejects two possible versions of narrow content in this scenario. On the first version, Visua represents a set of properties which includes C_1 and C_2, as well as many other properties ($C_1...C_n$) which would cause Visua to go into state S at possible worlds $W_1...W_n$. This view treats the content as either disjunctive (C_1 OR C_2 OR…) or conjunctive (C_1 & C_2 &…). An initial problem for this kind of view is that disjunction makes content indeterminate, whereas if we are considering Visua

only at W_1 it is initially plausible that Visua would represent a single property, and conjunction makes content significantly bloated, and again we might prefer determinacy of content.

The second view avoids these sorts of worries and takes Visua to represent an intermediate property, C_3, which subsumes both properties C_1 and C_2. The content ascribed to Visua and Twin Visua would relate to this single property C_3. Again, this view is familiar from Chapter 1 and Section 4 above, having the same form as the property *crackdow*. However, Egan contends that attributing C_3 to Visua and Twin Visua would be strange. The reason it would be strange seems to be a consequence of the fact that ascribing C_1 on Earth and C_2 on Twin Earth is more natural than introducing an apparently ad hoc property C_3, in part because C_1 and C_2 are, by hypothesis, such wildly different properties. Any property C_3 that subsumed C_1 and C_2 would be less natural than either C_1 or C_2, because C_1 and C_2 are specific environmental properties that could in principle be detected by scientific investigation, whereas C_3 is not. If C_3 related to these different properties, it would be strange because, so the intuition goes, C_1 and C_2 just don't seem to have anything in common. This point seems plausible when taken for perceptual mechanisms like Visua which are understood to detect specific environmental properties.

Egan supplements this criticism of narrow content by asking us to consider a publisher being commissioned to produce a textbook for users of the device, and the decision of whether to create world-specific textbooks which explain what Visua would represent at each world, or a single textbook which ascribes to Visua a narrow content which is either disjunctive/conjunctive or subsuming (C_3). Egan's case for wide content being more useful is that having to hand a description of what Visua represents at a specific world is much more useful than a description that says Visua will detect any of $C_1 \ldots C_n$, or instead C_3. This is because users would likely be interested only in the specific distal property at their environment. Being told what other properties may be detected or that a more general property would be detected would therefore be less useful, perhaps irrelevant, to users of the textbook.

5.3 In defence of narrow content

There are three main responses to Egan's case against narrow content that I will consider here. First, Egan's scenario may not relate to the internalist's supervenience claim outlined in Chapter 1, which would mean that her critique of narrow content can be ignored. Second, the mechanisms in the scenario are not cognitive subjects, and so the contents of these mechanisms may be subpersonal, as such, wide content may be appropriate whereas narrow content may be more suitable for personal level contents attributed to subjects. Finally, the critique of subsuming narrow content may be successful in the particular example, but when we consider the contents of perception and thought, subsuming contents are much more plausible, thus the example fails to show that all applications of such narrow content would be ad hoc. I consider each of these responses in more detail in the rest of Section 5.

5.3.1 Supervenience and kinds of narrow content

Egan's main focus is in showing why narrow content can't be representational. However, the internalist needn't be worried about this for two reasons. First, consider the nature of narrow content as intra-world narrow. Twins share content within a world, or at worlds that are nomologically similar such that things seem the same at the relevant worlds. In these cases, Twins will be in states with the same narrow content. It is not clear how Egan's scenario fits into this picture. Since we are not considering whole subjects, we cannot tell whether things would seem the same despite the different environmental properties being represented. Consequently, it is possible to hold that Visua in each scenario may have a certain kind of narrow content that does not relate to how things appear to the subject or to real world properties that are represented. This leads to the second point. The case at hand may best be understood by appealing to a version of narrow content as functional or conceptual role as articulated by Block (1988) and Loar (1988). Such narrow content is not representational, but relates instead to an abstract account of a state's relations to other states. This may be suitable in the case of Visua where the perceptual subsystem will bear relations to other perceptual states, playing a systematic functional role in that system. This would lead to a further addition to the pluralist view, allowing the distinction between narrow contents that relate to

abstract functional roles, and those which relate to cognitive content and are genuinely representational.

Two further related issues are relevant at this juncture. Noting the peculiarities of the example, where Visua is a device within a larger perceptual system, this raises a question of what sort of content devices within perceptual mechanisms bear. Furthermore, when considering Egan's criticism of narrow content as a general attack on narrow content, it may be appropriate in the specific example to hold that narrow contents are problematic. However, as already indicated above, when we consider normal subjects it is plausible that there are various kinds of properties that they can represent, not all of which will relate to specific environmental properties of the kind Visua is taken to detect. I consider these two remaining points in more detail below.

5.3.2 Personal and subpersonal content

A fundamental issue of interest is what sort of device Visua is and what sort of content it should be attributed. Evans (1982), and more recently Bermudez (2009), hold that the content of certain subpersonal states are not consciously accessible by subjects, although such content may play a relevant role in cognitive tasks. In contrast, personal level states will have content that is consciously accessible. An example of consciously accessible content relates to states with phenomenal character. For instance, a perceptual experience of a red ball will have phenomenally conscious content relating to the redness of the experience. In light of this, states that have such content might include mechanisms involved in perceptual processing, which may also relate in some way to states which we lack conscious access to. For example, we do not have conscious access to all of the kinds of representations attributed to the visual system by a theory like Marr's, and consequently many of these perceptual representations will be subpersonal, however, we do have access to the content of other representations, i.e. the ones with personal level content. In terms of the subpersonal/personal distinction, Egan's example may fail to support her

point since a case can be made that even if some subpersonal content is wide, we might think that personal level content is sometimes, if not always, narrow.[42]

One reason to think that the mechanisms in Egan's example are good candidates for possessing subpersonal content is because Visua and Twin Visua's contents do not seem to be accessible to consciousness. Such content could be wide content, relating to the distal properties represented by each mechanism at the relevant world. Granting this, however, does not cause problems for proponents of narrow content.

Recall from Chapter 2 and Section 4 above, where I held that there is a kind of narrow content relating to perceptual experience and to phenomenal character in particular. We can assume that the contents of Visua and Twin Visua will not share phenomenal character, since by hypothesis C_1 and C_2 are wildly different properties. Thus, we should assume that they would not cause states with the same phenomenal character. This is a problem for Egan's contention that we should reject narrow content on the basis of her example.

I have noted a distinction between personal and subpersonal content, and conceded that some subpersonal level content may be wide, while narrow content is perhaps best placed to play the role of personal level content as it reflects how things are represented as being. However, both subpersonal and personal states could have narrow contents and I do not need to speculate here on their connection, if any. Thus, while Visua may have subpersonal narrow content, which does not relate to how things appear to be or to real-world properties, such content may be purely functional in nature.

A problem for Egan is that her scenario is very unclear, both as it stands and for the implications it purports to have for whole perceptual systems and subpersonal and personal level content. If we suppose that Visua lacks cognitive content, being

---

[42] Kriegel (2012) makes a similar point, and argues that wide content is always subpersonal, while narrow content is always at the personal level. Kriegel's account of personal level narrow content is different from mine, relating to the notion of self-representation and phenomenal intentionality. I will not consider that view in any detail here.

simply a subpersonal system that detects changes in something other than an appearance property, then the two-component view can coherently hold that Visua has a wide content that represents the relevant property at the actual world, and perhaps a narrow content that reflects the functional role of Visua within the wider perceptual system of the organism she is a part of. However, this poses no threat to the cognitive and representational role of narrow content when we consider different scenarios involving the narrow states and contents of whole perceptual systems and subjects. Consequently, Egan's scenario and interpretation pose no threat to the nature and role of narrow content, and as I have noted there is a case that some form of narrow content may play a role within her scenario.

Consider the standard Twin Earth scenario. Take S's prototype concept WATER that represents the same set of superficial properties as T's prototype concept. Suppose that S and T's states also have a kind of wide content, relating to the externalist's notion of content where S represents H_2O only, and T XYZ. When we are considering the contents of S's mental representations, from S's perspective she may be unaware that at some level of description there is a kind of content that relates exclusively to H_2O, and similarly for T and XYZ. Such content would apparently not be consciously accessible, that is, it would be subpersonal content. The narrow contents of S's prototype, on the other hand, will likely be personal content as it is consciously accessible to S. This can be understood in terms of the basis of this narrow content in the appearances of the properties represented, which relate directly to how S and T represent things as being from their perspectives.

Returning to Egan's scenario, Visua and Twin Visua are taken to represent different properties. Suppose that C_1 and C_2 do appear the same to the relevant organisms. In this case, then there could be a kind of narrow content relating to appearance properties. If they do not appear the same, then the contents are likely subpersonal, since if Visua and Twin Visua are intrinsic duplicates, then by hypothesis things must appear the same to them if they bear contents that are at the personal level, since, as I argued in Chapter 2, phenomenal character supervenes on intrinsic properties. Consequently, an internalist can accept that Visua and Twin Visua would

have wide contents, because the paradigmatic narrow content developed in Chapters 1 and 2 seem closer to the notion of personal content, reflecting how subjects represent things as being from their perspective. This suggests that an internalist might not even need to argue against Egan that Visua and Twin Visua would bear wide content. Again, the example suffers from not being particularly clear, especially with regard to implications for the content of the perceptual representations of normal subjects, and the content of concepts.

5.3.3 The contents of perception and thought

A final point to note then is that when we consider the contents of perceptual representations and concepts of normal subjects, the example fails to show that any of the varieties of narrow content considered are problematic. I have already granted that in the specific case of Visua and Twin Visua representing C_1 and C_2, the internalist can accept that representing C_3 or conjunctive/disjunctive contents would be implausible and ascribing wide contents more useful. However, these kinds of narrow contents are much more palatable in the case of perceptual representations and the contents of certain concepts.

As discussed in Section 4, perceptual representations seem to be capable of having a range of contents that are not restricted to the specific distal causes of those states. For instance, it is plausible that dispositional properties are represented in so far as perceptual states might have general, as opposed to singular, contents. A perceptual state S might have a wide content relating to the distal cause of that state, such as red, yet S might also represent a more general property of *appearing red*, which is a kind of dispositional property that relates to narrow content. As I also indicated above, a representational theory for general concepts may countenance more properties than basic scientific properties like changes in depth (C_1). Certain properties are plausibly identical at worlds $W_1...W_n$, and many of them subsume different underlying natural kind properties. Take the property of being an algorithm that does not depend on the specific fundamental properties at $W_1...W_n$. Representation of such a property may not be suitable for a device like Visua, which merely tracks specific environmental properties, but is a typical property represented by general

concepts. Returning to the property of appearing red, this property is sharable by kinds with different fundamental properties but which cause in S and her intrinsic duplicates experiences with the same phenomenal character. Dispositional redness would therefore be similar to C_3 in being a more general property than the underlying fundamental properties, and in a sense subsumes them. In Egan's example, the proposed subsuming properties do appear strange. This strangeness, however, is only because Egan's scenario makes them so, and some of the properties represented by human perception and thought can be understood as subsuming different natural kind properties without naturalness being a problem.

6. Conclusion

In Chapter 3 I considered two approaches to putting pressure on internalism and narrow content. Social externalism purports to show that concepts and cognitive content can relate directly to facts about the linguistic community. In response, I showed that Burge's arthritis argument and Goldberg's minimalist view face several difficulties. Most importantly, they appear to be at odds with work in psychology. The second approach appeals to work in perceptual psychology and purports to show either that the science is fundamentally externalist, or that the ascription of narrow content is highly problematic. Against Burge I showed that his construal of the science is consistent with the ascription of narrow contents, thus a two-component theorist could endorse both wide and narrow contents. I also suggested that his case for perceptual externalism doesn't provide support for externalism about concepts. Against Egan's critique of narrow content I noted that at the computational level there is a case that some forms of narrow content may not be required, but that Egan fails to rule-out all narrow content, and that her case does not provide a basis for rejecting narrow contents more generally, such as for general concepts. Having shown that the view developed in Chapters 1 and 2 is not threatened by social externalism or approaches to individuation in perceptual psychology, in the next two chapters I consider the explanatory role of narrow content.

Chapter 4

Internalism and Propositional Attitude Ascriptions

1. Introduction

In Chapters 1 and 3 I outlined how Putnam (1975) and Burge's (1979) arguments for externalism partly rely on assumptions about propositional attitude ascriptions. From these cases externalist conclusions were drawn by highlighting which beliefs (concepts) are ascribed to subjects in Twin Earth, and other, scenarios. In Chapter 4 I consider the nature of propositional attitude ascriptions in more detail with a view to showing what is wrong with basing a case for externalism on attitude ascriptions, and that narrow content plays an important semantic role. To this end, I show how the account of narrow content developed in Chapters 1 and 2 is consistent with a Fregean theory of attitude ascriptions according to which the truth conditions of some attitude ascriptions relate to the narrow contents ascribed. Such a view faces a number of objections and I consider two of the main objections raised against Fregean accounts, with particular attention to how an internalist may respond.

The chapter is structured as follows. In Section 2, I highlight problematic assumptions about propositional attitude ascriptions as construed by Putnam and Burge. I then outline an internalist alternative, which shows that acceptance of externalist conclusions from cases involving attitude ascriptions is not mandatory. In Section 3 I show that internalists may hold that narrow content plays a semantic role in ordinary attitude ascriptions in so far as Fregeanism about attitude ascriptions is true. In linking the role of narrow content to Fregean accounts, it is necessary to consider the debate between Fregeans and Millians. In defence of a Fregean account as a basis for narrow content, I consider two objections. First, I respond to Travis' (2008) objection to the role of opacity in Fregean attitude ascriptions. Second, in

Section 4, I consider Soames' (2002) externalist and minimalist account of attitude ascriptions, which, if correct, would show that Fregean accounts are wrong about the semantics of attitude ascriptions. Finally, in Section 5, I contrast my account with existing internalist accounts from Segal (2007) and Chalmers (2007), and explain how my approach provides an alternative way to respond to an objection raised by Soames' (2002) against the role of narrow content in attitude ascriptions.

2. Externalist arguments and the semantics of attitude ascriptions

The Twin Earth and arthritis cases Putnam (1975) and Burge (1979) use to drive their externalist conclusions rest partly on the role and nature of attitude ascriptions. If we have reason to resist the accounts of attitude ascriptions in these scenarios, then we can resist the conclusions and thereby provide further support to the internalist responses considered in Chapters 1 and 2. In this section I explain how the externalist accounts are partially motivated by assumptions about natural language meaning and the concepts expressed by terms in public language, and then draw on Chapters 1 and 2 in order to cast doubt on the assumptions. In addition, I will show that the externalist cases are consistent with ascribing attitude contents which relate to the account of concepts and narrow content developed in Chapters 1 and 2.

2.1 Externalist arguments and attitude ascriptions

Consider once more Putnam's Twin Earth scenario. We are to consider a subject on Earth, Oscar, and his intrinsic duplicate Twin Oscar on Twin Earth. At their respective worlds the watery substance is H_2O and XYZ. Attitude ascriptions are used as one way to contrast what intrinsic duplicates believe and how we talk about their beliefs. For instance, we are urged to accept that an utterance of (1) is true, while (2), which uses the same sentence to attribute a belief to Twin Oscar, is false:

(1)   Oscar believes that water is wet

(2)   Twin Oscar believes that water is wet

Read literally, (1) and (2) would appear to attribute the same belief to Oscar and Twin Oscar that the kind referred to with 'water', by the speaker, is wet. However,

by hypothesis we cannot use 'water' to express Twin Oscar's belief because 'water' in our mouths refers to H_2O. Thus, it is held that we need to use 'twin water' to report a true belief of Twin Oscar's. Note that Twin Earthers can use 'water' to truly report on Twin Oscar's beliefs about twin water, but can't use 'water' to say true things about Oscar's water beliefs, since in their language 'water' refers to twin water. Consequently, given the meaning of 'water' in English, if we want to report Twin Oscar's belief, we should say something like 'Twin Oscar believes that twin water is wet'.

Putnam's thesis was about the meaning of 'water' in natural language. Two important assumptions can be highlighted at this juncture. The first assumption is that when a speaker uses an attitude ascribing sentence like (1), terms like 'water' are natural kind terms that pick out the local natural kind properties that the speaker is acquainted with. The problem here is that Putnam has not provided independent support for this treatment of natural language semantics. As noted in Chapter 1, it is plausible that 'water' or WATER may represent a non-natural kind. Suppose then that 'water' needn't refer exclusively to H_2O. Where a concept like WATER is individuated by narrow content, what the attitude ascribing sentence may express relates in part to the relevant narrow content of the speaker's concepts. This allows that if the speaker's WATER concept represents a set of non-natural kind properties, then using 'water' in (1) would not be restricted to the notion of a natural kind term, since it picks out non-natural kind properties.[43]

The second assumption builds on this point about the concepts of both speaker and ascribee. Note that the position assumes that Oscar and Twin Oscar have beliefs about different properties (H_2O, XYZ), however, in Chapters 1 and 2 I made the case that such subjects may share concepts that represent the same properties, for instance superficial appearance properties. Two other possible concepts were identified, where Oscar might represent a heterogeneous kind that subsumes both H_2O and XYZ, and where Oscar represents the local natural kind by way of a description.

---

[43] I say more about this view below. For now, all that is needed is to resist the assumption that we can't use 'water' to say anything meaningful about Twin Oscar's beliefs.

Hence both assumptions can be resisted. Consequently, we can deny that intrinsic duplicates in Twin Earth scenarios must be ascribed beliefs (and concepts) about the specific kinds and properties at their actual worlds. Rejecting these two assumptions would allow for internalist readings of (1) and (2), as I will show below.

Before developing an internalist proposal in more detail it is worth highlighting the similar assumptions at work in Burge's (1979) case for semantic externalism as this likewise revolves around assumptions about attitude ascriptions. As noted in Chapter 3, Burge's (1979) argument for psychological externalism rests on ascribing propositional attitudes to a subject, Alf, who has incomplete mastery of the term 'arthritis'. The standard case involves Alf incorrectly believing that his arthritis has begun to afflict his thigh. Burge contends that it is part of our ordinary attitude ascribing practices to use (3) to reflect Alf's propositional attitude in this case:

(3)   Alf believes that he has arthritis in his thigh

Motivation for (3) is given by noting that Alf has many true beliefs involving the term 'arthritis', such as that his father had arthritis, that arthritis is a painful condition, and that he has arthritis in his wrists and ankles. An utterance of (3) is then taken to be appropriate since Alf apparently has sufficient competence with the term in order for (3) to accurately report what he believes. The case for externalism is established by comparing Alf in a counterfactual scenario in which 'arthritis' is used by the linguistic community to refer to a broader condition that may also afflict the thigh, which we can call 'tharthritis'. In this counterfactual situation, (3) would be a true ascription since 'arthritis' refers to the broader condition and Alf is correct that he has this condition in his thigh. Externalism follows since Alf remains internally identical in each situation, and the relevant difference in his propositional attitude stems from the meaning of the attitude ascribing sentences used to express his beliefs as determined in the local linguistic community. Consequently, social externalism about propositional attitudes is true, since we cannot express the content of Alf's thoughts without recourse to the specific linguistic facts in the subject's actual

linguistic community which directly affect our attitude ascribing practices, and, crucially, the contents of the propositional attitudes.

Various reinterpretations of Burge's thought experiment have been proposed (e.g. Crane 1991; Segal 2000), many of which focus on the idea that it is wrong to say that (3) adequately reflects what Alf believes, and so reject the view that Alf has a false belief concerning the concept of arthritis in the first scenario. Instead, it is held that in both cases Alf thinks of the condition in his thigh in exactly the same way, as something more like tharthritis, and so has the same true belief in each scenario reflected by (4):

(4)   Alf believes that he has tharthritis in his thigh

As noted in Section 2 of Chapter 3, Burge rejects the view that we can identify such a concept, by resisting the kind of view associated with prototypes where concepts are individuated in part by a subject's dispositions to classify kinds into certain categories. I showed that this is at odds with work on concepts in psychology and as such we have reason to resist Burge's social externalist view of mental individuation. In addition, Burge's position is similar to Putnam's approach where Putnam held that 'water' in natural language on Earth represents H_2O, whereas the term means something else in a counterfactual scenario, and differences in belief follow from differences in the semantics of the attitude ascribing sentence. Hence, the same points raised against Putnam apply here.

What seems right about Burge's view, as noted in Chapter 3, is that we don't seem to reinterpret a subject whenever they appear to have a non-standard concept. However, I showed that in the arthritis case there is sufficient reason to deny that Alf meets minimal conditions on possessing a concept that represents arthritis exclusively. Instead, Alf plausibly has a concept that represents the more heterogeneous kind tharthritis. Since English doesn't have the term 'tharthritis', a normal speaker of English will not have access to a relevant term, and may need to qualify the attitude ascription, since (4) would only be an option as a result of philosophers introducing

the term 'tharthritis' into the language. However, the lack of a relevant term should not mislead us into holding that since an ordinary speaker might utter (3), then we can generate the externalist conclusion. Like Putnam, Burge commits to the assumptions that the relevant attitude ascribing sentences have a fixed externalist semantics, and that the relevant subjects represent different properties. Resisting both these assumptions allows us to block the externalist conclusions. Before showing how attitude ascriptions are compatible with the account of narrow content developed in Chapters 1 and 2, there is a third assumption worth highlighting.

2.2 The naïve view of attitude ascriptions

What is crucial for the externalist conclusions in these cases is that attitude ascriptions are reliable indicators of the beliefs and concepts of the ascribees in the relevant scenarios. Thus, for the conclusions about differences in belief to follow from the nature of the attitude ascribing sentence, it must be the case that attitude ascriptions genuinely ascribe a subject's belief or concept. Such a position relies on what I will call the 'naïve view' of propositional attitude ascriptions:

(Naïve view)   True attitude ascribing sentences ascribe the specific propositional
             attitude of the ascribee

Bach (1997) has argued against something very close to the naïve view, which he calls the specification assumption. This is the idea that attitude ascriptions specify the precise content of a subject's belief or other propositional attitude. On such a view, the truth of an attitude ascription depends on whether the proposition expressed is identical to the content of the subject's relevant propositional attitude. The basis of the specification assumption is the relational view of propositional attitudes, according to which, for example, Oscar's belief expressed as 'water is wet' consists in Oscar's standing in the belief relation to the relevant proposition expressed by

'water is wet', for instance the proposition comprised of the properties *water* and *wetness*.[44]

The problem Bach identified for the standard account revolves around a variation of a puzzle about belief introduced by Kripke (1979) through the case of Peter and his beliefs about Paderewski. The worry is that an utterance of (5) may not specify the specific belief of Peter's:

(5)   Peter believes that Paderewski has musical talent

For instance, if Peter hears the name 'Paderewski' while listening to a performance on the radio, (5) might under-specify Peter's belief which we might want to represent as (6):

(6)   Peter believes that Paderewski, the pianist, has musical talent

Again, the same thing might occur if Peter believes something even closer to (7):

(7)   Peter believes that Paderewski, the pianist he heard on the radio, has musical talent

On the standard view, the truth of the ascription depends on whether the content expressed by the embedded sentence specifies the subject's attitude content. However, given the problem of under-specification of what subjects like Peter believe, Bach contends that the proposition expressed by the ascribing sentence need not specify the content exactly, but rather the subject's belief must be such that it depends on the truth of the expressed proposition. Thus, so long as (5), (6), and (7) express something the truth of which is required for Peter's belief to be true, then they each may adequately *describe* a true belief of Peter's, without fully specifying his belief. From Bach, we can conclude that ascriptions may not properly specify

---

[44] Note that a Fregean could hold that the relevant proposition is not composed of individuals and properties, but rather senses. I set this view aside for now, since the objections raised against the property view could be raised against this alternative.

content in many (if not all) cases since the specific way a subject believes a given content might not be easily expressible, or knowable to the speaker.

Given the rejection of the two externalist assumptions above, Bach's criticism of the standard view may not be strong enough. Where our interest is in what S represents, and how S represents things as being, then we need to be careful not to make externalist assumptions about natural language and about the concepts of the subject whose beliefs we are talking about. Bach held that attitude ascriptions don't strictly specify what is believed, but can nonetheless be true. Here, we may need to be more careful, since the ascription 'Oscar believes water is wet' may not express something which describes Oscar's belief in Bach's sense. Where the externalist contends that this ascription does accurately reflect Oscar's belief, Oscar is taken to have a belief relating to H_2O being wet. However, if Chapters 1 and 2 are accepted, then this may not accurately reflect what is believed, nor express something the truth of which is crucial for Oscar's belief. This is because if Oscar's WATER prototype represents watery stuff (a set of superficial watery properties common to H_2O, XYZ, and so on), then the standard wide content does not reflect the relevant content of Oscar's belief, as this wide content pertains only to H_2O.

The objection to raise against the naïve view is that in addition to the underlying belief of the ascribee potentially being more complex than the attitude ascribing sentence, we need to be careful not to assume that the terms we use to talk about belief are restricted to an externalist semantics. I noted above that an internalist may respond to Putnam's and Burge's views by drawing attention to the way that subjects' concepts may represent a set of non-natural kind properties. What I did not say is that we must always try to use language in such a way that we specify these properties precisely. The internalist can therefore accept Burge's point that we don't normally reinterpret, since this is consistent with rejecting the naïve view and with holding that a speaker may use the standard attitude ascribing sentence without using it to ascribe a wide content to the ascribee.

It's time to take stock. On the one hand, I have suggested that we should resist the assumptions that attitude ascribing sentences have an externalist semantics, that the subjects in the relevant cases must represent different properties, and the naïve view of attitude ascriptions. On the other hand, I have noted that Burge is correct in that we don't normally introduce new terminology and utter sentences like (4) to try to reinterpret precisely what a subject believes. This latter point may look like a problem for the internalist who wishes to allow that attitude ascriptions can relate to a subject's narrow contents and narrowly individuated concepts. In the next section I outline what an internalist alternative might look like which accepts that we shouldn't aim to try to produce sentences like (4).

2.3 An internalist alternative

An internalist alternative can be outlined which allows that uses of attitude ascribing sentences like 'Oscar believes that water is wet', may be used in such a way to express the ascribee's idiosyncratic concepts and narrow contents, without requiring the speaker to introduce novel terminology. In Chapter 2 I noted alternatives to the view that Oscar possesses a natural kind concept that refers exclusively to H_2O. Furthermore, I have noted that there is insufficient support for the view that 'water' in English functions exclusively as a natural kind term that rigidly refers to H_2O in all possible worlds. Consequently, sentences like (1) and (2) might be used in order to express beliefs which do not relate to natural kind concepts possessed by Oscar and Twin Oscar. Suppose then that we want to express the way in which Oscar and Twin Oscar think of their respective watery substances in the same way. Used in this way, (1) and (2) could be used to truly report on the beliefs shared by Oscar and Twin Oscar. Given the range of concepts identified in Chapter 2, a speaker who utters (1) may have several different things in mind as being what they intend to communicate about Oscar's belief. In order to make this content explicit, we could introduce sentences which use terms closer to what the speaker has in mind. Thus, if a speaker intends to reflect Oscar's prototype concept WATER, roughly equivalent to water's being a kind of watery substance (being a clear, colourless liquid, etc), (8) could reflect the proposition intended to be expressed:

(8)   Oscar believes that the clear, colourless liquid that falls as rain is wet

Alternatively, we could substitute 'watery stuff' for the description, which can then be used to say something true of Twin Oscar as well, giving (9) and (10):

(9)   Oscar believes that the watery stuff is wet
(10)   Twin Oscar believes that the watery stuff is wet

It's worth noting that where 'watery stuff' picks out the set of superficial watery properties, then this may include being wet. This would apparently render (9) and (10) trivially true, in effect that Oscar believes, for instance, that the wet stuff is wet. However, what is important here is the general position, since we can change the case to make it non-trivial. For example, in (4), Alf is said to believe that he has tharthritis in his thigh, where 'tharthritis' is again a short-hand for a more complex description of the properties he associates with 'arthritis'. Here, the report is not trivial, since having this condition in one's thigh is not part of the meaning of 'arthritis' as understood by Alf.

As noted above, the internalist does not have to rely on using novel terminology in order to express a proposition that relates to the shared narrow content of Oscar's and Twin Oscar's beliefs. (9) and (10) (like (4) above) are intended to reflect only what a speaker may intend to express, while the sentence they actually use may just be (1) and (2), involving the term 'water'. Given the rejection of the assumptions that were operative in Putnam's and Burge's accounts, this position is available to the internalist since we do not have to agree with the externalist that the attitude ascribing sentence must be read with an externalist semantics (such as taking 'water' to pick out H_2O). If 'water' can be used to refer to watery stuff, i.e. the set instantiating the watery properties, then (1) and (2) are consistent with a version of internalism.

I have considered the case where Oscar and Twin Oscar's concepts relate to superficial properties. However, in Chapters 1 and 2 I allowed that there is a kind of

narrow content that relates to natural kind properties and to subsuming heterogeneous kind properties. In the case of natural kind concepts, Oscar and Twin Oscar will represent different properties. The question remains as to whether an utterance of (1) and (2) can be used to express something that Oscar and Twin Oscar both believe.

One way to show this is to appeal to the notion of relativized propositions developed in Chapter 1. Relativized propositions are such that different thinkers can entertain the same proposition, but once we consider the proposition at the relevant context of assessment thinkers may represent different properties picked out by the proposition. For instance, suppose there is a relativized proposition associated with 'water is wet' comprised of place-holders for 'water' and 'wet' which we can represent as an empty or un-filled proposition. The relativized proposition is associated with a mode of presentation, determined by the subject's underlying conceptual structures associated with 'water' and 'wet'. The mode of presentation relates to how the subject represents things as being, as given by the cognitive content of their concepts. The modes of presentation function to pick out the relevant natural kinds at a relevant context that relate to how the subject represents things as being. Where Oscar and Twin Oscar both represent the natural kind property associated with 'water' in terms of superficial, predominantly appearance, properties, their modes of presentation will pick out whatever natural kinds have these properties at the relevant context. Thus, Oscar and Twin Oscar can believe the same relativized proposition expressed by 'water is wet', and in doing so represent different properties, where on Earth 'water' picks out the property of H_2O, and on Twin Earth the property of XYZ.

If this is right, then we should allow that what is expressed by an attitude ascription can depend on what the speaker intends, which can range over the kinds of water concepts identified in Chapter 2 (watery stuff as a functional-appearance kind, water as a natural kind or a heterogeneous kind, and a scientific description of water as composed of H_2O). When we consider the case where Oscar and Twin Oscar share a mental representation of watery stuff as a heterogeneous kind that subsumes different natural kind properties, the attitude ascription will relate to the same content, and as

in the case of superficial properties, there is no need to introduce relativized propositions.

I have outlined the basic framework of what an internalist account might look like. However, the literature on attitude ascriptions is rich, and there is very little discussion relating to wide and narrow contents. Thus, an externalist may contend that the proper treatment of attitude ascriptions is at odds with this internalist picture. In order to sharpen the debate, in Section 3 I will introduce the debate between Fregeans and Millians, and show how the internalist view comports with a version of Fregeanism. The case against the internalist account can therefore be put partly in terms of whether Fregeanism is true. To this end, after outlining the Fregean/Millian debate and showing how narrow content fits into the Fregean picture, I then consider an initial objection to Fregeanism, before considering a more detailed Millian alternative in Section 4 which threatens the Fregean internalist account developed in Section 3.

3. Frege's puzzle, opacity and transparency

In this section I show how narrow content naturally fits within a Fregean account of attitude ascriptions, and consider an objection to Fregeanism from Travis (2008) which would raise doubts about the plausibility of attitude ascriptions expressing narrow contents.

3.1 Frege's puzzle

The referential theory of names (hereby referred to as the 'Millian' theory) holds that the meaning or semantic content of a name is the individual the name refers to whilst the semantic content of a predicate is the property referred to. The semantic content of the sentence 'Superman flies' can then be understood as deriving from the contents of its parts which can be represented by a proposition which contains both the individual referred to and the property attributed to the referent.

A feature of a Millian semantic theory is that co-referring names should be substitutable whilst preserving the truth value of sentences in which they occur, so that (11) and (12) are both true:

(11)   Superman flies
(12)   Clark Kent flies

This follows because each sentence expresses the same proposition since 'Superman' and 'Clark Kent' refer to the same individual and the same property is attributed to that individual. Frege (2000) introduced the notion of sense partly as a response to the apparent failure of a Millian semantic theory to preserve truth value in belief contexts when co-referring names are substituted. The problem for the Millian view is that it appears that (13) can be true whilst (14) false:

(13)   Lois believes that Superman can fly
(14)   Lois believes that Clark Kent can fly

A further worry for a Millian semantic theory is that (15) would be literally false, while it appears to say something true about Lois:

(15)   Lois does not believe that Clark Kent can fly

The question of the proper treatment of such attitude ascriptions is the focus of a recent debate between Fregeans and Millians. Since Kripke's (1980) critique of Frege's notion of sense, the standard Fregean treatment of attitude ascriptions accepts the Millian view that the meaning of a name is its referent, and that this holds in attitude contexts, yet Fregeans contend that the truth conditions of attitude ascriptions depend in part on the senses, or modes of presentation, associated with the ascribing sentence. On the Fregean account, we preserve Frege's judgements of truth values in cases like (13) and (14) by claiming that modes of presentation are part of the semantic content of the attitude ascription which may affect the truth conditions of the ascription.

One way to understand the debate between Fregeans and Millians is in terms of whether attitude ascriptions are opaque or transparent. Attitude ascriptions are opaque when the truth conditions depend in part on the associated mode of presentation, and transparent when the truth conditions depend only on the Millian semantic content. Opacity alters the truth value of a sentence by virtue of the name expressing or implying a mode of presentation that affects the truth conditions of the attitude ascription. In such cases, the semantic contribution of the name includes the mode of presentation in addition to its referent. Some Millians accept opacity in certain cases, and there is some disagreement among Millians as to the implications of opacity. For instance, Braun (2006) denies that opacity ever affects truth value, while Soames (2002) argues that Millians should recognise two kinds of propositions in opaque contexts and so two kinds of truth conditions associated with transparent and opaque readings of an attitude ascribing sentence (these kinds of Millian responses will be discussed below). As a result of opacity, Fregean truth conditions (TCs) can be expressed as follows:

(Fregean TCs)   The utterance 'S believes that X is F' is true if and only if S stands in the relation of belief to the proposition {X, F} and thinks of X via mode of presentation M[45]

Similarly, Millian truth conditions for belief can be expressed as:

(Millian TCs)   The utterance 'S believes that X is F' is true if and only if S stands in the relation of belief to the proposition {X, F}

If modes of presentation are expressed in certain attitude ascriptions, then Fregeanism is true. However, as noted, some Millians either reject opacity completely, or deny that attitude ascriptions are ever entirely opaque. Before considering both of these objections, in the next section I will show how the Fregean

---

[45] Again, it's worth noting that some Fregeans take the constituents of propositions to be senses or modes of presentation, rather than individuals and properties. However, the view I develop relates to the latter view, and it is that kind of account that I will focus on throughout.

account provides the basis for a semantic role for narrow content in attitude ascriptions.

3.2 Fregeanism and narrow content

If attitude ascriptions express modes of presentation, then there is a possible role for narrow content in terms of playing a semantic role in attitude ascriptions and reflecting the subject's manner of representing the relevant referent. Consequently, the fate of narrow content in attitude ascriptions can be linked to the case for a Fregean theory of attitude ascriptions. Such an approach for providing a role for narrow content has been suggested by Recanati (1993) and more recently Chalmers (2007). I will return to alternative internalist accounts in Section 5, where I explain the advantages of the view I develop below. Before that, I will consider a general approach to Fregean attitude ascriptions.

One of the most prominent Fregean accounts has been developed by Richard (2006a) who argues that attitude ascriptions express linguistically enhanced propositions that are comprised of a Millian proposition and the uttered sentence itself. On this view of propositions, the content of a proposition is given by the referents of the uttered sentence, so the content of an utterance of 'X is F' is given by what those words refer to, which includes the individual X, the property F and the relation that F bears to X.

Consider an attitude ascription of the form 'S believes that X is F'. The Millian proposition provides the content of the belief being ascribed to the subject, S, whilst the sentence translates, or stands proxy for, the subject's mental representation which realizes the particular belief (or it represents a sentence the subject would endorse as expressing something they believe). The truth of the ascription depends on whether S has a mental representation suitably translated by the ascribing sentence. Thus, on this view, an attitude ascription relates to both the belief content and the subject's mental representation which realizes the belief. Richard does not construe modes of presentation in anything like the manner discussed in Chapter 2. However, a natural position for an internalist to take is that the attitude ascribing sentence relates to the Fregean notion of a mode of presentation, which relates to the cognitive content of

the subject's concepts. Where modes of presentation are expressed, the resulting mode of presentation reflects the way a subject represents things as being.

An internalist needn't accept Richard's position that a Millian proposition is always expressed by an attitude ascription which provides the content of the ascription. It's worth noting that Chalmers (2007) seems to defend an internalist version of Richard's approach and holds that a kind of wide content associated with Millian propositions is always expressed. However, it is possible for a semantic internalist to reject this view and contend that there are cases where only narrow content expresses the content of the ascribee's belief, even though the attitude ascribing sentence may also express a wide content. For instance, I have conceded that names have a kind of wide content. Thus, where an attitude ascription picks out an ascribee, the attitude ascribing sentence will have a wide content relating to the ascribee. However, the remainder of the utterance might only express a narrow content relating to the content of the relevant belief. To see this, note that an utterance of 'Oscar believes that water is wet', could express a wide content associated with 'Oscar', but 'water is wet' could relate entirely to the narrow content of Oscar's belief. Nothing that I say in what follows requires commitment to this stronger view, since in order for narrow content to play a semantic role in attitude ascriptions, it is sufficient to show that narrow content is semantically expressed. Hence, I will focus on the Fregean notion of the mode of presentation associated with the referent of a name. Thus, where Lois' concept SUPERMAN has a mode of presentation that differs from that associated with her concept CLARK KENT, the case for narrow content being expressed by the attitude ascription will rely on the attitude ascription being opaque, and hence for the mode of presentation to be relevant to the semantic content of the ascription.

3.3 Travis' challenge to opacity

Two objections can be raised against such a Fregean theory which would prevent a role for narrow content in the semantics of attitude ascriptions. The first objection I consider is based on concerns about opacity voiced by Travis (2008) who has rejected opacity on the basis of a lack of criteria to determine whether attitude

ascriptions are ever genuinely opaque. In Section 3 I consider this issue in more detail. I consider a second objection from Soames (2002) in Section 4.

3.3.1 Opacity and semantic intuitions

To recap, Fregeans contend that the truth of an opaque attitude ascription depends on whether the ascribing sentence expresses a mode of presentation that suitably reflects the subject's way of representing the referent. Where (5) above is taken by Fregeans to express something true, some Millians contend that it would be literally false, since Lois believes the referent of 'Clark Kent' can fly. A sentence like (16) can be introduced to attempt to lend further credibility to an opaque reading:

(16)   Lois does not believe that the reporter of her acquaintance, Clark Kent, can fly

If it is not obvious from the context that the speaker is referring to Lois' way of thinking of Clark Kent, we may emphasise this by using certain linguistic cues, such as adding additional descriptive information (further properties associated with the referent) which add to the plausibility of the opaque reading. For instance, instead of adding 'of her acquaintance' to (16), we could say 'the geeky reporter she knows as Clark Kent'. Such cues may suggest that the speaker intends to use the sentence opaquely so that the truth-value depends on Lois' mode of presentation. As noted, Millians suggest that (5) is literally false, and could also dig their heels in and say that (16) was also strictly speaking false. I will return to this sort of response to opacity below. What is important here is that Frege introduced modes of presentation on the basis of the apparent failure of a Millian semantics to account for intuitions that sentences like (5) and (16) would be true. The resulting view entails that certain attitude ascriptions are opaque, and that their truth conditions depend in part on the relevant mode of presentation expressed. It is therefore crucial for Fregeanism that the notion of opacity can be reasonably well articulated.

Where (16) aimed to support opacity by introducing further descriptive information, it is also possible to draw attention to other features of the context to support the idea of opaque attitude ascriptions. For instance, Recanati (1993, 358) considers cases

where the context is such that the hearer is aware that the speaker intends to reflect the subject's way of representing things (the mode of presentation the subject associates with the name). Suppose that John truly believes that Tully was wealthy but does not believe that Tully is Cicero. In uttering an ascription like (17), a speaker can express the ascribee's perspective in a context where speaker and hearer are both aware that Cicero was wealthy:

(17)   John believes that Cicero was poor

From the context, since speaker and hearer are aware that John has a true belief that Tully is wealthy, and does not believe that Tully and Cicero are the same individual, then uttering (17) could draw attention to John's representation or mode of presentation associated with 'Cicero'. A strict Millian or transparent reading of (17) would be odd in this context, since speaker and hearer both know that John associates distinct representations with 'Cicero' and 'Tully'. The strict reading would say something false of John, but given the context, it certainly appears to say something true, and would be intended by the speaker to say something true.

3.3.2 Travis' objection

In this way, opaque uses of attitude ascriptions may depend on features of the context, in this case the shared knowledge of speaker and hearer. However, whether these kinds of moves genuinely support opacity is an important issue. Travis (2008) has recently argued that the notion of opacity and the issue of determining when an ascription is opaque are both highly problematic. Travis is concerned with a view held by some Fregeans that opaque attitude ascriptions can't be construed as having a transparent reading. Some Fregeans appear to endorse such a view, such as certain readings of (4) where they contend that the transparent reading is not available. Where Fregeans allow both opaque and transparent readings, Travis' objection is not as forceful, although it may still put pressure on the Fregean view since if the standard sorts of attitude ascriptions are not univocally opaque, then modes of presentation will not affect the truth conditions of the ascriptions. Consequently, narrow content might not contribute to truth conditions, thus a response to Travis-

type concerns needs to be provided. What a Fregean can accept is that Travis is surely right that it is not always clear when an ascription is supposed to be opaque, and that the form of the attitude ascribing sentence provides no such basis for opacity. In this section, after considering Travis' worry in some detail, I will explain how a Fregean who allows both opaque and transparent readings can highlight when an opaque reading is appropriate, and explain on what basis we can identify an opaque reading as the correct reading of a given attitude ascription. Thus, if we can say why in certain cases a transparent reading is to some extent not relevant, then even if there is available a transparent reading, the Fregean will be able to say that it should not provide the semantic content of the attitude ascription.

Travis' central thesis is that we don't have a good handle on what opacity consists in and hence no basis for identifying when an ascription would be opaque. What motivates Travis' case against opacity is the general concern that "…issues of opacity are typically much cloudier than they are taken to be" (Travis 2008, 189). Travis asks us to consider a subject, Brangwyn, who knows of an individual under the name 'Grant' but does not know that the individual also goes by the name of 'Gustaf'. We are then to consider an attitude ascription like the following:

(18) Brangwyn believes that Grant was at Harrod's but she does not believe that Gustaf was at Harrod's

Travis identifies two related problems for (18) being opaque. First, (18) is not something speakers would normally say. The problem is that "it is abnormal to use two different names in such close proximity to name the same person" (Travis 2008, 194). Just why this should be the case is not clear. Note that Millians might do something similar where they hold that Brangwyn believes that Grant was at Harrod's and she believes that Gustaf was at Harrod's, which would be a similar abnormal use of two co-referring names in a sentence. The second point provides a possible explanation why this is a problem for opacity, where Travis thinks that a reasonable response to an utterance of (18) would pick up on the second part of the sentence and induce a hearer to ask a question such as "But I thought you just said

she did think he was there" (Travis 2008, 194). The situation imagined is where speaker and hearer both know that 'Grant' and 'Gustaf' refer to the same individual. The problem with Travis' appeal to intuitions about what ordinary speakers may or may not judge is that we can conceive of situations in which the form of (18) would not be implausible. For instance, we can change the example to (19):

(19) Lois believes that Superman can fly but does not believe that Clark Kent can fly

It seems perfectly plausible that (19) might be uttered by someone explaining the story of Superman to someone who hasn't heard it before. The use of both names in close proximity doesn't seem problematic in this case. Furthermore, the same point seems to apply with regard to real historical figures. Returning to the above example, John knows of a famous Roman orator via the name 'Cicero', but does not know that Cicero was also called 'Tully', and John falsely believes that Cicero was poor. (20) would again be unproblematic, and appears to induce opacity:

(20) John believes that Cicero was poor but does not believe that Tully was poor

If the emphasis is placed on '*Tully*', then this can be read in such a way that the speaker could be highlighting two errors John makes in thinking that Cicero was poor, and in not recognising that Cicero is identical to Tully. The truth conditions thus relate to the relevant mode of presentation. When we set out the relevant background, the opaque reading appears plausible, and we generate a semantic content enriched by the relevant mode of presentation.

A further worry with Travis' objection is that in rejecting opacity every attitude ascription must be transparent, and the resulting truth condition relates only to the Millian content. This produces the kinds of problems that motivated Frege's notion of sense in the first place. For instance, where John knows the names 'Tully' and 'Cicero' but does not know they refer to the same individual we get (21):

(21)   John does not believe that Cicero was wealthy

Intuitively, (21) is true, because John does not associate the name 'Cicero' with an individual who is poor.   However, since John does believe of Cicero that he is wealthy, albeit under the name 'Tully', then there is sense in which (21) is false. However, only permitting transparent readings of (21) would appear to require rejecting intuitions that support opacity.   This is a problem for Travis, since his rejection of opacity, as in (18), is motivated by intuitions about what attitude ascriptions would be used to express in various contexts.   Consequently, if intuitions are to play a role in deciding whether attitude ascriptions are ever opaque, then intuitions about (21) are just as relevant as intuitions about (18).

3.3.3 The basis of opacity in the speaker's intentions

While Travis' concerns about opacity can be allayed, a further variation of Recanati's (1993) example shows that even those who defend opacity might disagree on which sentences are opaque.   For instance, Recanati (1993, 333) contends that an utterance of (22) need not be opaque, since if it was uttered in a context where speaker and hearer are both aware that Cicero was wealthy, then using 'wealthy' need not serve as a cue that the speaker is referring to John's mode of presentation:

(22)   John believes that the wealthy Cicero was poor

Instead, Recanati contends that the sentence could be used to highlight the error in John's belief, where mentioning the 'wealthy Cicero' conflicts with John's belief that Cicero is poor.   However, it's not obvious that an utterance of (22) couldn't be used to express John's mode of presentation.   An utterance of (22) could be read opaquely in so far as the speaker is drawing attention to the conflict between the way Cicero is and the way John believes Cicero to be.   Suppose that the context is as before, so that both speaker and hearer are interested in John's perspective.   An utterance of (22) seems capable of being used opaquely.   One might worry that (22) cannot be opaque because the truth conditions do not require that the sentence expresses John's mode of presentation.   For instance, the sentence's truth depends on the falsity of

164

associating the property of being poor with the referent of 'Cicero'. Yet the sentence is not obviously transparent either. If read transparently, the sentence would appear to literally state that John attributed contradictory properties to the same individual, i.e. that Cicero is both wealthy and poor. This suggests that whether ascriptions are used opaquely or transparently is not as clear as often suggested, and so Travis' worry about opacity applies equally to transparency. However, an utterance of (22) can perhaps best be understood by contrast with an utterance of (21) in which the speaker aims to reflect the subject's perspective. We can understand (22) as reflecting the speaker's perspective on the world, with 'wealthy' reflecting the speaker's and hearer's knowledge of the facts of history. Yet (22) is not straight forwardly transparent, since if read in the literal Millian manner, then it expresses a contradictory belief, and this is not what a use of (22) would be intended to do.

One way around this problem is to adopt a more complex account of attitude ascriptions, which allows a single attitude ascription to be capable of being used to express a variety of perspectives, depending on how the ascription is used, as in (22). What this suggests is that whether an ascription is opaque or transparent depends in part on the intentions of the speaker and perhaps the shared knowledge between speaker and hearer. Thus, it seems that Travis (2008) is right that opacity is not very well understood, but wrong not to draw the same conclusion about transparency.

If opacity is linked to what the speaker intends to express then we can make sense of the view developed in Section 2 above. There, I held that an utterance of 'Oscar believes that water is wet', could relate to Oscar's narrow contents, without introducing novel terminology. We can now say that when a speaker intends to express Oscar's mode of presentation, then on the Fregean account the ascription is opaque, and the truth conditions relate to the relevant mode of presentation (and hence narrow content). Thus, extending the insights of opacity of names to the externalist scenarios, we can hold that attitude ascriptions that relate to kinds and properties may also be opaque, and that the truth conditions depend on the mode of presentation the subject associates with the relevant kinds/properties represented.

In linking opacity to the use of an attitude ascription, it is worth noting that Fregeans can allow that there are transparent uses of attitude ascriptions. For instance, an utterance of (3) can be transparent, for instance where Lois' mode of presentation is irrelevant to the truth of the ascription. Yet, it could also be opaque, if the context was such that how Lois is thinking of Superman was part of what the speaker intended to communicate. A transparent use of (3) is compatible with Fregeanism, and can be accounted for using the same components identified above. For instance, if a speaker intended to communicate only the point that Lois associates the property of flying with Superman, without any intention to reflect how Lois believes this, then the ascription can be used transparently. On transparent uses, the truth conditions are identical to those proposed by Millians, since the subject's modes of presentation are irrelevant to the truth or falsity of the utterance. However, although (4) and (16) could in some context be read transparently (in the context of a philosophy paper, for instance), it seems unlikely that anyone would use (4) or (16) transparently since in most contexts in which speaker and hearer are discussing what Lois believes, both will be aware that (4) and (16) would be highly misleading. Nonetheless, transparent uses are available, and with Recanati (1993) we can agree that an utterance of (17) could be used transparently to report that John associates the property of being poor with Cicero, without the speaker intending to communicate anything about the specific way in which John believes this, as they are interested primarily in reflecting the speaker's and hearer's perspective on the facts of history. However, if the stronger internalist view, outlined in Section 2 above, is defended, then transparent uses might not relate to Millian contents, since the terms occurring in the attitude ascription might all relate to narrow contents, with the exception of names. Again, I do not commit to this stronger view here.

There is one final objection to consider against Fregean semantics for attitude ascriptions. Where I have defended the Fregean view that opacity affects the semantic content of an opaque attitude ascription, some Millians deny that Fregean content is semantic content, even if it can be what a speaker intends to communicate. One way in which Millians have defended this view is by holding that the literal semantic content of the sentence type in natural language is fixed. Thus, even if

there is an opaque reading, there is always a standard transparent reading that provides the semantic content of the ascription. This would be true even where the speaker primarily intended to talk about the subject's mode of presentation, hence where an opaque use was intended rather than a transparent use. This would provide the Millian with a basis for defending the transparent reading without denying that there are opaque readings. Thus, where we have a response to Travis that opacity is not as difficult to induce as he contends, a Millian can still argue that there is always a transparent reading, and that this provides the semantic content of the utterance. This is a natural basis for Millianism, since if we can read-off the semantics from the literal meaning of the terms used, then transparency could be the default position by virtue of every utterance of a relevant attitude ascribing sentence having a default Millian semantic content. Soames (2002) has defended such a view of attitude ascriptions, and I consider this view in Section 4.

4. Millianism and semantic minimalism

In Section 3 I showed how opacity can be defended by drawing attention to uses of attitude ascriptions where the speaker explicitly intends to express the ascribee's mode of presentation, and where this semantically enriches the relevant truth conditions. Another challenge to Fregean semantics is the proposal that semantic content is always Millian, as determined by the literal meaning of the attitude ascribing sentence. The problem for Fregeanism is that a Millian can accept that modes of presentation are indeed intended to be expressed, but that the semantic content is determined entirely by the Millian content of the attitude ascription. The threat this poses to Fregeanism is that if the Millian content is the semantic content of the utterance, then modes of presentation are not semantically relevant, even if the speaker intends to express the mode of presentation. Consequently, if modes of presentation do not affect the semantics of attitude ascriptions, then neither will narrow content. In this section I consider Soames' (2002) Millian account of attitude ascriptions which defends a version of this view. I show that Soames' position can be understood as being based on semantic minimalism, and I outline three alternatives to this view. Although none of these responses provide an argument against minimalism, the point is that the rejection of Fregean semantics by Soames

can be resisted by highlighting relevant semantic foundations that contrast with Soames' preferred semantic framework.

4.1 Semantic minimalism and Millian attitude ascriptions

As noted, some Millians concede to Fregean intuitions that an utterance of (4) might imply something false, such as that Lois associates the name 'Clark Kent' with the property of flying. Different explanations have been provided in the literature. One such view is developed by Soames (2002) who accepts that certain uses of attitude ascriptions may involve a Fregean mode of presentation, but he denies that this is ever semantically expressed. Instead, at most there are uses of attitude ascriptions that are pragmatically enriched with a mode of presentation, but the semantic content is always given by the Millian content of the attitude ascribing sentence.

Soames (2002) agrees with Fregeans that we should allow that an utterance which aims to reflect the subject's mode of presentation can depend for its truth on whether the mode of presentation reflects a way the subject takes things to be, and that this may be the more appropriate truth condition of the utterance in a given context. However, Soames contends that given the nature of semantics, the literal Millian semantic content of sentences like (4) is always expressed, whether or not an utterance of (4) also pragmatically implies something false and which the speaker primarily intended to communicate. Soames' strategy is to make a distinction between the proposition semantically expressed by an ascription and the propositions that may be pragmatically conveyed/imparted by an utterance of the sentence in a given context. Thus, on Soames' view we must recognise two propositions: the semantically expressed proposition tied to the literal meaning of the sentence (the Millian content), and the asserted proposition which may be descriptively enriched with a relevant mode of presentation (the Fregean content). The semantically expressed content of the sentence is always the same irrespective of the context or speaker, while the asserted content may vary from context to context.

Soames' view can be illustrated with an example similar to (4) which he takes to show that Fregeans get the semantics of attitude ascriptions wrong. Suppose that

Tom, Dick, and Harry know an elderly gentleman they call 'Peter Hempel' and are aware of the philosopher called 'Carl Hempel', yet are unaware that Peter Hempel the elderly gentleman of their acquaintance is Carl Hempel the philosopher. Harry forms the belief that Carl Hempel died last week, and Tom may report this belief to Dick by uttering (23):

(23)   Harry believes that Carl Hempel died last week

Soames (2002, 212-215) contends that an utterance of (23) may be intended to communicate a proposition that is descriptively enriched with information common to Tom, Dick, and Harry, for instance (24):

(24)   Harry believes that the philosopher Carl Hempel died last week

This descriptively enriched proposition may more closely reflect what Harry believes, capturing the relevant mode of presentation with the term 'philosopher', and so would be true. Thus, (23) might be uttered in order to communicate a descriptively enriched proposition like (24), which may be closer to the way in which Harry represents things as being. It is worth noting the similarity with the internalist view outlined in Section 2, where the descriptively enriched proposition expressed by 'Oscar believes the watery stuff is wet', was taken to more closely reflect what a speaker may intend to express with 'Oscar believes water is wet'. In this way Soames aims to incorporate Fregean views about what a speaker may intend to express when using an attitude ascription, including the descriptive information which reflects the subject's mode of presentation of the referent. The trouble for Fregeans arises when we consider (25):

(25)   Harry believes that Peter Hempel died last week

According to Soames, Tom would utter (25) only if he wanted to convey something like the descriptively enriched proposition (26):

(26)　Harry believes that the elderly gentleman of our acquaintance, Peter Hempel, died last week

Supposing that Harry does not believe (26), Soames allows that an utterance of (26) would assert something false. However, since (23) and (25) express the same proposition and have the same semantic content, they each express something true since the one is true only if the other is true. Consequently, Soames holds that substituting co-referring names does not alter the proposition semantically expressed or the semantically determined truth conditions. Thus, strictly speaking, utterances of (25) and (26) would be literally true since Harry believes *of* Peter Hempel that he died last week. This may not be what Tom primarily intends to communicate, but given the Millian semantics operative in these cases, these sentences nonetheless express true propositions.

4.2 Objections to semantic minimalism

In this section I will raise three lines of objection that put pressure on this view, and ultimately provide a foundation for a Fregean semantics that rejects Soames' claim that Millian content is always semantically expressed. Soames' central claim is that Fregean views get the semantics of attitude ascriptions wrong. This is because in treating cases like the Hempel example as having only one truth condition tied to the opaque reading, they fail to properly respect the sense in which sentences have a fixed semantic content determined by relevant features of a Millian semantic framework. On this view, across all contexts a sentence makes the same semantic contribution. This position is the essence of semantic minimalism, the view that there is a fixed minimal semantic content for each expression of natural language. Soames (2002) contends that much of natural language can be understood in terms of a directly referential theory, in the spirit of Kripke's (1980) theory of names. On this view the semantic content of a name like 'Aristotle' is its referent. The same point applies to kind terms like 'water', where the semantic content is the referent of 'water', which according to Soames is determined along externalist lines. Thus, whether Soames' critique of Fregean semantics succeeds depends on whether his semantic framework is correct. If the semantics of natural language is not directly

referential in this way, or not fundamentally externalist, then Soames' critique of Fregeanism will fail. Consequently, modes of presentation, and hence narrow contents, may be semantically expressed.

The following three rival semantic frameworks take a different stance on what the semantic content of an utterance is. First, semantic contextualism (e.g. Recanati 2010) is a direct competitor to semantic minimalism, and there is a rich debate between minimalists and contextualists. Second, conceptual semantics (e.g. Jackendoff 2002) contends that the semantic content of expressions of natural language relate to the speaker's conceptualization of the world, and thus the semantic content is not a variety of representational content, relating to reference. Third, the account of concepts and narrow content developed in Chapters 1 and 2, together with the claim in Section 2 above that a speaker may intend to express narrow content with a given attitude ascribing sentence, provides the basis for the view that the speaker's semantic content is fully determined by the narrow content of her concepts. I will consider each position in turn.

While a full treatment of the debate between semantic contextualists and minimalists is beyond the scope of this discussion, it is important to highlight because Soames' case against Fregeanism turns on his semantic framework. Without this framework, the examples above fail to motivate the view that we must recognise two propositions. According to semantic contextualism, pragmatic processes contribute to the semantic content of certain expressions. Consider the sentence 'it's raining'. According to Recanati (2002) the proposition expressed by an utterance of this sentence contains unarticulated constituents, relating to the fact that it must be raining at a certain place and time. Since rain must occur at a place and time, then the relevant constituents of the proposition are unarticulated, although tacitly expressed by an utterance of 'it's raining'. Proponents of semantic minimalism such as Cappelen & Lepore (2005) reject this view. However, my point here is not to address this debate directly. Instead, I wish to point out how the contextualist approach may provide a basis for defending the Fregean framework.

In support of the view that (25) could semantically express (26), a contextualist might appeal to something similar to unarticulated constituents. For instance, where Tom may implicitly intend to communicate the content of (26) when uttering (25) which lacks explicit reference to this additional content, the descriptive information is an unarticulated constituent of his utterance of (25). The semantic content of (25) would be given by (26) since the additional semantic content is not articulated by Tom. If this is right, then the contention that (23) and (25) semantically express the same proposition may be false, since if there are unarticulated constituents that contribute to the proposition semantically expressed, then (23) and (25) are uttered in order to semantically express the pragmatically enriched propositions (24) and (26). Rather than seek to adjudicate this debate, it is enough to point out that there are a range of views on these issues, which turn on the minimalism/contextualism debate. The point to note here is that we needn't accept Soames' pragmatic/semantic proposition distinction, since the case against Fregeanism presupposes the distinction. If we restrict the discussion to the data relating to attitude ascriptions and Frege's puzzle we are not forced to adopt the Millian view which turns on issues outside the scope of this chapter. Consequently, Soames' claim that Fregeans get the semantics wrong rests on a debate outside of the discussion relating to attitude ascriptions, and Fregeans may simply reject the relevant semantic framework that supports Soames' position.

Two specifically semantically internalist views are also worth considering. Conceptual semantics (Jackendoff 2002) is a fully internalist semantic theory, which construes semantic content in a non-representational manner. On this view, both pragmatic and semantic content derive from information stored in long-term memory such that the semantic content of an utterance is determined entirely by features of the speaker's conceptual system. What is important on this view is that there is no distinction between semantic content and pragmatic content. According to Jackendoff (2002), pragmatic enrichment contributes to semantic content. Hence, if this view is right, then there would be no basis for Soames' distinction between semantically expressing and pragmatically asserting different propositions, and instead we could say that the content expressed reflected the semantic content the

speaker intended to express, as determined by the speaker's conceptual system. For instance, the semantic content of (14) 'Lois believes that Clark Kent can fly' would be fully determined by the speaker's concepts. Consequently, were the speaker to intend to express Lois' mode of presentation, then the semantic content would be fixed by the speaker's concepts, giving rise to the Fregeans' semantic claim that there is just one truth condition in this case. Such a view would clearly block Soames' dual proposition approach, and is a direct competitor to Soames' semantic minimalism. This response is stronger than the contextualist's view, drawing on a radically internalist approach to semantics that relates semantic content to an individual's concepts understood as non-representational. The main point to note, again, is that there is a way to resist semantic minimalism and the semantics / pragmatics distinction as Soames considers it to be.

Finally, a second, less radical, internalist account can be outlined by combining the account of concepts from Chapters 1 and 2, together with the claim in Section 3 that opacity relates to what the speaker intends to express with a given attitude ascribing sentence. Such a view does not need to reject externalism, since it is compatible with allowing wide content for certain expressions, as well as for general terms, but with prioritizing narrow content in some way. Like conceptual semantics, this approach holds that an utterance of (14) may have a kind of semantic content determined entirely by the speaker's concepts (in particular by the relevant narrow content of those concepts). However, in Chapters 1 and 2 narrow content was taken to be genuinely representational, hence the relevant narrow content of the utterance may relate to real individuals and kinds in the world. I noted in Section 2 that in this chapter I am only concerned with defending the weaker version of internalism which concedes that attitude ascriptions may express wide contents, such as those relating to the term 'Lois'. This is consistent with allowing that what the speaker intends to express provides the semantic content of the utterance. For instance, the proposition expressed may include a combination of wide and narrow contents. The point to note is that without presupposing minimalism, we are not forced to hold that an utterance of (14) must have Millian semantic content, since if the semantic content can be determined by what the speaker intends to express, as determined by the

contents of the concepts she deploys, then the Fregean account can withstand Soames' criticism.

Given the three responses provided, we can put pressure on the general idea that what an attitude ascription semantically expresses is tied to the Millian content rather than the content associated with what the speaker intends to express. If any of the views outlined above are correct, then rather than two propositions associated with semantic and pragmatic contents, there might be a single semantic content relating to a descriptively enriched proposition that reflects the intuitive truth conditions of the ascription and is what the speaker intends to communicate in that context.

What I have not shown, however, is that any of these views are correct accounts of the semantics of natural language. This requires much more space than is available here, and may turn on empirical issues in linguistics and the philosophical study of the semantics of natural languages. However, if we reject Soames' semantic theory, then a central feature of Soames' case for rejecting Fregeanism is thereby blocked. This is a relevant dialectical point to highlight, since Soames' examples only get their force when combined together with his semantic theory, so we can't take his examples as theory-neutral data to be used against the Fregean view. Only by interpreting his example via his semantic theory do we get anti-Fregean implications. Hence, the reason I note that Fregeans may wish to reject Soames' semantic framework is that without that as a background for interpreting his examples, then his conclusions do not follow, since we only get semantic and pragmatic contents associated with attitude ascribing sentences with that picture in place. Thus, the interpretation of these cases is only really a worry if we have independent reason to think that the semantic picture is correct, and this is something that Fregeans, or opponents to Soames' semantic theory, may wish to criticise. The alternative accounts of semantic content I outlined in this section are just a sample of the accounts which would not have the anti-Fregean implications for attitude ascriptions which Soames' examples purport to have. What I hope to have shown is that a Fregean account can be linked to any of these views, and, importantly, that there are several ways in which Soames' rejection of Fregean semantics can be resisted.

5. Internalism and the Perfect Earth objection

In this final section I outline some of the advantages of the view developed in this chapter over existing internalist accounts in the literature. I then show how the view provides a response to Soames' (2002) Perfect Earth argument against two dimensional accounts of narrow content.

5.1 Comparison with other internalist accounts

Two of the main internalist accounts of attitude ascriptions in the literature come from Segal (2007) and Chalmers (2011), who, as noted in Chapter 1, provide very different accounts of narrow content and suitably different accounts of attitude ascriptions. To recap the view developed so far, if the Fregean view is correct, and attitude ascriptions can express modes of presentation, then we can motivate narrow content by noting that the modes of presentation expressed are equivalent to narrow content (as argued in Chapter 2). This is consistent with allowing that the content of transparent uses relates to wide content.

Segal (2007) develops a Fregean account of attitude ascriptions that draws on Davidson's (1968) account of how we report on what another has said. On Segal's (2007) view, what the speaker expresses is related to his own way of putting things. In particular, the speaker aims at relating the subject to what the speaker expresses by the belief report, and according to Segal the truth of the ascription depends on whether the content expressed by the speaker is sufficiently similar to a content of the ascribee. Thus, take the sentence 'S believes that x', uttered by T. On Segal's account, it has the form: that x. S believes that. The first sentence states what T has said, and the second states that S believes what T has said.

While this approach is consistent with the account of narrow content developed in Chapter 2, the advantage of the Fregean view is that there is an established framework, and pitching narrow content to opacity requires less work than showing that attitude ascriptions have the peculiar form attributed to them by Segal. Chalmers (2007) avoids the worry with Segal's non-standard approach to attitude ascriptions and defends an overtly Fregean view. On Chalmers' view, attitude

ascriptions express both Fregean and Russellian contents associated with narrow and wide contents. Chalmers' account has much in common with Richard's (2006a) which was that attitude ascriptions express linguistically enhanced propositions. The difference is that where for Richard the uttered sentence that translates the ascribee's mode of presentation is part of what is specified, Chalmers takes this to be narrow content as understood in terms of epistemic space – a set of epistemically possible scenarios that reflect the subject's idealized a priori knowledge. Furthermore, Chalmers rejects the notion of translation (or standing proxy for the ascribee's mode of presentation), and uses the notion of coordination instead, such that an attitude ascription is true if the sentence uttered expresses a narrow content that is coordinate with the ascribee's narrow content. Coordination is to be understood in terms of a similarity between epistemic spaces, so that the epistemic space of the speaker bears sufficient similarity to the ascribee's.

There is a superficial similarity with my view. I accept that both wide and narrow contents may be expressed by attitude ascriptions, however, where Chalmers appeals to coordination between narrow contents, my account makes more of the differences in the mental representations of speaker and hearer which the ascribing sentence aims to pick out. In particular, I note that 'water' might be associated with a range of mental representations that relate to different kinds of narrow content. So, where Chalmers tends to make the mode of presentation expressed relate to epistemic space, my view draws on work on concepts to allow that the truth conditions may relate to different conceptual structures, and the different narrow contents these are associated with. My view is not dependent on the success or failure of the epistemic framework. My appeal to a variety of narrow contents provides a further advantage of the view, as I will show in Section 5.2.

5.2 Soames' Perfect Earth objection to two dimensionalism

The account developed in this chapter provides an alternative response to Soames' (2002) Perfect Earth objection that is raised against two dimensional accounts of attitude ascriptions. In particular, I show how Soames might respond to Kallestrup's

(2011a) defence of the two-dimensionalist view and how my view provides a different way of avoiding Soames' objection.

Soames (2002) develops an example intended to show that certain semantic internalist accounts are unable to account for attitude ascriptions in relation to the subject *Perfect Mary*.   While Soames directs his critique at two dimensional semantics, it also applies to any descriptive account of narrow content such as that outlined in Chapter 1 where narrow content relates to the representation of natural kind properties at a subject's actual environment.   Take the descriptive account of natural kind concepts where 'water' represents a natural kind property at the subject's actual environment.   On this account, we can use 'Oscar believes that water is wet' to talk about Oscar's belief about H_2O.   According to Jackson's (2003a) version of the view, 'water' is substituted with 'the actual watery stuff'.   Soames asks us to consider the case of Perfect Mary on Perfect Earth.   The objection Soames raises is that on the descriptive view a subject on Earth is unable to account for Perfect Mary's belief, since if 'water' represents the actual watery stuff, then an Earth speaker will effectively claim that Perfect Mary's beliefs are about the actual world, and not Perfect Earth, since 'actual', when uttered by us, represents the actual watery stuff on Earth.   Thus, a speaker of English will utter (27):

(27)   Perfect Mary believes that water is wet

However, on the descriptivist view, this is understood as:

(28)   Perfect Mary believes that the actual watery stuff is wet

The problem is that on the descriptive view we can never truly report what Perfect Mary believes, that is, in the sense of relating Perfect Mary's belief to the watery stuff on Perfect Earth. Kallestrup (2011a) responds to Soames' argument by noting that 'actual' functions as an indexical on the descriptive view. Thus, when we use 'water' in an attitude ascription and this functions to pick out the actual watery stuff, the implicit term 'actual' functions to indexically pick out the watery stuff at the

ascribee's actual world. Thus, the speaker uses 'water' to tacitly represent the watery stuff at the ascribee's actual world, and hence 'actual' is sensitive to the context of the ascribee, and needn't relate to the speaker's actual environment.

As is stands this looks like an effective response to Soames. However, Soames may respond in the following way. Recall Soames' semantic minimalism, where there is a fixed semantics for natural language which determines the semantic content of each expression of natural language. On this view, the semantic content of 'actual' could be fixed in English. Thus, when a speaker of English utters 'actual', or implicitly asserts 'actual', the content always relates to Earth, the speaker's actual world. What can be remarked here, however, is that we can object to Soames' minimalism on the basis of the three lines of objection noted in Section 4.2 above relating to different semantic frameworks. Such views may allow that the semantic content of 'actual' is not fixed in the way required by Soames. Thus, while Soames could make this response to the descriptive view, there is ample scope for a reply which rejects Soames' semantic framework and assumptions about the semantics of 'actual' in English. However, I will not consider here the proper treatment of 'actual' in English. What is important here is that Soames has a response to the indexical account of 'actual'.

The approach to narrow content developed in this thesis provides an alternative response to Soames' scenario. Recall from Chapter 1 that there is reason to hold that narrow content may come in more than one form. In particular, in addition to descriptive representation of local natural kind properties, it was suggested that we should allow for representation of non-natural kind properties. Thus, where a speaker utters (27), this needn't be taken to relate to (28) that contains a description, but instead relates 'watery stuff' to a heterogeneous kind that takes as its extension the heterogeneous property common to H_2O, XYZ, and so on, all of which satisfy the relevant description. Alternatively, on the relational view, the heterogeneous kind may be represented by considering what did or would trigger Perfect Mary's WATER concept.

The proponent of this view should not be concerned by the claim that this entails that Perfect Mary has specific beliefs about the natural kind on Earth and other possible worlds unconnected with Perfect Earth. As noted previously, in relating 'water' to the actual or possible triggers of S's WATER, the view does not entail that S's concept takes as its content a massive disjunction of different natural kinds. The property view of representation relates concepts to the properties they represent. On this view, since the concept WATER fails to represent a unique natural kind, it takes as a property a heterogeneous kind.

As the cases above show, it is also possible to construe 'water' as being used to pick out a set of superficial properties via relations to the relevant properties. In this case, where we talk of Perfect Mary's belief, we needn't accept with Soames that the speaker must relate Perfect Mary to a local natural kind property. Instead, as for Oscar, the speaker may relate Perfect Mary to a set of superficial properties instantiated on Earth, Twin Earth, Perfect Earth, and so on. In addition, in Chapter 1 and 2 I suggested that the descriptive view is not particularly well supported as a general account of the narrow content of non-natural kind concepts, although it is consistent with concept pluralism. Therefore, if we pay attention to the variety of concepts through which subjects like Perfect Mary represent the world, this opens up the possibility that there are a range of possible truth conditions for attitude ascriptions expressed by the same sentence token (27). Thus, the ascription should not automatically be taken to relate to a belief regarding a unique natural kind denoted by 'water', since, as I have been urging, the ascribing sentence might not neatly match the concepts through which the subject has the relevant belief.

6. Conclusion

Chapter 4 had two main aims, one negative and one positive. As externalism is sometimes based in part on claims about propositional attitude ascriptions, the first part of the chapter aimed to highlight key assumptions that can be rejected, in particular, that the terms we use to express propositional attitudes like belief only express wide contents. In support of a positive role for narrow content in attitude ascriptions, I outlined a Fregean account and then responded to two objections to

Fregeanism from Travis and Soames. Finally, I showed how the Fregean account, together with the response to Soames' minimalism, provides a way to respond to Soames' case against narrow content's role in propositional attitude ascriptions in the Perfect Earth case. Consequently, narrow content should be understood as having a semantic role in propositional attitude ascriptions by providing a range of truth conditions that are sensitive to the narrow contents of subjects.

Chapter 5

Internalism and Psychological Explanations of Behaviour

1. Introduction

In Chapter 4 I developed and defended the role of narrow content in propositional attitude ascriptions. Narrow content has also been taken to play a role in psychological explanations of behaviour. However, there is no standard account of what this role consists in, furthermore, existing accounts face serious objections raised by externalists. The most common approach is to relate narrow contents to proximal causes which are taken to be essential for any account of behaviour which appeals to psychological states like belief. However, two of the most prominent accounts from Fodor (1991) and Jackson & Pettit (1988) face objections that externalists can raise in response. The most important objection contends that proximal causes, and so narrow contents, are causally irrelevant to psychological explanations, since wide belief states (or wide contents) are sufficient. I contend that the best response to this objection is to find an alternative explanatory role for narrow content. To this end, I develop and defend a connection between narrow content and the rational role of reasons in providing psychological explanations of behaviour, and I show how this provides a suitable response to the most pressing objection facing internalist explanations of behaviour.

The structure of the chapter is as follows. In Section 2 I outline the nature of psychological explanations of behaviour and the issues at stake in the internalism/externalism debate. In particular, I explain the challenge faced by internalists who aim to defend a role for narrow states or narrow contents while accepting wide intentional explanations of behaviour as genuinely relating to different behaviours in Twin Earth-type scenarios. In Section 3 I consider the two

internalist proposals identified above. I show that there are important objections which cause serious problems for the views, specifically in relation to the reliance on proximal causes and narrow states (Burge 1989; 1995), and objections concerning explanatory generality and causal relevance (Yablo 2003). In Section 4 I develop an alternative basis for an explanatory role for narrow content by noting that a satisfactory account of psychological explanation must pay attention to the roles of reasons and rationality. I show how linking the role of narrow content to the explanatory role of reasons, as understood by Davidson (1963) and McDowell (2006), provides a way to respond to the objections from generality and causal relevance. Finally, I show how this account provides a response to Fodor's (1994) externalist approach to Frege cases according to which a kind of wide content is explanatorily sufficient for explaining behaviour.

2. Psychological explanations of behaviour

In this section I outline three approaches to psychological explanations of behaviour and highlight the central problem for internalism. In particular, I introduce the basic framework of belief-desire psychology and accounting for behaviour in terms of mental causation, Davidson's appeal to reasons as rational and causal explanations of action, and the role of modes of presentation in explaining behaviour in Frege cases. While these approaches have tended to be discussed separately, the issues are inter-related, and this will be important in the view I develop in Section 4. Also in this section, I explain the central features of the debate between externalists and internalists regarding the causal relevance of wide and narrow states. I then show that the central problem for internalism is that the descriptions we use to explain behaviour are glossed in terms of wide belief states. As such, the challenge facing the internalist is to provide a central explanatory role to narrow states or narrow contents.

2.1 Belief-desire psychology, reasons for action and Frege cases

When a subject S goes to the fridge to get a cola, we can understand such behaviour in terms of S's desire to drink cola and her belief that there is cola in the fridge. Such explanations permeate our talk about why subjects behave in certain ways.

That is, we typically appeal to beliefs and desires to explain or rationalize why S did what she did. According to Fodor (1987), we use belief-desire psychology to make useful generalizations about behaviour, such that if S believes that there is cola in the fridge and desires a cola, then all things being equal, S will get a cola from the fridge. Underlying this view of commonsense psychological explanations of behaviour is the idea of mental causation where mental states like belief provide causal explanations of behaviour. S's mental states, her belief-desire pair, explain why S does A (e.g. gets a cola from the fridge) by virtue of the causation of behaviour by mental states. Thus, belief-desire pairs are taken to be causes of behaviour, such that when we provide a description of why S did A, we are citing the cause of A. On this view, causal explanations of behaviour track causation of behaviour. However, some philosophers think that there are varieties of causal explanation that do not relate to the causation of behaviour, but are still causal explanations of behaviour. A version of this view is considered below.

Following Davidson (1963), we can understand S's belief-desire pair as providing S's reason for performing the relevant action[46], *A*. As with commonsense belief-desire psychology, on Davidson's view, reasons serve as causes of actions, thus reason-giving explanations are fundamentally causal explanations of behaviour. What Davidson's approach adds to the mental causation debate is that in giving S's reasons for doing A, we must be able to pick out something about the reason that provides insight as to why S did A. For Davidson, a key point is that:

> A reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action. (Davidson 1963, 23)

Thus, in terms of S's belief-desire pair, the description we use should tell us how S conceived of her own action. Davidson provides two further conditions on a reason, R, providing a rationalization of S's action:

---

[46] Here I will use 'action' and 'behaviour' interchangeably, understanding particular actions by S as aspects of S's behaviour. While some philosophers clearly distinguish between actions and behaviour, the differences will not be relevant to my discussion. Any objections to the terminology could be responded to by re-casting the Davidsonian approach in terms of behaviour.

> R is a primary reason why an agent performed the action A under the description [D] only if R consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that A, under the description [D], has that property. (Davidson 1963, 25)

The notion of having a pro attitude is to be favourably disposed to actions of that type. For instance, in the vocabulary of belief-desire psychology we might say that in addition to S's belief that there is water, S also has the desire to drink some water. S therefore has a pro attitude to actions involving S drinking water. Davidson is saying that R rationally explains S's action only if R relates to S having a favourable attitude to an action with a certain property captured by the description D, and where S has a belief that the action has the relevant property captured by description D. Davidson contends that we don't need to specify the primary reason itself, since to do so would be 'fatuous', as it is implied by the descriptions we tend to use. However, the description's causally explaining A does depend on its picking out in some sense S's primary reason.

One way of understanding Davidson's point about what the agent saw in her action is the notion of a mode of presentation. The notion of a mode of presentation of a referent was discussed in Chapters 1 and 2 where they were associated with concepts and their cognitive content. Cognitive content reflects how S represents things as being from her perspective. A natural understanding of Davidson's point is therefore that the reason we provide for S doing A must tell us in some sense how S thinks of things as being, and we may use the notion of a mode of presentation or cognitive content for this purpose. The explanatory role of modes of presentation in behaviour is central to the debate revolving around Frege cases. Consider a case from Segal (2009b), where Oedipus wants to go to Hesperus, believes that a flight to Phosphorous is about to leave, but does not believe that Hesperus is Phosphorous. It is intuitively plausible that given Oedipus' beliefs and desire, then Oedipus would not board the flight, since he does not believe that Hesperus is Phosphorous. According to Fregeans, the mode of presentation, or cognitive content of S's concept, is crucial to the explanation of why S does A, in part because it is the cognitive content that reflects how S thinks of things as being. For it to follow that S would

not board the flight to Phosphorous, the Fregean appears to assume that cognitive content plays an explanatory role in behaviour, such that S does A because of the cognitive content of her relevant beliefs/concepts. Some philosophers (e.g. Braun 1995; Fodor 1994) deny that we need to appeal to modes of presentation, and contend that wide content is explanatorily sufficient for psychological generalizations. In response, internalists may argue that wide content explanations fail to explain the relevant Frege cases. I return to these cases in Section 4.4.

The three approaches above all relate to the psychological explanation of behaviour, yet they focus on different issues. The debate between externalists and internalists has tended to focus on the first and third set of issues relating to mental causation and Frege cases, although these debates are kept apart. However, the issues discussed so far are not mutually exclusive, and I will draw on aspects of each of these three approaches in the following discussion.

2.2 Wide behaviour and the problem for internalism

As noted, a subject's behaviour is explained by providing a description that cites one of S's beliefs together with a desire. Typically, the behaviour is intentionally described, that is, it relates to the specific objects or environmental properties interacted with. Consider the following belief-desire pair: S is thirsty and believes there is water in the cup. When S picks up the cup and drinks the water, her behaviour relates to water. When we change the scenario to where the cup contains twin water, we would say S picks up the cup and drinks twin water, which is a different behaviour. The problem for internalism is that such descriptions invoke wide belief states. For instance, in the first scenario, an explanation of S's drinking water will be glossed as S *believes there is water* in the cup, while the second scenario will be glossed as S *believes there is twin water* in the cup. Such psychological explanations of S's behaviour thereby appear to be fundamentally externalist. If we accept that the behaviours are different in these scenarios, and that the intentional descriptions relate to wide beliefs states, then internalists must explain what role, if any, narrow belief states or narrow contents play.

Internalists have tended to focus on two of the approaches outlined in Section 2.1. A first approach appeals to Frege cases and the explanatory role of modes of presentation. Where the internalist holds that modes of presentation relate to narrow contents, the case for an internalist account of behaviour involves defending the Fregean thesis (for instance Segal (2009b)). For present purposes I will set the elements of this view aside, as a complete discussion of this approach would require more space than is available here. What will be relevant below is the appeal to the mode of presentation associated with S's beliefs or concepts, and the role this plays in explaining behaviour.

A second approach has been developed by Fodor (1991) who provides an argument for the explanatory role of narrow content based on the idea that the notion of causation in psychology relates to proximal or local causes at the level of internal states. Fodor accepts that the intentionally described behaviours are different, but contends that the explanation of behaviour cannot rely on the wide belief states. If this is correct, then the challenge this poses to the externalist is to explain how wide belief states can causally explain S's behaviour, since in the water and twin water scenarios S's proximal causes can be the same, whilst her wide belief states are different. I will consider Fodor's view, and a response the externalist can make, in more detail in Section 3 below. Jackson & Pettit (1988) defend a hybrid account of psychological explanations of behaviour. While they agree with Fodor that at one level proximal causes are central to causally explaining behaviour, they aim to show how this can be so whilst wide belief states are also causally relevant. On this view, wide belief states are causally relevant because they state that one of a range of narrow proximal states was causally responsible for S's behaviour. However, Yablo (2003) has provided a defence of the causal relevance of wide states which would threaten to undermine the hybrid view by making the narrow states causally irrelevant. If successful, this objection would again leave the internalist in the position of having to explain why narrow states or contents were relevant to psychological explanations of behaviour.

The third approach identified above, relating to the rational role of reasons in causally explaining actions has not featured in internalist accounts of behaviour. I will show in Section 4 that this provides a more promising defence of internalism in psychological explanations of behaviour, and that it provides a clear response to Yablo's objection to internalist accounts. Before developing that view, I consider the views of Fodor, Jackson & Pettit in more detail, and explain that externalists have effective responses to these views which provide the internalist with motivation to look elsewhere for a role for narrow states or contents in explaining behaviour.

3. Proximal causes, generality and causal relevance

Section 3 provides a critical overview of two of the main internalist accounts of behaviour and highlights the main problems they face.

3.1 Proximal causes and causal powers

As noted in Section 2, Fodor highlights the centrality of proximal causes in mental causation and commonsense explanations of behaviour. In subsequent work, Fodor (1991) presents an argument for narrow content which contends that psychology classifies mental states by their causal powers which supervene on intrinsic properties, and that this serves as the basis for identifying a central role for narrow content. Central to Fodor's view is the idea that taxonomy in psychology, as in science generally, is by causal powers. On this view, we should individuate mental states by their causal powers. According to Fodor, the relevant notion of a causal power relates to intrinsic properties. For instance, what determines the causal powers of S's causally efficacious mental states are S's intrinsic properties. The underlying conception of causation relates to proximal or local causation. We can understand this notion in terms of object causation. For instance, where S kicks a ball into a window, the ball's breaking the window is causally responsible for the window breaking. The relevant causal power of the ball relates to its solidity (an intrinsic property relating to the air pressure inside the ball) so that when it impacts on the window, it will cause the window to break, thus the football causes the break by virtue of its intrinsic properties. Note that other properties of the ball are explanatorily irrelevant, such as the colour or price of the ball. Fodor applies this

model of causation to mental states. For instance, suppose S is thirsty and believes there is a cup of water in front of S, so that S picks up the cup and drinks the water. The causally relevant properties of S's belief are intrinsic properties, relating to the internal states of her body/brain. Thus, what causes S to lift the cup and drink the water are her internal, narrow, bodily states.

This provides a solution to the problem relating to wide intentional descriptions of behaviour. Suppose we consider a Twin Earth scenario where S is thirsty, but this time there is twin water in the cup. Our wide intentional description might explain S's behaviour of lifting the cup and drinking the twin water in terms of Twin S's belief *that there is twin water in the cup*. The problem for the internalist is that S and her Twin are intrinsically identical, and hence the same proximal causes are in play in each scenario, but the descriptions we use invoke different wide beliefs, relating to water and twin water. On Fodor's view, although intrinsic duplicates perform different behaviours on Earth and Twin Earth, drinking water and drinking twin water, the same proximal causes are ultimately what explain these actions at each world. Assuming that psychology is interested in proximal causation for mental taxonomy, we should therefore say that intrinsic duplicates share a causally efficacious mental state, which at different worlds can result in different behaviours[47]. Hence, we can identify a notion of a narrow mental state (or narrow mental content) which relates to the identity in causal powers of duplicates' intrinsic states. What seems right about Fodor's account is that there is an important role for proximal causes on an object view of causation, and the worry for wide intentional descriptions is that they are incapable of providing an adequate account of this form of mental causation.

---

[47] An internalist could individuate behaviour narrowly so that intrinsic duplicates perform the same behaviour. This would avoid the need to accept the claim that we need to account for different wide behaviours. While there is plausibly a kind of behaviour shared by intrinsic duplicates, such as bodily movements, the worry with this is that the externalist may accept this but hold that wide behaviour is more important to psychology, since behaviour is typically glossed in intentional terms. Thus, in order to show that narrow content is of central importance in both cases, the internalist should not simply rely on sameness of narrow behaviour as the basis for the explanatory role of narrow content in psychological explanations of behaviour.

An important assumption in Fodor's account is that mental states are individuated by their causal powers, where causal powers are understood as relating only to proximal causation. The case for narrow mental states succeeds only if causal powers supervene on intrinsic properties, and that intrinsic duplicates should be classified as having the same kind of mental states because their states share causal powers. If, on the other hand, there is a case that the causal powers of intrinsic duplicates might differ, or that there is some other way for wide intentional descriptions to take priority without dismissing proximal causation, then Fodor's argument will be undermined. In response to Fodor, in Section 3.2 I consider Burge's view that causal powers can be construed widely, and show that this provides a serious worry for Fodor's defence of internalism.

3.2 Wide causal powers

Externalists can accept that proximal causation is necessary at some level of description of causation, however, Fodor's claim is that taxonomy of mental states is by causal powers that relate to these proximal causes. The problem with this is that it is possible to understand causal powers as in some sense being wide, so that where the behaviours of duplicates are different, their causal powers are also different, which would thereby undermine the case for the explanatory role of narrow states. According to Burge (1989; 1995), intentional explanations of behaviour are fundamentally externalist. In contrast to Fodor's view of causal powers as supervening on intrinsic properties, Burge considers cases which he takes to show that there is an externalist notion of causal power and that this is operative in psychology. If this claim is right, then much of Fodor's view will be blocked.

Burge's case for wide causal powers involves noting how externalism is assumed in the approach to individuating states in empirical psychology. One example involves the individuation of hearts. According to Burge, hearts are a biological kind. Biological kinds are typically individuated in part by appealing to evolutionary origin and natural function, hence being a heart involves a pattern of causal relations to various other kinds in the environment in which the organ evolved. When we consider the causal powers of such kinds, Burge is suggesting that we need to

consider the organ within the evolutionary environment in which it evolved, so that we relate the causal powers to the specific kinds with which the organ typically, and historically, has had causal commerce with. Such a view marks a striking contrast with Fodor's species of narrow proximal causal powers. While Fodor connects narrow causes to his case for narrowly individuated mental states, Burge can utilise his broad causal powers in the service of externalist individuation of mental states. As the case for mental state individuation is not of concern in this chapter, I will restrict the discussion to the accounts of behaviour and the appeal to narrow and wide causes.

When we consider the Twin Earth cases, the same point can be made. Where S drinks water and her twin T drinks twin water, their states have different causal powers in so far as S's states have causal powers that relate to water and T's to twin water. The challenge to Fodor's position, and hence to the defence of internalism, is that the notion of a causal power isn't restricted to proximal causation. If it is right that scientifically interesting kinds like hearts and beliefs can be taken to possess causal powers that are individuated externalistically, then citing wide intentional descriptions would be sufficient for explaining behaviour since this is the right level of description for behaviours characterized intentionally.

It is worth noting that Burge does not need to deny that proximal causation is necessary. Both hearts and beliefs will require relevant internal/proximal causes. The issue, however, is that the notion of a causal power is not particularly well-defined, and there seems to be a notion of a wide causal power available to the externalist that seems plausible. The problem for the internalist then is that an externalist can hold that although proximal causes might be relevant, wide causes are what are causally relevant to psychological explanations of behaviour. Thus, as Burge contends, for taxonomic purposes intentional psychology can run mental states and behaviours together without denying that at some level causation must be proximal. The worry for internalism is that in terms of mental causation, there is no problem with wide states being causally relevant to causing actions intentionally described. Thus, the externalist can allow that at some level causation is proximal,

relating to bodily states, however, in terms of psychological explanations of behaviour, wide belief states are causally relevant. On the Burge-type view, they are causally relevant because the relevant causal powers are wide, so S's behaviour is causally explained by the wide belief state relating to water, while T's causal powers relate to twin water and hence her behaviour is causally explained by a different wide belief state.

It is worth flagging here that a two-component theory could accept that there are two kinds of behaviour connected to narrow and wide causes. However, although a two-component theorist could accommodate wide and narrow behaviour there seems to be something important about narrow causes – even on Burge's picture – since at some level systems must have proximal causes that do explanatory work. With regard to the differences between Fodor and Burge, a case can be made that there is a stand-off in the debate about wide and narrow causes. In so far as externalist individuation of kinds like hearts is viable, which the two-component picture can grant, simply appealing to narrow proximal causation does not seem like the best approach to demonstrate the explanatory primacy of narrow causes or narrow contents. Furthermore, there is a worry that in focusing on proximal causation Fodor's account fails to tell us anything about the rational role of beliefs in explaining behaviour. As noted in Section 2, belief-desire psychology was supposed to relate to commonsense explanations which, according to Davidson, are reason-giving explanations. Focusing only on proximal causes moves away from this key aspect of explanations of behaviour, and does not provide insight into the reasons why S did A since we are not given any account of why the relevant belief-desire pair provides a rational explanation of S's behaviour. Thus, in the rest of this chapter I make the case for this primacy of narrow content by considering Jackson & Pettit's view which makes narrow content primary and an objection from Yablo which threatens that picture. I then show, against Yablo, that an alternative case can be made for a similar thesis by showing how cognitive content is crucial to certain kinds of psychological explanations of behaviour.

3.3 Hybrid explanations: narrow and wide causes

Jackson & Pettit (1988) develop a hybrid theory of explanations of behaviour which assign different causal roles to wide and narrow content. On this view, internal (narrow) states are directly causally responsible for behaviour, and relate to what are called 'process' explanations. Wide contents are also taken to be causally relevant in a certain constrained sense, but without themselves being the direct causes of behaviour, serving instead as 'program' explanations. This approach is motivated by what Jackson & Pettit call the *doppelganger challenge*, which is seen as a problem for the externalist to explain what role wide contents play given an underlying view of causation in terms of proximal causes. The worry is that:

> …the ascription of [wide] content does not play an appropriate role in the explanation of behaviour. It is argued that it is the assumption of narrow content, content which does not depend on how things are outside the subject, which plays the appropriate role in explaining behaviour. (Jackson & Pettit 1988, 382)

As discussed above, some externalists will reject this construal of the relevant issues. However, Jackson & Pettit pitch themselves against this view, and they intend to provide an explanatory role to wide content, in addition to that of narrow content. This dialectical position is important for their account since they see themselves as aiming to provide an explanatory role for wide content given the (obvious) explanatory role of narrow states. While they intend to allow that wide states are causally relevant, this is highly constrained in that the relevant wide state:

> …plays no distinctive role, the explaining it does being best done by a combination of narrow psychological states together with environment. (Jackson & Pettit 1988, 390)

The basic idea is that the fundamental causal work is done by narrow proximal causes, and that the wide belief states typically invoked by descriptions of behaviour play an explanatory role in relation to how they function to draw attention to the relevant narrow states and the environment. The hybrid view provides a distinctive role for wide content. Jackson & Pettit provide the following example to illustrate this:

> I observe that Fred's sighting of that cup causes him to reach for it, and explain his behaviour in terms of his belief that it contains water. This explanation says that *a range of states* playing the belief that-that-cup-contains-water role would have caused him to reach for it, and that *one of them* did cause him to. (Jackson & Pettit 1988, 396) [My emphasis]

On this view, narrow states provide process explanations of behaviour in terms of the relevant internal processes, whilst wide belief states causally program for this behaviour, identifying a range of relevant states, without doing any causing. The wide state of *believing that the cup contains water* is not itself causally efficacious, but it is explanatory in that it functions to pick out a range of narrow states, one of which actually did cause the relevant behaviour. Wide explanations can therefore be seen as laying out the general causal terrain where their explanatory role is essentially related to the underlying proximal causes.

Unlike Fodor who relies on the appeal to causal powers and mental taxonomy, Jackson & Pettit rely on the inherent plausibility that proximal causation is at some level essential to causation, together with a claim about how intentional descriptions of behaviour, that cite wide states, function to draw attention to the narrow states. The problem is whether the externalist needs to accept that this is the extent to which wide states are causally relevant. If, contrary to this view, wide states may be construed as being causally relevant in some other sense, then the externalist may be able to reject the centrality of narrow states. One reason to doubt the view of Jackson & Pettit is that it is not obvious that wide descriptions function in this way. For instance, when we cite S's belief *that there is water in the cup*, and contrast this with her twin's belief *that there is twin water in the cup*, the externalist might appeal to the Burge-type view and draw a connection to the distinct belief states and the distinct objects interacted with.

The worry is that the view does not add enough to narrow content's being central to causal explanations to prevent an externalist digging their heels in and taking the wide state to provide an adequate explanation. For instance, the externalist might ask what has been left out of our explanation of S and T's behaviour when we only cite their wide belief states. S drinks water because she believes there is water in the cup,

and T drinks twin water because she believes there is twin water in the cup. This seems to be explanatory, in that it appeals to a relevant psychological state, a wide belief state, and explains the relevantly different behaviours in terms of these beliefs. It is therefore central to Jackson & Pettit's defence of the role of narrow content that the only reason the wide cause is casually relevant is because of its role in picking out the proximal causes. As was the case in Section 2, there appears to be something of an impasse at this stage. Above, it was noted that there is a plausible sense in which causal powers are wide, and this allowed the externalist to accept that proximal causes are important, whilst contending that wide states are causally relevant. Jackson & Pettit accept that wide states are causally relevant, but the problem is that an externalist may deny that the causal relevance of wide states is limited to the extent to which they pick out a range of narrow states.

3.4 Yablo on explanatory generality and causal relevance

Just how the externalist would want to say that wide states are causally relevant has not yet been made clear, and Jackson & Pettit could reply that without such an account the role of narrow states remains central. However, Yablo (2003) has provided a range of objections to the internalist's appeal to narrow states, which he takes to show that wide states alone are indeed causally relevant. Thus, in Section 3.4 I outline Yablo's rejection of the relevance of narrow states, which would support the externalist view that there is no need to appeal to proximal causes, and hence that citing wide states is explanatorily adequate for psychological explanations of behaviour.

3.4.1 Proximal causes, generality and locality

Yablo (2003) provides a partial reconstruction of the reasoning behind an account like that of Jackson & Pettit (1988). Recall that Jackson & Pettit's dialectical position was that narrow causes are taken for granted and that there is a question about the causal relevance of wide states. The hybrid view provided an account of the causal relevance of wide states which was essentially connected to the relevant narrow states that do the underlying causal work. Yablo (2003) picks up on a similar way of reasoning discussed by Williamson (2000) and notes that the purported

problem for externalism about explaining behaviour can be understood as arising from two main points:

(Locality) Causation of behaviour is proximal, requiring that the cause of some action is a relevant proximal state internal to the agent. Causes operate via a relevant causal mechanism or process, and so causation is 'here and now'

(Generality) There are common causes for different wide behaviours such as drinking water/drinking twin water, for instance a brain state common to intrinsic duplicates S and T

This distinction neatly captures the basis for Jackson & Pettit's hybrid view. On one hand (locality), proximal causes directly cause actions. On the other (generality), identical causes are common to different wide intentional behaviours. Thus, the problem for externalism is to explain how wide states are causally relevant given locality and generality. Following Williamson (2000), Yablo explains that we can reject the claim concerning generality by pointing out that it is plausible that a given action type A, such as drinking water, caused by some brain state M_1 could have been caused by some other brain state M_2, since the differences between M_1 and M_2 are not such as to prevent S from doing A. This seems right if the generality principle is that some *specific* narrow state M_x is the cause. This is consistent with Jackson & Pettit's hybrid view, since wide states relate to one of a range of possible narrow states, and intentional descriptions do not pick out a specific narrow (brain) state M_x. Contrast this with an account of generality in which a single mental state Q explains actions of type A, regardless of the underlying narrow states:

(Maximal Generality) Explanations of behaviour are maximally general when we can explain actions of type A with a common mental state Q

Explanations of this type are *more* general, since we can say that Q is the cause of A, rather than any of $M_1…M_n$, each of which might have been the cause of A. Since any of narrow states $M_1…M_n$ are sufficient for causing A, then, as Williamson notes, narrow causes are *less* general than wide causes. Consequently, if maximal generality is a reasonable constraint on an account of psychological explanations of behaviour, then we should deny that there are common causes to actions of type A, since different behaviours characterised by different wide intentional descriptions needn't have common causes. The objection then is that on a view like Jackson & Pettit's, it's not the case that the same narrow cause must be responsible for actions of type A as performed by subjects (whether intrinsic duplicates or not), since if A would have been performed had M_1 caused S to do A, and M_2 caused T to do A, then there is in fact not a common narrow cause. Maximal generality, however, can preserve the generalization that actions of type A are caused by mental states of type Q, and this would lend support to externalism in the familiar Twin Earth scenario. For instance, water-drinking behaviour (A_1) can be causally explained by water-beliefs (Q_1) while twin water-drinking behaviour (A_2) can be causally explained by twin water-beliefs (Q_2). Thus, where we think that psychological explanations should allow for generalizations of this nature, then only the externalist view can accommodate this constraint.

With regards to locality, Williamson's strategy is to appeal to paradigmatic wide states like knowledge and perception and to consider whether in such cases we need to appeal to proximal causes. If we can provide a causal explanation of behaviour without invoking the relevant set of narrow proximal causes, then we could reject locality. This would answer the point raised at the end of the previous section regarding why an externalist account which appealed only to the causal relevance of wide belief states would be adequate. According to Williamson, wide states like knowledge are 'prime' in the sense of lacking any internal component common to, say, a case of knowledge (the good case) and a case of belief without knowledge (the bad case). Suppose in the good case S does A because S *knows that P*. On this view, S's knowing that P causally explains why she does A. One reason for there being no common component in the good and bad case is that if a range of narrow states are

equally suitable for providing the relevant internal component for S to know that P, then brain state M_1 could be relevant to when S knows that P and M_2 could underlie a case where S does not know that P. The relevant difference is the external facts, not which brain state S is in. So, if prime states like knowledge can feature in psychological explanations of behaviour, then locality can be rejected in so far as the explanatory force of the prime mental state relates to the wide belief state and not to any of the possible range of narrow states which S could be in while being in such a wide state.

If this response works, then the worry for Jackson & Pettit is that the casual relevance of wide states might be secured without there being an essential connection to the underlying narrow states. What their view required was that a wide state was only causally relevant because it effectively states that one of a range of narrow states actually caused A. According to Williamson, prime states like knowledge are causally relevant without picking out a set of narrow states. However, while Yablo accepts the point about generality, he contends that Williamson doesn't give a satisfactory account of locality and the issue of proximal causation. This is surely correct in so far as Williamson appears to be aiming at a way of bypassing proximal causation so that in citing the relevant prime wide state we have fully explained S's action. The worry with this view is that it doesn't seem plausible to do without the underlying proximal causes altogether since prime states still involve relevant narrow states which must do *some* causal work. It doesn't seem that we can completely ignore proximal causes. For instance, while it may be right to say that prime states depend crucially on certain external factors and not internal factors, this does not show that citing a prime state does not thereby get its force partly from the relevant internal (brain) states that must occur with any given prime state. Thus, Yablo is right that Williamson's appeal to prime states is not enough to deal with proximal causation in terms of internal (brain) states.

One way to reject the claim that states of knowledge are irreducible is to highlight that in relevant scenarios, such as where S falsely believes that P and still does A, once we have the same behaviour type but a different wide mental state, it is

tempting to appeal to a common (narrow) factor in these cases as the relevant cause. For example, consider two cases, first where S drinks water because she believes that there is water in the glass, however, suppose S doesn't know there is water in the glass because her belief is not sufficiently justified (or formed in the right way). Second, S drinks water because she knows there is water in the glass. Only the latter case involves a prime mental state, yet without further justification for accepting Williamson's point we lack motivation for accepting that there is no common factor that explains S's behaviour. It is tempting to appeal to S's mental representations as the common factor, which provides a basis for a role for narrow content.

Here is where we are. Williamson provides a good case for preferring maximal generality if the internalist's only case for narrow causes is that some set of proximal causes are causally necessary. Where internalists appeal to a set of proximal causes, if it is correct that any member of this set would have resulted in the same action A, then it would seem that citing this set of possible proximal causes provides a less general explanation than citing a wide intentional mental state Q. What this asks us to choose between is the relevance of proximal causes and maximal generality. However, as yet we do not have a case for the irrelevance of narrow states. To this end, Yablo develops an alternative line of reasoning intended to respond to locality, so that citing wide causes *is* sufficient and that there is good reason to take them to be causally relevant without also requiring that we must cite relevant proximal causes. I will show that the objection is a good one, and that it raises serious problems for internalist accounts, which serves as motivation for an alternative internalist approach to defending a role for narrow causes.

3.4.2 Generality and the explanatory irrelevance of narrow states

As noted above, Williamson fails to satisfactorily account for locality. Where wide intentional states were taken to be causally relevant, what to say about proximal causes was not explained. To this end, Yablo develops a line of reasoning where locality is "…not refuted; it is not even mentioned; it is overruled" (Yablo 2003, 321). The aim is to show that in citing wide causes our interests in the causal explanation are satisfied. If this approach succeeds, then wide states do not get their causal

relevance by virtue of stating that a range of narrow states are causally responsible. The case for the irrelevance of narrow states combines Williamson's idea of generality with a second point developed by Yablo. This second point relates to an objection Yablo takes to be assumed by those who appeal to proximal causes in psychological explanations of behaviour. Yablo notes that one objection an internalist might make is that wide causes contain "extrinsic irrelevancies" (Yablo 2003) because they include redundant components in addition to proximal causes. Wide states like the belief *that P* involve the wide content associated with 'that p'. For instance, S's belief *that the cup contains water* relates to the wide content, which on one view of representational content is given by the actual cup and its contents (i.e. H_2O). The idea of irrelevance is that wide content, a central component of the wide belief state, is irrelevant to the actual proximal causes of S's behaviour (e.g. drinking the cup's contents) because it is "too far away" (Yablo 2003) to make a causal difference to S's actions.

In order to motivate the idea that narrow proximal causes are explanatorily irrelevant, Yablo points out that narrow causes also contain irrelevancies and hence that this is no reason to reject wide causes. The irrelevancies in question relate to the fact that, as in Jackson & Pettit's view, the wide intentional description highlights a set of proximal causes, only one of which was the actual cause. Now, since the causal explanation says that any of states $M_1...M_n$ could also have caused A, then the causal explanation contains extrinsic irrelevancies, namely those states that did not cause A. However, the internalist could respond that this does not show that the narrow states can be overruled. While it would be right that on Jackson & Pettit's view the wide explanation cites a set of possible causes which contains irrelevant states, this does not of itself show that the wide explanation should take priority. Furthermore, it seems open to the internalist to make a similar point in response to Yablo and hold that wide causes are in fact explanatorily redundant if the narrow causes provide a suitable explanation, since there is no need to invoke wide causes. Thus, an internalist might say that Jackson & Pettit are too concessive to externalism.

In response, Yablo can motivate the point by drawing attention to maximal generality, and link this to the point about relevance. Recall that Williamson's contention was that citing states $M_1 \ldots M_n$ was less general than citing a wide state Q, because in citing Q we can get maximal generality since states of type Q are taken as causes of actions of type A, whereas the internalist account must allow that a range of states could result in actions of type A. Combining these points allows Yablo to consider narrow causes casually irrelevant because the same kind of criticism about the irrelevance of wide content made by the internalist that underpins the appeal to narrow states can be applied to the internalist.

If narrow states can be overruled, or ignored, then we have some reason to think that appealing to wide states provides better explanations, because we get maximal generality and cite causally relevant wide states. For Jackson & Pettit, the wide cause is causally relevant because it cites a set of proximal causes. For Yablo, the wide cause cites a wide state only, with no requirement that it makes reference to proximal causes. In support, there is no scope to object to the wide state on the basis that the wide content is causally inert as it's 'too far away', since the set of narrow states likewise contains causally inert components. We are left with a trade-off between maximal generality, linking single wide states of type Q to actions of type A, and accounting for proximal causation and a set of narrow states $M_1 \ldots M_n$. If given a choice between maximal generality and attention to proximal causes, there is little the internalist can say to motivate preference for proximal causation without apparently begging the question against the externalist.

In the next section I develop and defend a different view which I contend provides a stronger case for internalism, as well as a suitable response to Yablo. The view I develop is intended to prevent any kind of externalist appeal to wide states, since if I am right then narrow contents are indispensable to psychological explanations of behaviour, without this relating simply to the notion of proximal causation, and wide states can play a role similar to that as identified by Jackson & Pettit.

4. Reasons, rationality and the explanatory role of narrow content

In Section 3 I considered two internalist accounts of why we need narrow causes in psychological explanations of behaviour. The most pressing objections to these views related to generality and causal relevance. The objections seem compelling in so far as the internalist appeals to a set of proximal causes, whereas the externalist appeals to causally relevant wide intentional states which may satisfy a preference for maximal generality in giving an explanation of actions of type A in terms of a single mental state Q. In Section 4 I contend that there is an alternative basis for defending internalism derived from criteria identified in Section 2 in relation to the role of reasons in giving rational explanations of why S performed an action A. Focusing on reasons provides a way of putting narrow content at the centre of explanations of behaviour. In particular, I base the view on Davidson's (1963) claim that we should be able to say something about what the subject saw in her actions which can rationalize why she did what she did, and McDowell's (2006) suggestion that subjects must have the capacity, in principle, to stand back and evaluate their reasons. I show how this can provide the basis for the claim that reasons that causally and rationally explain behaviour must be narrow (i.e. have a narrow content). I then show how this account provides a clear response to Yablo's objections based on generality and causal relevance.

4.1 Rationality and responsiveness to reasons

In Section 2 I introduced Davidson's (1963) account of reasons as causes of action. According to Davidson, giving reasons provides both a rational and causal explanation of action. On that view, the reason that we cite must meet several important conditions. The first is that:

> A reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action. (Davidson 1963, 23)

When we give a typical description of a subject's action, such as *S drank some water*, the reason we cite, such as *S was thirsty and believed there to be water in the cup*, must relate to how S sees her action. In addition, the primary reason why S does A must meet two further points:

R is a primary reason why an agent performed the action A under the description [D] only if R consists of *a pro attitude of the agent towards actions with a certain property*, and a belief of the agent *that A, under the description [D], has that property*. (Davidson 1963, 25) [My emphasis]

As noted in Section 2, the notion of having a *pro attitude* is to be favourably disposed to actions of that type. Thus, we might say that where S is thirsty, S's pro attitude relates to actions that involve S drinking water. Davidson is saying that a reason R rationally explains S's action only if R relates to S having a desire that relates to certain actions. The second point relates to the description, D, used to express R. Davidson holds that R is captured by D when S has a belief that the action has the relevant property (e.g. results in drinking) and this is adequately captured by D. Thus, for the description *S desires to drink some water and believes there is water in the cup* to be S's reason for drinking water, it must be the case that it meets these conditions. Where ordinary psychological explanations of behaviour cite a description, the description's causally explaining A depends on its picking out in some sense S's primary reason, although it can imply it as we needn't explicitly identify S's primary reason. Combining these points provides the following conditions on an account of reasons that causally explain action:

(1) The reason must tell us what S saw in her action

(2) The reason must relate to S having a pro attitude towards the relevant action

(3) The description of the reason must reflect that S believes the action to have a certain property and that S believes it under this description

I want to focus on the first and third points: what S saw in her action, and how D relates to what S believes about her action. Where we use the wide description *S is thirsty and believes there is water in the cup*, for this to be a primary reason it should say something about what S saw, and reflect S's belief about her action. Thus, when we consider the water example, it should be the case that S believes that her action is an instance of water-drinking, and that the description *S drank the water because she was thirsty and believed there was water in the cup* tells us what S saw and what she believed.

The question I want to raise is whether a typical wide intentional description would imply S's primary reason if it needs to meet conditions (1) and (3) above. This provides a basis for saying what descriptions would not provide S's primary reason. For instance, if a given description does not tell us what S saw in her action, and doesn't capture how S thinks of the relevant action type, then we have not been given a reason that will provide a rational, and hence causal, explanation of S's doing A. Consequently, an argument against wide intentional descriptions being primary reasons can be developed from Davidson's criteria on rational explanations of action. If successful, the view that wide intentional descriptions provide S's primary reasons can be resisted, since the internalist will deny that the description captures what S believes and tells us what S saw in her action. If we have reason to doubt that the wide intentional description does meet these conditions, then wide intentional descriptions may fail to rationally explain S's actions.

Before developing a case in favour of internalism, I want to introduce a further condition on S's reason providing a rational explanation of her action from McDowell's (2006) account of reasons and rationality. McDowell's notion of rationality in his account of reasons for action makes use of the idea of "…responsiveness to reasons *as such*" (McDowell 2006, 2). This allows for a distinction between acting for reasons, on the one hand, and acting for reasons *as such*, on the other. The former is suitable for animals that lack rationality in McDowell's sense, so that a reason for an animal's action of fleeing might be *that danger is near*, without this being a reason as such for the animal. The latter notion of acting for reasons as such involves the reason serving as a subject's rational justification. McDowell (2006, 2) stresses the importance of a rational subject being able, in principle, to stand back and evaluate the reason. For instance, S should be able to assess whether the apparent danger provides her with good reason to flee, something apparently not available to non-human animals which consequently lack rationality in McDowell's sense.[48]

---

[48] McDowell's (2006) aim was to argue that since perceptual beliefs or experiences can serve as rational justifications of action, this could only be the case if perception was in some sense conceptual. The basic idea was that for experience to provide a rational ground for action, serving as a reason for action, then it must be recognised for the reason it provides, which entails that it is conceptual as it

From this view we can take it that the ability to reflect on one's reasons endows a subject with the capacity to recognise the reason for the reason it is. That is, in having access to her reasons S can reflect on what her reasons are and, perhaps, why they would serve as reasons for doing A. Consider McDowell's example. S can reflect on her reason for running away, *that danger is near*, which requires that she can entertain a belief with this content, or equivalently, that she can entertain a thought with this content by virtue of possessing the relevant concepts. Given this view, it seems to follow that recognising a reason for the reason it is implies that a subject is able to distinguish differences between possible reasons for some action A, for instance between *the cliff edge is dangerous* and *the loose rocks at the cliff edge are dangerous*. Differences in content entail differences in reasons, and if S has the capacity to reflect on the content that rationally explains her action, then S should be able to distinguish between possible contents that would not explain her action and those that would. That is, given responsiveness to reasons as such, S is able to identify, in principle, what her reason for doing A was. Thus, if the former reason, *the cliff edge is dangerous,* was S's reason, then S should be able to reflect on this and recognise that this was her reason, and not *the loose rocks at the cliff edge are dangerous*.

Davidson's and McDowell's points can be seen to overlap. From Davidson, I identified two points: what S saw in her action, and S's belief that her action has a relevant property under the description D. From McDowell, we can now add that the description must also relate to the reason that S is rationally responsive to, such that S would be able to stand back and reflect on her reason, as captured by D. Where D tells us what S saw in her action, D will also provide us with the reason that S is responsive to, hence the reason which caused S's action. Suppose that S's reason for fleeing was that the cliff edge is dangerous. What S might reasonably see in her action is that it is the evasion of the dangerous cliff edge. So, the description *S ran away from the cliff edge because she believed that it was dangerous* tells us what S

---

relates directly to rationality. Now, my point does not depend on the debate about conceptual and nonconceptual content, or if conceptual capacities are operative in perception. The point I develop is the idea of recognising reasons as reasons, so that what causally explains S's action is what is available to her for rational reflection.

saw in her action and gives us the reason she was responsive to and can stand back and reflect on. This gives the following rational constraint on psychological explanations of behaviour:

(Rationality)   Rational explanations of action require that the reason R, expressed by description D, captures what the agent saw in their action, and provides the reason the agent is rationally responsive to

These conditions on D providing a rational explanation of why S did A provide the basis for an argument in favour of internalism as the primary explanatory basis of psychological explanations of behaviour which I will develop in the next section. Before that, it's important to explain why externalists and internalists alike should be interested in rationality when considering explanations of behaviour. To anticipate my response to Yablo's appeal to maximal generality below, I will consider two general points in favour of placing the emphasis on, or at least recognising the importance of, rationality. First, psychological explanations that cite mental representations or their contents in explanations of behaviour are tied to rationality in the following manner. Typically, we assume that subjects are rational, in that they will not knowingly hold conflicting beliefs, and, importantly, if S desires that P and believes that doing A will satisfy her desire, then S doing A relates to S's reasons and hence rationality. Second, the defence of wide causes by motivating the principle of maximal generality involves making sense of S doing A by virtue of her mental state Q. What explanatory power Q has must relate, in part, to the rational domain, since citing Q involves citing one of S's beliefs, which, as just noted, typically explains why S does A in terms of her reasons. This appeal to rationality is as yet neutral between wide and narrow content, and should be accepted by both externalists and internalists. However, as I will now show, once we consider the importance of rationality, a case can be made that only narrow content can provide the required link to rationality and S's reasons for action.

4.2 Primary reasons and cognitive content

In the example above, we considered two contents that S can distinguish which relate to the dangerousness of the cliff edge, and the dangerousness of the loose rocks at the cliff edge. When we consider the water example in the context of a Twin Earth scenario we can question whether S will be able to distinguish between two reasons with different wide contents. For instance, consider scenario 1 where the cup contains water (H_2O) and scenario 2 where the cup contains twin water (XYZ). If S picks up the cup and drinks the contents in each scenario, this results in different actions A_1 and A_2: in scenario 1 S picks up and drinks water, in scenario 2 S picks up and drinks twin water. Prima facie, it seems that S could fail to distinguish between her reasons for performing these different actions. For instance, H_2O and XYZ are, by definition, superficially identical, so when S perceives H_2O and XYZ, then even if the contents of her perceptual state are wide in some sense, S would fail to recognise any difference in the contents of her perceptual states. Suppose for present purposes that the contents of S's wide states related to her reasons for performing these actions. This follows from a reasonable assumption that our perceptual contact with the world in some sense guides our actions. It is plausible that S would not be able to tell the difference between the different perceptual contents or between the actions. For instance, suppose S reflects on her reasons in each scenario, which may involve accessing the contents of her perceptual states. Since the two liquids are superficially identical, S would plausibly be unable to distinguish between the liquids. Likewise, S would seemingly fail to recognise a difference in the distinct actions of drinking water and drinking twin water, since from S's perspective she may take herself to be performing the same action type, such as drinking a liquid.

So far I have talked about perceptual contents providing S's reasons, and S being unable to distinguish between the contents of her perceptual states. It is possible to provide further support to this claim by appealing to the account of concepts and cognitive content developed in Chapter 2. Suppose S has a prototype concept WATER. Prototype concepts represent superficial properties, such as appearance properties, and these properties are shared by H_2O and XYZ. S's WATER prototype therefore has a narrow representational content which relates to this appearance

property. When S encounters H_2O and XYZ in scenario 1 and 2, the same prototype is triggered, and S represents the same superficial properties in each case. Furthermore, the cognitive content of prototypes derives from perceptual templates which encode information derived from perceptual experience, specifically in terms of narrow phenomenal properties of the experience. Such cognitive content is narrow because it supervenes on S's intrinsic properties. Now, where S's WATER prototype is triggered in both scenarios, S thinks of each liquid in the same way, via the identical narrow cognitive content: S thinks of both H_2O and XYZ in terms of the same superficial appearance and functional properties (the watery properties, for short).

The case for internalism can now be made as follows. If we understand S's reasons as relating to the cognitive content of her concepts, then in both scenarios the same narrow cognitive content reflects how S thinks of water and twin water. The connection between cognitive content and S's reasons can be motivated by considering what aspect of the content of her concepts S is responsive to. One way to determine this is to see what S would judge as being in the cup in each scenario. Following the typicality effects experiments discussed in Chapter 2 where subjects make category judgements based on superficial properties, we have reason to think that S would judge that the contents of the cups have the same superficial appearance properties. Thus, S's reason will likely relate to these superficial properties, and in turn to the cognitive content of her prototype concept WATER, since this reflects how she thinks of these properties.

It should be noted that it is plausible that S's perceptual states and concepts have wide contents. For instance, where S perceives water, the wide content relates to H_2O, and where this is slowly[49] switched with some XYZ the wide content relates to the XYZ. Furthermore, S's concepts may have wide content which relate to the specific substance in the cup at the relevant time and context. However, it is plausibly the cognitive content of S's general concepts which S has cognitive access

---

[49] The literature on switch cases is extensive, and I assume here that the change in wide content is unproblematic.

to, that is, which reflects the way S represents things as being around her. Consequently, in both scenarios, S will be in a state with the same narrow cognitive content.

We can now return to the problem with wide content. Suppose we try to capture S's reasons for drinking water/twin water using the typical wide intentional description *S believed there was water in the cup* (D_1) and S *believed there was twin water in the cup* (D_2). Descriptions D_1 and D_2 must tell us what S saw in her actions (A_1 and A_2) and provide the reasons S was responsive to. By hypothesis, D_1 and D_2 imply different reasons, in so far as S is being held to have a belief specifically about each substance in each scenario. Where cognitive content is narrow, S would not think of these substances in different ways since the cognitive content of S's concepts tells us how she represents things as being. I outlined why we should think S is responsive to the narrow cognitive content of her concepts. If this is right, then when S drinks water and drinks twin water S's reason in both scenarios is plausibly the same, given in terms of the way that she thinks of both liquids in terms of identical superficial appearance properties. Therefore, the reason S is responsive to will be something like *there is watery stuff in the cup*, where 'watery stuff' relates to the cognitive content of S's WATER prototype. Consequently, D_1 and D_2 fail to meet our conditions, and so do not provide rational, and hence causal, explanations of why S did A_1 and A_2.

I conclude that it is narrow cognitive content that is most causally relevant in psychological explanations of behaviour, and not wide mental states or contents. Nonetheless, wide intentional descriptions may imply something like S's narrow cognitive content, and hence they could be true or imply something true. The truth of the wide description would rely on it implying something that more closely reflects S's actual reasons – her primary reason, given by the cognitive content of her concept. A standard wide description may sometimes fail to convey this, but as long as it implies something like *S thought that there was watery stuff in the cup*, then it will do for normal purposes. The upshot of this is that we have the basis for an alternative hybrid theory. Unlike Jackson & Pettit (1988), wide descriptions are

causally relevant because they imply a primary reason, not because they identify one of a range of narrow states that was causally responsible for A. I will show in the next section that this provides the basis for a response to Yablo.

4.3 Explanatory generality and causal relevance revisited

In Section 3.4 I raised two related objections to internalist accounts of psychological explanations of behaviour. The first relates to the generality of psychological explanations. Following Williamson (2000), Yablo (2003) contends that explanations of behaviour that cite a range of sufficient proximal causes (narrow states) are less general than explanations that cite a single wide mental state Q (such as believing that P). The second relates to the case for the causal relevance of wide causes, which in turn relates to the irrelevance of narrow causes. The irrelevance of narrow states was based on two points. First, Yablo responds to the concern that wide causes are 'too far away' and hence contain extrinsic irrelevancies by highlighting that narrow causes also contain irrelevant components, such as the redundant narrow states in the set of narrow states picked out by Jackson & Pettit. Second, if we explain the causal relevance of wide states in terms of a wide mental state Q, which provides a maximally general account of the causes of actions of type A, then Q alone may be sufficient for providing an adequate explanation of actions of type A. This is because if we prefer maximal generality, then our citing Q does not invoke whatever narrow states Q may relate to on an occasion.

The response I will develop here notes that there is a trade-off between maximal generality, where a single wide state Q explains actions of type A, and the account of rational explanations of behaviour developed above, where rational explanations of action involve citing S's reason for doing A which rationally and causally explains actions of type A. The view developed in Sections 4.1-4.2 appeals to the rational role of reasons, and identifies narrow cognitive content as playing the role of S's primary reason. This does not involve saying that a set of narrow states would cause A and that one of them actually did. It may rely on this fact, since narrow content supervenes on intrinsic properties, but the explanation relies on rationality, not on proximal causes. In a sense, there is more generality in this account than that

associated with Jackson & Pettit's (1988) view, since we cite the relevant reason or cognitive content for doing A, rather than a set of possible narrow states. We can say that it is a given narrow cognitive content that caused A, and although this may supervene on a similar set of narrow states, the explanation is pitched in terms of rationality, and not proximal causes. However, that my view is less general than the externalist view follows from allowing the same reason to explain actions A_1 and A_2. For instance, the same narrow cognitive content can explain why S drinks water and twin water. Whether this is a disadvantage of my account is not obvious, since it has not been shown that we require unique mental states Q_1 and Q_2 to rationally explain distinct actions A_1 and A_2. In fact it would be surprising if this were the case, since two subjects could plausibly perform the same action type without having the same reasons, which means that we should not expect a single state Q to relate to actions of type A. Hence, we shouldn't be surprised if the same reason could explain two different actions.

Setting this point aside for now, a case in favour of internalism can be made even supposing that maximal generality is a desirable feature of psychological explanations of behaviour. Given that our interest is in what it was about S's reasons that caused her to do A, rationality should be central to psychological explanations that invoke reasons for performing actions. An internalist can accept that Williamson's maximal generality principle allows wide states to be more general than narrow contents. However, I contend that in the trade-off between maximal generality where only one wide state Q explains actions of type A, and rationality, where narrow cognitive content best provide S's primary reason, we should prefer the latter over the former. Choosing maximal generality at the expense of rationality would render the externalist's account unable to link psychological explanations to the rational role of reasons. Thus, we should tolerate some variation in causes of actions of type A, and indeed sameness of reasons for different actions A_1 and A_2, if we can allow that the reasons we cite as causes provide rationally intelligible psychological explanations of behaviour.

4.4 Against Fodor on Frege cases

In Section 2 I introduced Frege cases which relate to psychological explanations of behaviour in cases where subjects do not know that two names refer to the same thing. In these cases, Fregeans contend that what explains behaviour is the relevant mode of presentation of the referent, such that an explanation only succeeds where the explanation relates to the mode of presentation the subject associates with the referent. Millians, on the other hand, contend that all modes of presentation are explanatorily irrelevant, and hence that the wide semantic content associated with either term can provide a suitable explanation.

Eschewing his earlier appeal to narrow causes of behaviour, Fodor (1994) defends a Millian account of Frege cases. The basic idea is that, in general, holding everything else equal, we can explain why S does A by appealing to the wide Millian content of her mental representations. Thus, ceteris paribus, S will do A because A believes that P (where that P is a relevant wide content). In a similar spirit to Yablo (2003), Fodor (1994) endorses the view that scientific explanations typically aim for maximal generality, such that if a single mental state or content Q can explain actions of type A, then this should be preferred to introducing additional explanatory devices to account for unusual cases. What is important, according to Fodor, is that the Frege cases are exceptions to general psychological explanatory practices which get by perfectly well citing wide Millian contents. The differences between Fregean and Millian accounts can be understood by considering the following two scenarios, from Segal (2009b) [50]:

(4) Oedipus wants to fly to Hesperus

(5) Oedipus believes that the USS Evening Star is about to depart for Hesperus

(6) So, Oedipus boards the USS Evening Star

(7) Oedipus wants to fly to Hesperus

(8) Oedipus believes that the USS Evening Star is about to depart for Phosphorus

---

[50] Segal's (2009b) objection to Fodor's account differs from the one I present here. For another objection to Fodor (1994) on Frege cases see Wakefield (2000).

(9) So, Oedipus boards the USS Evening Star

Fregeans hold that while (4)-(6) is acceptable, (7)-(9) is problematic, since (9) doesn't seem to follow from (7) and (8). According to Fodor's view, however, we may ignore the appearance of a problem with (7)-(9) because such cases are exceptions to general explanatory principles which allow us to explain behaviour by citing the wide Millian content of S's beliefs. Recall from Chapter 4 that Millians hold that occurrences of co-referring names in propositional attitude ascriptions have the same semantic content. Thus 'Hesperus' in (5) has the same content as 'Phosphorus' in (8). What Fodor is saying is that since explanations can ordinarily get by perfectly well by appealing to wide Millian content, then we can effectively ignore cases such as (7)-(9), in preference for maintaining maximal generality in our explanatory practices.

The importance of rationality in typical explanations of behaviour, however, causes a problem for Fodor. Above I suggested that rationality is presumed in psychological explanations of behaviour that cite mental representations as causes of behaviour. This is because mental representations are linked to the rational domain, and, as such, there is a presumption of rationality in psychological explanations of behaviour. The worry with ceteris paribus clauses then is that it would seem that Oedipus' behaviour in (7)-(9) is not rationally intelligible, and in preferring generality, Fodor gives up the link to rationality and hence, ultimately, to an important explanatory component of the subject's mental representations. If rationality is going to be side-lined for generality, then this looks like an unattractive consequence of Fodor's position. Now, it is not the case that we are then forced to choose between rationality and generality. On my account, explanations that cite the narrow contents of S's states can provide suitably general explanations of behaviour, though perhaps not maximally general explanations. But as I noted above, there are sufficient problems with maximal generality which raise doubts about the plausibility of the principle as a general condition of psychological explanations of behaviour.

5. Conclusion

In this chapter I hope to have shown that narrow content is central to psychological explanations of behaviour by virtue of the rational role of reasons in causally and rationally explaining why S does A. This provides a suitable basis for internalism about psychological explanations of behaviour which has advantages over the views discussed in Section 3, in part because it moves away from the reliance on the proximal causes model of explanation. A key problem for Jackson & Pettit's (1988) hybrid theory was that it is susceptible to Yablo's (2003) concerns about generality and locality. My alternative account focused on reason-giving explanations. From Davidson (1963) and McDowell (2006), two main constraints on an adequate rational explanation of behaviour were identified. These are that a description D should tell us what S saw in her action, and should provide the reason that S is responsive to and hence which caused her to do A. I showed that wide intentional descriptions do not provide plausible rational explanations of actions of type A. This provided a problem for externalist accounts of psychological explanations of behaviour. Furthermore, my view provides a suitable response to Yablo's objections and allows the internalist to opt for rationality over maximal generality. Finally, my account lends support to a Fregean account of Frege cases, and shows that Fodor's (1994) appeal to ceteris paribus clauses can be undermined in so far as psychological explanations needn't prioritize generality over rationality, since rationality should be fundamental to psychological explanations.

Conclusion

In part one of this thesis, I argued for a pluralist view of narrow content, grounding it in two main forms of representation and in concept pluralism. The central argument of Chapter 1 was that representation comes in a variety of forms, both externalist and internalist. What is important in understanding the representational nature of narrow content is to correctly characterise the supervenience claim and the relevant representational framework. Where existing accounts focus on just one form of representation, the view I propose suggests that we should countenance a number of different kinds of approaches to representation, including descriptive representation and informational (or relational) representation. Furthermore, narrow content relates to different sorts of properties, encompassing natural and non-natural kind properties, and correspondingly different mental representations and truth conditions. Chapter 2 provided a suitable foundation for the pluralist account by showing that mental representations can relate to a range of conceptual structures, each of which may relate to conceptualizing the world in different ways, and in terms of different properties. I showed that concept pluralism is consistent with internalist individuation, and that the cognitive content of concepts relates in part to phenomenal internalism through the mechanism of perceptual templates. I defended this view against Weiskopf's (2007) argument against narrow cognitive content, as well as the notion of phenomenal externalism (Tye 2000), and the idea that narrow content may be too complex to play the role of cognitive content (Vaughan 1989). I then showed, in Chapter 3, that the internalist should not be worried about arguments for externalism based on social externalism (Burge 1979; Goldberg 2002), since these appear to be at odds with ways of individuating mental representations on the concept pluralist approach. I also showed that perceptual psychology is consistent with narrow content, and that the approaches to individuation from Burge (2010) and Egan (2010) do not put pressure on the notion of narrow content in psychology.

Some of the advantages of my account relate to avoiding the problems noted in Chapters 1 and 2 for existing accounts. In particular, my account provides a viable basis for narrow content in an approach to mental representations derived from recent work in philosophy and empirical psychology, while avoiding problems associated with Prinz' (2000; 2002) similar appeal to work in the psychological literature. Approaching narrow content in this way also avoids the kinds of issues that arise in relation to two-dimensional semantics, especially Chalmers' (2003) epistemic framework.

In part two, I developed the semantic and psychological role of narrow content in propositional attitude ascriptions and psychological explanations of behaviour. The account developed in part one played an important role in both of these areas. In Chapter 4, I showed that narrow content can relate to the truth conditions of Fregean attitude ascriptions. There were two main consequences of this view. In showing that attitude ascriptions may express narrow contents, it was possible to show that two arguments for externalism (Putnam 1975; Burge 1979), and one against internalism (Soames 2002), could be rejected in so far as they rely on the assumption that attitude ascriptions only express wide contents. Second, given the variety of mental representations and narrow contents that may be associated with a term like 'water', I showed that there is no simple reading-off mental content from attitude ascribing sentences. Consequently, to ascertain the truth conditions of attitude ascriptions we must consider the narrow content expressed. Chapter 5 made the case that psychological explanations of behaviour are best accounted for in terms of the reasons why subjects perform certain actions, and that only narrow content can play this role. There were two main advantages of this approach over existing internalist accounts from Fodor (1991) and Jackson & Pettit (1988). First, in moving away from the role of narrow states, I showed how it is that narrow content plays a central role in explaining behaviour. Second, appealing to the role of reasons provided a way to respond to Yablo's (2003) objections that can be raised against the hybrid theory of Jackson & Pettit, as well as providing a response to Fodor's (1994) appeal to ceteris paribus clauses to reject Fregean explanations of behaviour.

I was unable to consider the complete range of accounts of wide and narrow content, and other important issues thrown up by the externalism/internalism debate. This is partly due to the fact that a detailed study of wide and narrow content would require a much longer work (perhaps a lifetime of study), and partly because the issues discussed here form a coherent line of thought. What drives the discussion is the nature and role of mental content in philosophy and psychology. As noted in the introduction to this thesis, Fodor (among others) has embarked on a project of showing that all content is wide, and that this is sufficient for accounting for mental representations and for psychological explanation in psychology. Thus, part of my aim has been to rebut much of this picture by showing the importance of narrow content in understanding mental representations and psychological explanation, without needing to reject the notion of wide content.

Bibliography

Bach, K. (1997). Do belief reports report beliefs? *Pacific Philosophical Quarterly* 78 (3):215-241

Bach, K. & Elugardo, R. (2003). Conceptual minimalism and anti-individualism: a reply to Goldberg. *Noûs* 37 (1):151-160

Barsalou, L.W. (1987). The instability of graded structure: implications for the nature of concepts. In U. Neisser (ed.), *Concepts and conceptual development: ecological and intellectual factors in categorization*. Cambridge University Press. 101-140

Barsalou, L.W. (2008). Situating concepts. In P. Robbins & M. Aydede (Eds.), *Cambridge handbook of situated cognition* (pp. 236-263), New York: Cambridge University Press

Bermudez, J.L. (2009). The distinction between conceptual and nonconceptual Content. In McLaughlin, P. (ed.) *The Oxford handbook of philosophy of mind*. Oxford: Oxford University Press

Besson, C. (2012). Empty natural kind terms and dry earth. *Erkenntni*s 76 (3):403-425

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy* 10 (1):615-78

_____(1990). Inverted earth. *Philosophical Perspectives* 4:53-79 (1990)

Block, N. & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review* 108 (1):1-46

Borg, E. (2009). Must a semantic minimalist be a semantic internalist? *Aristotelian Society Supplementary Volume* 83 (1):31-51

Braddon-Mitchell, D. & Jackson, F. (1997). *Philosophy of mind and cognition*. Blackwell

Braun, D. (1995). Causally relevant properties. *Philosophical Perspectives* 9:447-75

_____(2002). Cognitive significance, attitude ascriptions, and ways of believing propositions. *Philosophical Studies* 108 (1-2):65-81

_____(2006). Kripke's revenge. *Philosophical Studies* 128 (3):669 - 682

Brogaard, B. (2007). Attitude reports: do you mind the gap? *Philosophy Compass* 3 (1):93- 118

Brown, C. (1993). Belief states and narrow content. *Mind and Language* 8 (3):343-67.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy* 4 (1) pp. 73- 122

_____(1982). Other bodies. In Andrew Woodfield (ed.), *Thought and object*. Oxford University Press

_____(1986a). Individualism and psychology. *Philosophical Review* 95 (January):3-45

_____(1986b). Intellectual norms and foundations of mind. *Journal of Philosophy* 83 (December):697-720

_____(1989). Individuation and causation in psychology. *Pacific Philosophical Quarterly* 707 (4):303-22

_____(1995). Intentional properties and causation. In C. Macdonald & Graham F. Macdonald (eds.), *Philosophy of psychology: debates about psychological explanation*. Blackwell

_____(2003a). Perception. *International Journal of Psychoanalysis* 84 (1):157-167

_____(2003b). Phenomenality and reference: reply to Loar. In Martin Hahn & B. Ramberg (eds.), *Reflections and replies: essays on the philosophy of Tyler Burge*. MIT Press. 435-450

_____(2007). *Foundations of mind*. Oxford University Press

_____(2009a). Five theses on de re states and attitudes. In Joseph Almog & Paolo Leonardi (eds.), *The philosophy of David Kaplan*. Oxford University Press. 246-324

_____(2009b). Perceptual objectivity. *Philosophical Review* 118 (3):285-324

_____(2010). *Origins of objectivity*. Oxford: Oxford University Press

Byrne, A. & Pryor (2006). Bad intensions. In Manuel Garcia-Carpintero & Maci (eds.), *Two-dimensional semantics: foundations and applications*. Oxford University Press. 38-54

Byrne, A. & Tye, M. (2006). Qualia ain't in the head. *Noûs* 40 (2):241-255

Cain, M.J. (2010). Linguistics, psychology, and the scientific study of language. *Dialectica* 64 (3):385-404

Cappelen, H. & Lepore, E. (2005). *Insensitive semantics: a defense of semantic minimalism and speech act pluralism*. Blackwell Publishing

Chalmers, D. (1996). *The conscious mind: in search of a fundamental theory*. Oxford University Press

_____ (2002). The components of content. In David J. Chalmers (ed.), *Philosophy of mind: classical and contemporary readings*. Oxford University Press

_____ (2003). The nature of narrow content. *Philosophical Issues* 13 (1):46-66

_____ (2004a). Epistemic two-dimensional semantics. *Philosophical Studies* 118 (1-2):153-226

_____ (2004b). The representational character of experience. In Brian Leiter (ed.), *The future for philosophy*. Oxford: Oxford University Press

_____ (2006a). Perception and the fall from Eden. In Tamar S. Gendler & John Hawthorne (eds.), *Perceptual experience*. Oxford University Press. 49-125

_____ (2006b). Two-dimensional semantics. In E. Lepore & B. Smith (eds.), *The Oxford handbook of the philosophy of language*, Oxford University Press

_____ (2011). Propositions and attitude ascriptions: a Fregean account. *Noûs* 45 (4):595-639

_____ (2012). *Constructing the world*. Oxford University Press

Chalmers, D. & Clark, A. (1998). The extended mind. *Analysis* 58: 10-23

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press

Collins, J. (2009). Methodology not metaphysics: against semantic externalism. *Aristotelian Society Supplementary Volume* 83 (1):53-69

Crane, T. (1991). All the difference in the world. *Philosophical Quarterly* 41 (January):1-25

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy* 60 (23):685-700

⎯⎯⎯⎯(1968). On saying that. *Synthese* 19 (1-2):130-146

Davies, M. (1986). Externality, psychological explanation, and narrow content. *Proceedings of the Aristotelian Society, Supplementary Volumes*, Vol. 60 (1986), pp.263-283

Dennett, D. (1987). *The intentional stance*. MIT Press

Devitt, M. (2006). *Ignorance of language*. Oxford: Clarendon Press

Dretske, F. (1981). *Knowledge and the flow of information*. MIT Press

⎯⎯⎯⎯(1993). Mental events as structuring causes of behaviour. In John Heil & Alfred R. Mele (eds.), *Mental causation*. Oxford University Press

Dupré, J. (1981). Natural kinds and biological taxa. *Philosophical Review* 90 (1):66-90

Egan, F. (1992). Individualism, computation, and perceptual content. *Mind* 101 (403):443-59

⎯⎯⎯⎯(1995). Computation and content. *Philosophical Review* 104 (2):181-203

⎯⎯⎯⎯(2009). Wide content. In McLaughlin, P. (ed.) *The Oxford handbook of philosophy of mind*, Oxford University Press

⎯⎯⎯⎯(2010). Computational models: a modest role for content. *Studies in History and Philosophy of Science Part A* 41 (3):253-259

Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press

Farkas, K. (2003). What is externalism? *Philosophical Studies* 112 (3):187-208

⎯⎯⎯⎯(2006). Semantic internalism and externalism. In Ernest Lepore & Barry C. Smith (eds.), *The Oxford handbook of philosophy of language*. Oxford University Press

⎯⎯⎯⎯(2008a). Phenomenal intentionality without compromise. *The Monist* 91 (2):273-93

⎯⎯⎯⎯(2008b). *The subject's point of view*. Oxford University Press

Fisher, J. (2007). Why nothing mental is just in the head. *Nous* 41 (2):318-334

Fodor, J.A. (1975). *The language of thought.* Harvard University Press

⎯⎯⎯⎯(1982). Cognitive science and the twin earth problem. *Notre Dame Journal of Formal Logic* 23 (April):98-118

_____(1986). Individualism and supervenience. *Proceedings of the Aristotelian Society, Supplementary Volumes*, Vol. 60 (1986), pp.235-262

_____(1987). *Psychosemantics: the problem of meaning in the philosophy of mind*. MIT Press

_____(1990). *A theory of content and other essays*. MIT Press

_____(1991). A modal argument for narrow content. *Journal of Philosophy* 88 (1):5-26

_____(1994). *The elm and the expert*. MIT Press

_____(1998). *Concepts: where cognitive science went wrong*. Oxford University Press

Fodor, J.A. & Lepore, E. (1992). *Holism: a shopper's guide*. Blackwell

_____(1996). The pet fish and the red herring: why concepts aren't prototypes. *Cognition* 58 (2):243-76

Frege, G. (2000). Über sinn und bedeutung (on sense and reference). In Robert J. Stainton (ed.), *Perspectives in the philosophy of language: a concise anthology*. Broadview Press

Gallese, V. & Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 2005, 21 (0)

Georgalis, N. (1999). Rethinking Burge's thought experiment. *Synthese* 118 (2):145 64

_____(2003). Burge's thought experiment: still in need of defence. *Erkenntnis* 58 (2):267-273

Goldberg, S. (2002). Do anti-individualistic construals of propositional attitudes capture the agent's conceptions? *Noûs* 36 (4):597-621

Higginbotham, J. (2006). Languages and idiolects: their language and ours. In Barry C. Smith (ed.), *The Oxford handbook of philosophy of language*. Oxford University Press. 140-50

Horgan, T. & Tienson, J. (2002). The phenomenology of intentionality and the intentionality of phenomenology. In David J. Chalmers (ed.), *Philosophy of mind: classical and contemporary readings*. Oxford University Press: 520-533

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press

Jackson, F. (1998). Reference and description revisited. *Philosophical Perspectives* 12 (S12):201-218

———(2003a). Narrow content and representation - or twin earth revisited. *Proceedings and Addresses of the American Philosophical Association* 77 (2):55-70

———(2003b). Representation and narrow belief. *Philosophical Issues* 13 (1):99-112

———(2007). Reference and description from the descriptivist's corner. *Philosophical Books* 48 (1):17-26

Jackson, F. & Pettit, P. (1988). Functionalism and broad content. *Mind* 97 (July):318-400

Johnston, M. (1998). Are manifest qualities response-dependent? *The Monist* 81 (1):3-43

Kallestrup, J. (2011a). Actually-rigidified descriptivism revisited. *Dialectica* 66 (1):5 21

———(2011b). *Semantic externalism*. Routledge

Kaplan, D. (1989). Demonstratives. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*. Oxford University Press. 481-563

Koslicki, K. (2008). Natural kinds and natural kind terms. *Philosophy Compass* 3 (4):789-802

Kriegel, U. (2002). Phenomenal content. *Erkenntnis* 57 (2):175-198

———(2008). Real narrow content. *Mind and Language* 23 (3):304–328

———(2012). Personal-level representation. *Protosociology* 28:77-114

Kripke, S. (1979). A puzzle about belief. A. Margalit (ed.), *Meaning and use*. Reidel

———(1980). *Naming and necessity*. Harvard University Press

Larson, R. & Segal, G. (1995). *Knowledge of meaning*. MIT Press

Laurence, S. (1996). A Chomskian alternative to convention-based semantics. *Mind* 105 (418):269-301

Laurence, S. & Margolis, E. (1999). *Concepts: core readings*. MIT Press

_____(2007). The ontology of concepts: abstract objects or mental representations? *Noûs* 41 (4):561-593

Lepore, E. & Loewer, B. (1986). Solipsistic semantics. *Midwest Studies in Philosophy* 10 (1):595-614

Lewis, D. (1970). General semantics. *Synthese* 22 (1-2):pp.18-67

_____(1975). Languages and language. In Keith Gunderson (ed.), *Minnesota studies in the philosophy of science*. University of Minnesota Press

_____(1979). Attitudes de dicto and de se. *Philosophical Review* 88 (4):513-543

_____(1984). Putnam's paradox. *Australasian Journal of Philosophy* 62 (3):221-236

Loar, B. (1982). Conceptual role and truth conditions. *Notre Dame Journal of Formal Logic* 23 (July):272-83

_____(1988). Social content and psychological content. In Rosenthal, D. (1991) *The nature of mind*. Oxford: Oxford University Press

_____(2003). Phenomenal intentionality as the basis of mental content. In Martin Hahn & B. Ramberg (eds.), *Reflections and replies: essays on the philosophy of Tyler Burge*. MIT Press. 229-258

Ludlow, P. (2006). The myth of human language. *Croatian Journal of Philosophy* 6 (3):385-400

Lycan, W.G. (2001). The case for phenomenal externalism. *Philosophical Perspectives* 15 (s15):17-35

Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science* 72 (3):444-467

Marr, D. (1982). *Vision.* Freeman

McDowell, J. (2006). Conceptual capacities in perception. In G. Abel (ed.), *Kreativität*. Felix Meiner Verlag (2006)

McGinn, C. (1990). *Mental content*. Blackwell

Medin, D. & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. New York: Cambridge University Press

Mendola, J. (2008). *Anti-externalism*. Oxford University Press

Millikan, R. (1989). Biosemantics. *Journal of Philosophy* 86, pp.281-297

Newman, A. (2005). Two grades of internalism (pass and fail). *Philosophical Studies* 122 (2):153-169

Papineau, D. (1993). *Philosophical Naturalism*. Blackwell

_____(2006). Phenomenal and perceptual concepts. In Torin Alter & Sven Walter (eds.), *Phenomenal concepts and phenomenal knowledge: new essays on consciousness and physicalism*. Oxford University Press. 111-144

Peacocke, C. (1992). *A study of concepts*. MIT Press

Perry, J. (1979). The problem of the essential indexical. *Noûs* 13 (December):3-21

Pietroski, P. (2010). Concepts, meanings, and truth: first nature, second nature, and hard work. *Mind and Language* 25: pp.247-78

Prinz, J. (2000). The duality of content. *Philosophical Studies* 100 (1):1-34

_____(2002). *Furnishing the mind: concepts and their perceptual basis*. MIT Press

_____(2005). The return of concept empiricism. In H. Cohen & C. Leferbvre (eds.), *Categorization and cognitive science*. Elsevier

Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7 pp.131-193

Recanati, F. (1993). *Direct reference*. Blackwell

_____(2000). Deferential concepts. *Mind & Language*. 15 (4) pp.452–464

_____(2002). Unarticulated constituents. *Linguistics and Philosophy* 25:299-345

_____(2007). *Perspectival thought: a plea for (moderate) relativism*. Clarendon Press

_____(2010). *Truth-conditional pragmatics.* Oxford University Press

_____(2012). *Mental files*. Oxford University Press

Richard, M. (2006a). Meaning and attitude ascriptions. *Philosophical Studies* 128 (3):683-709

_____(2006b). Propositional attitude ascriptions. In Devitt, M. & Hanley, R. (eds.) (2006) *The Blackwell guide to the philosophy of language*. Blackwell Publishing

Russell, B. (1905). On denoting. *Mind* 14 (56):479-493

Salmon, N. (1986). *Frege's puzzle*. MIT Press

Sawyer, S. (2003). Conceptual errors and social externalism. *Philosophical Quarterly* 53 (211):265-273

_____(2007). There is no viable notion of narrow content. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), *Contemporary debates in philosophy of mind*. Blackwell

Schroeter, L. (2004). The rationalist foundations of Chalmers's 2-D semantics. *Philosophical Studies* 118 (1-2):227-255

Searle, J. (1983). *Intentionality: an essay in the philosophy of mind*. Cambridge University Press

_____(2007). What is language: some preliminary remarks. In Savas L. Tsohatzidis (ed.), *John Searle's philosophy of language: force, meaning, and mind*. Cambridge University Press

Segal, G. (1989). Seeing what is not there. *Philosophical Review* 97 (April):189-214

_____(1991). In defence of a reasonable individualism. *Mind* 100 (399):485-94

_____(2000). *A slim book about narrow content*. MIT Press

_____(2004). Reference, causal powers, externalist intuitions, and unicorns. In Richard Schantz (ed.), *The externalist challenge*. De Gruyter

_____(2007). Cognitive content and propositional attitude attributions. In Brian P McLaughlin & Jonathan D. Cohen (eds.), *Contemporary debates in philosophy of mind*. Blackwell

_____(2009a). Narrow content. In Brian McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *The Oxford handbook of philosophy of mind*. Oxford University Press

_____(2009b). Keep making sense. *Synthese* 170 (2):275-287

Shoemaker, S. (2006). On the way things appear. In John Hawthorne (ed.), *Perceptual experience*. Oxford University Press

Soames, S. (2002). *Beyond rigidity: the unfinished semantic agenda of naming and necessity*. Oxford University Press

_____(2005). *Reference and description: the case against two-dimensionalism*. Princeton: Princeton University Press

Sprevak, M. (2010). Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science* (2010), 41, 260-270

Stainton, R. (2006). Meaning and reference: some Chomskian themes. In E. Lepore & B. Smith (eds.), *Oxford handbook of the philosophy of language*. Oxford University Press

Stalnaker, R. (1989). On what's in the head. *Philosophical perspectives* 3: 287-319

_____(1990). Narrow content. In C. Anthony Anderson & Joseph Owens (eds.), *Propositional attitudes: the role of content in logic, language, and mind*. Stanford: Csli

_____(1999). *Context and content*. Oxford University Press

_____(2001). On considering a possible world as actual. *Aristotelian Society Supplementary Volume* 75 (75):141-156

_____(2004). Assertion revisited: on the interpretation of two-dimensional modal semantics. *Philosophical Studies* 118 (1-2):299-322

_____(2008). *Our knowledge of the internal world*. Oxford University Press

Travis, C. (2008). Are belief reports opaque? In *Occasion sensitivity: selected papers* (2008). Oxford University Press

Tye, M. (2000). *Consciousness, color, and content*. MIT Press

Vaughan, R. (1989). Searle's narrow content. *Ratio* 2 (2):185-90

Vicente, A. (2012). Burge on representation and biological function. *Thought* 1 (2):125-133

Vicente, A. & Manrique-Martinez, F. (2014). The big concepts paper: a defence of hybridism. *British Journal for the Philosophy of Science*

Wakefield, J. (2002). Broad versus narrow content in the explanation of action: Fodor on Frege cases. *Philosophical Psychology* 15 (2):119-33

Weiskopf, D. (2007). Atomism, pluralism, and conceptual content. *Philosophy and Phenomenological Research* 79 (1):131-163

_____(2009). The plurality of concepts. *Synthese* 169 (1)

White, S. (1982) Partial character and the language of thought. *Pacific Philosophical Quarterly* 63 (October):347-65

Wikforss, Å. (2001). Social externalism and conceptual errors. *Philosophical Quarterly* 51 (203):217-31

_____(2004). Externalism and incomplete understanding. *Philosophical Quarterly* 54 (215):287-294

_____(2005). Naming natural kinds. *Synthese* 145 (1):65-87

_____(2007). Semantic externalism and psychological externalism. *Philosophy Compass* 3 (1):158-181

_____(2012). Color terms and semantic externalism. *Croatian Journal of Philosophy* 12 (3):399-420

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press

Wilson, R. (1994). Wide computationalism. *Mind* 103 (411) pp.351-72

Yablo, S. (1997). Wide causation. *Philosophical Perspectives* 11 (11):251-281

_____(2003). Causal relevance. *Philosophical Issues* 13 (1):316-28

Zemach, E. (1976). Putnam's theory on the reference of substance terms. *Journal of Philosophy* 73 (March):116-27