
THE UNIVERSITY *of* EDINBURGH

# Expressivism, Normative Content, and Propositions

James L. D. Brown

Doctorate of Philosophy
University of Edinburgh

For my parents

And for Sophie, of course

## Abstract

The thesis of this thesis is that expressivists can and should develop a theory of normative propositions that can play an explanatory role in their theory of normative thought and discourse. It has been widely assumed that expressivists cannot make explanatory appeal to normative propositions because propositional content is representational in the following sense: a proposition is or determines a way that reality must be when that proposition is true. If a normative proposition is or determines a way reality must be when that proposition is true, and believing a proposition involves taking that proposition to be true, it follows that believing a normative proposition involves taking reality to be some particular way. But expressivists deny that normative thought represents reality in this way. Rather, normative thought is best explained as having some distinctive nonrepresentational function, such as motivating our actions and coordinating our attitudes. As such, almost all expressivists have rejected the existence of normative propositions, except in a deflationary and hence non-explanatory sense.

However, by rejecting the existence of normative propositions, expressivists face a number of serious problems in relation to explaining normative thought and discourse. By positing propositions as the objects of our attitudes and speech acts, we can provide a straightforward and systematic characterisation of our thought and talk in terms of the things we are related to in believing, desiring, asserting, denying, and so on. For example, logically complex attitudes can be explained in terms of the logical complexity of their propositional content. Rational connections between attitudes, such as inconsistency and entailment relations between beliefs, can be explained in terms of the properties of their propositional objects. Different attitude types with the same content can be explained in terms of a subject's standing in different relations to the same proposition. And quantification over attitude contents can be explained in terms of a domain of propositional objects over which such quantification occurs. By rejecting normative propositions, expressivists must provide alternative explanations of these features of normative thought and discourse.

Although a number of philosophers have attempted to provide such an alternative, such attempts face a number of serious difficulties. Moreover, even supposing some adequate alternative is forthcoming, there is a remaining problem of explaining why it is that both normative and non-normative thought possess many of the same features but for completely different reasons. Problems such as these have more recently led some to suggest that expressivists should embrace the existence of normative propositions within

their theory of normative thought and discourse. This thesis takes up this idea and examines a number of different frameworks in which expressivists might develop a theory of normative propositions.

How could an expressivist ever fully embrace the existence of normative propositions? My simple answer is that expressivists should reject the assumption that normative propositions are representational. If a normative proposition is not or does not determine a way reality must be when it is true, then believing that proposition need not involve representing reality as being some particular way or other. I explore several different views about propositions and argue that they each admit of generalisation such that some but not all propositions are representational. I argue that some of these views are better than others for the purposes of expressivism. My discussion of propositions is also meant to contribute to theorising in the philosophy of mind and language about what propositions are. For if you think there might be something right about the expressivist idea that normative thought is not representational like descriptive thought, then you should think that some views of propositions are better than others in light of my arguments about why expressivists can and should embrace a nonrepresentational view about normative propositions.

## Lay Summary

On the 27 March 2020, the UK Government advisor Dominic Cummings travelled with his family from his London home to stay with his parents while he and his wife were ill with Covid-19. This trip, purportedly made so that their child could receive adequate care if both parents were incapacitated from the illness, was made at the height of a government-imposed lockdown that forbade all but essential travel. Given his senior role in advising the government, especially in relation to their communications strategy during a public health crisis, Dominic Cummings' actions caused widespread debate and outrage. Although the trip, or at least certain parts of it, seemed to have gone against the spirit and perhaps the letter of the government guidance, he has denied any wrongdoing. Many see things otherwise.

Given the risk of travelling across the country in full knowledge of his wife having Covid-19, it has been claimed that he was *wrong* to take this journey. Moreover, it has been claimed that it was *unfair* for him to make an exception of his own circumstances, when others have made far greater sacrifices in order to stick to the rules. Further, it has been claimed that he *ought* to feel some guilt or regret about his actions, and that the fact that there seems to be one rule for those in charge and another rule for everyone else is a *reason* to feel indignation. And it has also been claimed that we *should not* believe his version of the events, given all the available evidence. (Is it really plausible that he drove his family to a tourist hotspot to "test his eyesight" on a day that just so happened to also be his wife's birthday?) These are just some of the issues that are at the time of writing being widely discussed in the media and homes of the UK.

All of these claims are what philosophers call *normative* claims. They are claims about how we ought to act, what feelings and emotions we ought to have, and what we ought to believe. Centrally, they are claims that are in some way or another about *justification*. In claiming that one ought to feel indignation, one offers a justification for that attitude. In claiming that it was wrong to have taken the journey, one asserts that the action lacked justification. In answering the big questions about how we should live, we are making normative judgments. But they are also the stuff of everyday. (Should I spend my evening catching up on work or reading?). However, as well as asking such questions and forming answers to them, we can also take a reflective step back and ask: What are we doing when we ask normative questions? What are we doing when we make normative claims? These are *meta-ethical* or *meta-normative* questions, and they make up the topic of this thesis.

Contemporary meta-ethical views typically fall into one of two opposing camps. According to *descriptive* views, normative claims describe how things

are. On this view, the claim that Dominic's journey was unjustified is fundamentally no different to the claim that the journey took place on 27 March. Each claim describes the journey as having a certain feature, and in making such claims we are asserting that the journey did *in fact* have those features. The difference between the two claims, according to the descriptive view, is simply in the features we describe the journey as having. By contrast, according to *expressive* views, normative claims do not describe how things are. Instead, they express our attitudes and feelings. So, for example, we might think that the claim that Dominic's journey was unjustified expresses *disapproval* of his journey, in contrast to describing it as having some particular feature.

The expressive view has a lot going for it. For instance, when deliberating about whether I ought to perform some action, it seems that I am not simply trying to ascertain whether some particular fact is or is not the case. Rather, I am trying to settle the practical question of what to do. If normative claims are simply descriptions of how things are, as the descriptive view maintains, then this feature of normative deliberation looks puzzling. However, if normative claims express attitudes for and against various things, then this feature of normative deliberation can be explained straightforwardly. Further, if our normative claims are descriptions of how things are, then this raises questions about the nature of the facts so described. Specifically, it raises the question of how to make sense of facts about what is wrong, justified, required, and so on within a modern scientific worldview. Indeed, it raises the question of whether any such facts really exist. The question arises because the natural sciences tell us only about how things are, not how things ought to be, and so 'normative facts' do not seem to be recognised by the sciences. However, if our normative claims do not describe such facts, as the expressive view maintains, then no difficult questions arise about the place of such facts within a broader worldview.

But the expressive view has its problems as well. For instance, consider the claim that *if* Dominic's journey did not hurt anyone, *then* it was not wrong. Assuming the expressive picture, it is unclear what attitude this claim expresses. Because the claim about what is wrong appears in what is sometimes called an 'embedded context', here the consequent of a conditional claim, it does not seem to express disapproval (or some other negative attitude). In asserting the claim, one only commits to the conditional; unless one also accepts the antecedent claim that the journey did not hurt anyone, one need not express disapproval of anything. So if, as the expressive view maintains, normative claims express our attitudes, then we are owed an account of what attitudes are expressed by embedded normative claims. The

so-called 'embedding problem' gives rise to a number of related but more technical problems for the expressive view.

This thesis examines several theoretical frameworks for addressing these problems. Specifically, it aims to develop a theory of *normative belief* and a theory of *what normative beliefs are about*, or what philosophers call *propositions*, that are compatible with the expressive view. Traditionally, philosophers have maintained that belief is a descriptive attitude in the sense that believing some claim involves taking some description to be true. For example, this view holds that my believing that Dominic's journey was unjustified involves my taking a certain description to be true, namely that the description "was unjustified" is true of Dominic's journey. As such, it has generally been assumed that the expressive view must deny that claims about what we ought to do express beliefs. However, in this thesis I argue that proponents of the expressive view should reject this assumption. I do this by exploring several 'non-descriptive' ways in which to understand what it is to believe something. I argue that the resulting picture provides the most promising version of the expressive view, avoiding the problems mentioned above, as well as providing an illuminating and interesting theory of what we are doing when we make normative claims.

## Acknowledgments

Michael Dummett once wrote that he was always disappointed when a book lacked a preface. To him, it was like turning up to a dinner party and being guided straight to the dining room. As much as I enjoy dinner parties, I've never been one for hosting them. So I hope the reader will not be disappointed in my not offering any preface to this thesis. However, it would be wrong not to give due acknowledgment to the many people who have in one way or another contributed to the final product.

Initial thanks are due to my supervisors, Matthew Chrisman and Mike Ridge. Since arriving at Edinburgh, they have always treated me as a colleague rather than a student, and the confidence given to me through their encouragement cannot be understated. I moved to Edinburgh in large part to work with Matthew and Mike, and it has been a privilege to work with two philosophers who have done so much to shape my understanding of expressivism and metaethics. Thanks to Mike especially for the generosity of his feedback. And thanks to Matthew for the many accumulated hours discussing everything in this thesis and much else besides, and especially for always making me see the significance of my work in relation to the bigger picture.

Prior to Edinburgh, I started thinking about these issues while at University College London completing my MPhilStud. Thanks are due to my then supervisor Mark Kalderon for teaching me how to receive criticism, as well as the importance of not passively accepting the philosophical status quo. Special thanks are also due to Giles Howdle, James Laing, Debbie Roberts, and Silvan Wittwer for extensive comments and conversation on various aspects of the project, and especially to James for always pressing me to defend the value of such a project. In addition, I'd like to thank the following people for comments or conversation on material that in one way or another ended up in this thesis: Fausto Carcassi, Jessica Isserow, Silvia Jonas, Luke Kerston, Jiwon Kim, Sebastian Köhler, Nicholas Laskowski, James Lewis, David Plunkett, Andrés Soria Ruiz, Saranga Sudarshan, Jussi Suikkanen, Christine Tiefensee, Teemu Toppinen, Herman Veluwenkamp, and Jack Woods. This thesis also benefitted from feedback and questions from audiences at: Arché at the University of St Andrews; the Mind, Action, and Language Group at the University of Porto; the University of Edinburgh; and the University of Manchester.

Certain intellectual debts also ought to be acknowledged. Anyone familiar with Mark Schroeder's work on expressivism and propositions will be able to see how much I owe to his framing of the terms of the debate. And as will be clear to many, the expressivist picture endorsed in the later chapters of this

thesis owes much to Allan Gibbard's work on expressivism, as well as to Seth Yalcin's more recent developments of Gibbard's work. However, my biggest intellectual debt is to Simon Blackburn for first making me see the value and importance of the expressivist programme.

Finally, my biggest debts are owed to my parents, Keiron Brown and Sarah Jameson, and to my partner, Sophie Potter. My parents have been nothing but supportive of my best attempts to never leave the education system, despite probably having only a vague idea about what I do. But no one has been as supportive and encouraging, in all senses of those words, as Sophie, who, despite our on-off living in different countries over the past few years, has truly been there with me through everything. It is to her I owe the most.

This research was supported by an Arts and Humanities Research Council DTP Studentship with the Scottish Graduate School of Arts and Humanities, as well as a bursary from the Royal Institute for Philosophy. Some of the material from Chapters 2 and 3 was originally published as 'Expressivism and Cognitive Propositions' in the *Journal of the American Philosophical Association*, and some of the material from Chapters 2 and 6 derives from ideas developed in my MPhilStud thesis *Propositions and Nondescriptivism in Metaethics* (University College London).

# Table of Contents

# 1
# Introduction

What is the nature of normative thought and discourse, or thought and talk about what we ought to do, think, and feel? Traditionally, metaethical expressivists have denied that normative thought and discourse is primarily about whether particular propositions are true. Instead, they claim, it is about influencing each other's conative and affective attitudes. Accordingly, such a view has no place for normative propositions within an explanatory account of normative thought and discourse. However, eschewing normative propositions creates all sorts of difficulties for expressivists. A straightforward way to resolve these difficulties is for expressivists to embrace normative propositions. This thesis argues that expressivists can and should develop a theory of normative propositions to play an explanatory role in their theory of normative thought and discourse. What expressivists need, I argue, is a conception of normative propositions consistent with the idea that normative discourse is importantly distinctive from descriptive thought and discourse, or thought and talk about the way the world is. In this introductory chapter, I lay out the conceptual and theoretical underpinnings of the project. First, I motivate the project in the most general terms, situating it within the existing metaethical landscape (section 1.1). Next, I outline the theoretical background that frames the discussion of the thesis (section 1.2). Then, I explain some of the more detailed motivations for positing propositions and the problems that result from rejecting them (section 1.3). Finally, I provide a roadmap of the remaining chapters (section 1.4).

## 1.1 Two metaethical pictures

This thesis is about normative thought and discourse. What exactly is this? As I am using the term, normative thought and discourse is thought and talk that is in some way or another about the *justification* of our actions, attitudes, and beliefs. Suppose I break a promise to a friend in order to help a stranger. Was I *wrong* to do so? Did my *reasons* to help the stranger outweigh my reasons to keep my promise? Given my evidence at the time, *should* I have believed that the stranger needed help? *Ought* I to feel regret or guilt at my having broken

my word? These are all normative questions, and thought and talk concerning these questions is normative thought and discourse.

Contemporary metaethics is dominated by two competing pictures of our normative practices. According to the first picture, normative deliberation and discussion is just one part of the more general activity of inquiring about the world around us. When we make a normative judgment, we represent a way that reality might be. What makes normative judgments distinctive, this picture claims, is that they concern putative facts about the existence and distribution of normative properties and relations. This is in contrast, say, to facts about the location of physical objects in an environment or about its natural history. Call this the *representationalist* picture of normative thought and discourse.

Pressure is placed on the representationalist picture when we begin to ask about the nature of the facts that normative inquiry aims to discover. Specifically, there is a challenge to explain how normative facts, properties, and so on fit into the world as described by the natural sciences. Relatedly, there is also a challenge to explain how natural beings like ourselves acquire epistemic and semantic access to these facts. Further, there is a challenge to explain the practical significance of normativity. Normativity is something we care deeply about. It guides us in how to live. It matters. But if normative inquiry is just the investigation of one kind of fact among others, we are owed an explanation of why these facts seem to have a distinctive kind of authority in determining how to live.

Despite these challenges, the representationalist picture remains compelling for many philosophers. One of the central tasks for representationalists is to answer the challenges in such a way that vindicates our ordinary normative practices and beliefs. Many philosophers are optimistic that this task can be met, and there are many competing theories that purport to provide such vindication. Others, however, remain sceptical. These sceptics can be placed in one of two camps. First, there are those who accept the representationalist picture. According to this group, our normative judgments are correctly understood to be about the existence and distribution of normative properties and relations. However, as a matter of fact, there *are* no normative properties or relations, or at least none that are instantiated or obtain. As such, our normative practices and beliefs are fundamentally premised on an error. By contrast, the other group of sceptics see no such error in our normative practices. The error, these sceptics claim, resides in a false philosophical picture of the nature of normative thought and discourse. According to these philosophers, we need another picture.

This leads us to the second prominent picture of normative thought and discourse. According to this second picture, normative deliberation and discussion should not be understood as part of the more general activity of inquiring about the world around us. Rather, at its core, normative deliberation and discussion is *practical* deliberation and discussion about what to do, what to think, and what to feel. Our normative judgments aim to settle how to live, not simply how things stand in the world. Proponents of this second picture argue that it avoids the challenges that beset the representationalist picture. For instance, the practical significance of normativity can be explained in terms of the practical nature of normative practice itself rather than in terms of the part of reality it is about. Moreover, if normative judgments do not aim to represent normative facts, no puzzles arise about how such facts fit into our conception of the world and our place in it; such facts are simply absent in the explanation of normative thought and practice.

The dominant theoretical position that conforms to this second picture is *metaethical expressivism* (hereafter 'expressivism'). As I will be using the term, expressivism is broad church with a long history. At a general level, 'expressivism' can be used to carve out a cluster of different theories rather than any particular thesis or theory.[1] I propose to define an expressivist theory as any metaethical theory that endorses the following two theses: first, that normative thought and talk is fundamentally *nonrepresentational*; second, that normative thought and talk is fundamentally *practical*.[2] I will explicate these notions further in the next section. But the general idea should already be clear from the picture sketched above. As suggested there, expressivism provides an attractive view insofar as it explains the practical significance of normative discourse within a broadly naturalistic worldview while dissolving certain metaphysical and epistemological puzzles about normativity and avoiding error-theory.

Despite these attractions, many doubt that expressivism can successfully explain the semantic and logical properties, broadly construed, of normative thought and discourse (more on which below). The problems critics raise tend to lie in the details of the particular expressivist theory in question. But the general worry might be expressed as follows. Our best account of the semantic and logical properties of intentional thought and talk presupposes a

[1] Prominent expressivist theories include Blackburn (1984, 1993, 1998), Gibbard (1990, 2003), Horgan & Timmons (2006), Schroeder (2008), and (Ridge 2014); contemporary expressivism is a descendent of emotivist and prescriptivist theories such as Ogden and Richards (1923); Barnes (1933); Carnap (1935); Ayer (1936), Stevenson (1937), and Hare (1952).

[2] Compare Price's (2011) formulation of expressivism.

representationalist picture of such thought and talk. If we give up this picture, we give up our best means of explaining these features of normative thought and talk. Thus, any nonrepresentational picture of normative thought and discourse is going to have a difficult time explaining these features. Traditionally, expressivists have accepted the challenge on its own terms, accepting the received explanation of semantic and logical properties in the descriptive domain and providing alternative explanations in the normative domain. This thesis takes a different tack.

Rather than accepting the challenge on its own terms, I argue that our best account of the semantic and logical properties of intentional thought and discourse does *not* presuppose a representationalist picture. More specifically, I argue that a *propositionalist* account of the semantic and logical properties of thought and discourse — one that explains these properties in terms of the propositional content of our attitudes and speech acts — does not presuppose a representationalist picture. On this approach, what expressivists need is a reconceptualization of propositional contents and attitudes that is compatible with the expressivist picture. The undertaking of this task is the main project of this thesis. In this chapter, I argue that expressivists can *in principle* embrace normative propositions by denying that all propositions are representational (in a sense to be explained below). The remainder of the thesis then explores several different views about propositional content and argues that they each admit of generalisation such that some but not all propositions are representational. This, I argue, allows expressivists to embrace a propositionalist account of normative thought and discourse.

## 1.2 Theoretical background

With the basic landscape now sketched, I will now explain the key notions and arguments introduced in previous section. First, I will provide a more detailed characterisation of expressivism. Second, I will explain what I take to be the central theoretical role for propositions in terms of their role in the framework of *propositional attitude psychology*. Third, I will discuss some further issues that arise from this framework that will frame and constrain the subsequent chapters.

### *1.2.1 Expressivism characterised*

I have defined expressivism as the conjunction of two theses: first, that normative thought and discourse is nonrepresentational; second, that

normative thought and discourse is practical. I will now say something about these two theses. The negative thesis claims that normative thought and discourse does not represent or describe normative reality. Here, 'normative reality' is shorthand for normative facts, properties, relations, states of affairs, or whatever else might be included in one's normative ontology. This thesis variously goes by the name *nonfactualism*, *nondescriptivism*, and *nonrepresentationalism*.[3]

The positive thesis claims that normative thought and discourse has some kind of distinctive nonrepresentational function, such as motivating our actions and coordinating our affective reactions in various real and imagined circumstances. For the purposes of this thesis, I will remain relatively noncommittal about how exactly we should cash out the second thesis. This is for two reasons. First, it is because I want to be as inclusive as possible as to which expressivist theories might benefit from embracing normative propositions. Second, it is because, at least at the outset, it will not matter too much to the arguments discussed exactly how the second thesis is cashed out. So while I will often describe normative thought and discourse as having a 'directive' or 'practical' function, I will generally leave it open exactly how these notions should be cashed out. That said, I will at times adopt a more specific conception of the second thesis for the sake of concreteness in developing a theory or argument.

Throughout this thesis, I will treat 'factual', 'descriptive', and 'representational' as synonymous, and likewise for their respective cognates. Given the ubiquity of these terms in many different contexts, it is important to distinguish my use of these terms from other theoretical uses. For example, philosophers of mind sometimes debate whether our mental states are 'representational' in the sense that they constitutively involve concrete 'mental representations', where these are conceived as vehicles of semantic content realised in the brain.[4] This sense of 'representational' is orthogonal to the sense employed here. A mental state might be 'representational' in both, one, or neither senses. Second, philosophers of language sometimes debate whether denoting expressions are 'descriptive' in the sense that their denotation is fixed

[3] Throughout the thesis, I will often characterise the negative thesis in terms of normative properties and relations. This is because opponents of expressivism typically think in such terms. However, it's worth highlighting that the rejection of normative properties and relations as such is not what's at issue. As Devitt (2010: 139) points out, a nominalist might deny that normative claims are about the existence or distribution of normative properties, but this does not necessarily make her a nonfactualist in the relevant sense. So I take it that there is some reading of 'describing normative reality' available to the nominalist cum factualist that would put her in disagreement with the nonfactualist.

[4] See, for instance, Egan (2014) for a discussion of this kind of terminology.

by an associated description, rather than referring directly or by some other means. As I will be using the term, any expression whose meaning is explained in terms of its reference is descriptive.[5] Third, as I will be using the term, a 'factualist' view need not be cashed out in terms of an ontology of facts, rather than some other kind of ontology (e.g. objects or tropes). Rather than opt for a unified nomenclature, I will make use of each family of terms throughout. This is partly to avoid tedious repetition. But it is also because there is no agreed use of terminology in the literature and I will be engaging at different points with theorists who employ each kind of terminology.[6]

A further qualification about the negative thesis is that it should be understood as a theoretical claim about whether normative thought and discourse should be *explained* in terms of representing normative reality. This is because expressivists can coherently accept talk of 'representing' or 'describing' normative 'facts' as first-order claims *within* normative discourse, as opposed to meta-normative claims *about* normative discourse. This is possible insofar as the expressivist can provide deflationary interpretations of these notions. For example, a deflationary reading of 'fact' might claim that asserting [it is a fact that p] is more or less equivalent to asserting [p]. Such a view might claim that the utility of factual vocabulary in first-order discourse is its expressive power (e.g. in generalisations such as "everything she said was a fact"). Where 'p' is a normative claim, if (i) asserting [it is a fact that p] and asserting [p] are equivalent, and (ii) asserting [p] expresses a nonrepresentational state of mind, it follows that (iii) asserting [it is a fact that p] also expresses a nonrepresentational state of mind.[7] In this way, the expressivist can give a nonfactualist explanation of normative 'fact'-talk and similar notions.

---

[5] Thus, for instance, the 'nondescriptive' ideationalist theory of meaning proposed by Davis (2003) is still a descriptivist theory in the sense that the thoughts expressed by sentences have world-characterising truth-conditions.

[6] However, I do wish to avoid characterising expressivism in terms of the 'cognitive'-'noncognitive' distinction. This terminology is strongly associated with the notion of *belief*, where metaethical noncognitivism is the denial that there are normative beliefs. Because most contemporary expressivists hold that normative statements express beliefs in some sense or other, the noncognitivist label seems inappropriate (it would also exclude 'cognitivist expressivist' theories such as Horgan and Timmons [2006]). Moreover, a central theme of this thesis will be that expressivists can claim that normative statements express beliefs in propositions in exactly the same sense as descriptive statements express beliefs in propositions. Further, the term 'cognitive' is employed in chapter 2 in a broader sense of 'mental' or 'intentional'.

[7] Compare Blackburn's (1998: 78) metaphor of Ramsey's ladder.

This is somewhat oversimplified, and I don't mean to suggest that providing adequate deflationary interpretations is a trivial task. But if something like this is right, then expressivists should have no problem accommodating talk of normative facts at a first-order level. More generally, the task of accommodating the realist-sounding features of normative discourse is known as *quasi-realism* (see Blackburn 1993). It is perfectly consistent with this to claim that representational or factual notions play no *explanatory* role in accounting for normative thought and discourse. The point here is not to defend quasi-realism, but to get clearer on the commitments of the negative thesis, that normative thought and discourse is fundamentally nonrepresentational.

Finally, although this should already be clear, it is worth emphasising that my use of 'expressivism' differs to philosophers who define the term more narrowly. For example, Mark Kalderon defines expressivism as the thesis that "The content of a moral sentence wholly consists in the noncognitive attitude it conveys." (2007: 53) And Ralph Wedgwood defines it as the thesis that "the fundamental explanation of the meaning of normative statements […] is given in terms of the type of *mental state* that the statements made by uttering those sentences *express*." (2007: 35) As I am using the term, expressivism is not committed to either thesis. (I will talk about the significance of language in relation to expressivism below.) Although throughout the thesis I will examine particular expressivist theories, the general investigation will not be guided by any single conception of how expressivism should be developed. Moreover, although certain theories of propositions might lend themselves more naturally to certain expressivist theories, I hope that the resources developed in the proceeding chapters can be utilised by a wide variety of expressivist theories.

### *1.2.2 Propositional attitude psychology*

As it is used throughout this thesis, 'proposition' is a theoretical term of art. To understand it, we need to know what theoretical role the notion answers to. As I will be using the term, propositions are first and foremost the objects of attitudes. But what exactly is it to be the object of an attitude? I propose to understand this notion in terms of the broader framework of *propositional attitude psychology*. I will now say something about this broader framework and how this gives content to the notion of being the object of an attitude. Although metaethical debates often emphasise the theoretical importance of propositions, they rarely spell out in any detail what they take the role of propositions to be, and so it will be worth dwelling on this question in some detail.

Many philosophers believe that we can provide a systematic characterisation of a large subset of our mental states, speech acts, and actions by relating agents in various ways to entities called 'propositions' or 'propositional contents'. According to this approach, where *V* is a propositional attitude verb, *S* a subject, and *p* a proposition, attitude attributions take the following form:

$V(S, p)$

In other words, *S* stands in relation *V* to *p*. For example, suppose James believes that veganism is morally required. On the current approach, we can understand this attribution to involve ascribing a relation (*believing*) between a subject (James) and a proposition (that veganism is morally required), or:

*Believing*(James, [that veganism is morally required])

(I'll use the square bracket notation '[…]' throughout to indicate that the sentence or sentence variable within the brackets denotes a proposition.) Believing is a kind of propositional attitude, where propositional attitudes are those mental states that can be individuated by relating a subject to a proposition. Similarly, propositional acts are those cognitive or behavioural acts that can be individuated by relating subjects to a proposition, such as judging [that veganism is morally required] or asserting [that veganism is morally required].

As well as providing a systematic way of categorising our thoughts and speech acts, propositional attitude psychology has further explanatory aims. Specifically, it aims to explain the characteristic ways in which our attitudes stand to each other and interact. For example, suppose that *A* is the attitude of believing [that veganism is morally required] and *B* is the attitude of believing [that veganism is not morally required]. Propositional attitude psychology explains the relations between *A* and *B* in terms of the proposition that is the object of each attitude and the relation taken towards it. One such relation that needs explaining is the fact that *A* and *B* are inconsistent. Propositional attitude psychology explains this fact in terms of: (a) the more general fact that it is inconsistent to believe incompatible propositions; (b) the fact that *A* and *B* are beliefs; and (c) the fact that the proposition [that veganism is morally required] is incompatible with the proposition [that veganism is not morally required]. Further, propositional attitude psychology can appeal to (a)-(c) to explain why subjects who have both *A* and *B* will tend to revise their attitudes when this inconsistency is brought to light. Not only this, but the attitudes we attribute

to subjects will provide the basis for rationalising their actions more generally. These explanations are somewhat programmatic as they employ various notions that are in need of further elucidation. But they nonetheless form the basis of an explanation.

As I am using the term, 'propositional attitude psychology' is any approach to theorising about the mind that individuates and identifies mental states and speech acts by relating subjects to propositions. As with expressivism, this characterisation is deliberately vague and may be fleshed out in many different ways. For instance, we have said nothing yet about the nature of the relevant *relations* between subjects and propositions, the nature of *propositions*, or the nature of the *subjects* of propositional attitudes. We have said nothing about the empirical nature of our actual propositional attitudes and little about their metaphysical nature, other than that they are subject to the kind of individuation conditions given above. We have said little of the explanatory ambitions of propositional attitude psychology other than its aim to provide a systematic characterisation of certain aspects of our mental and behavioural lives. We have said nothing about the relation between the things we believe and the meanings of our words and sentences. And we have said nothing about the relation between propositional attitude psychology and our folk psychological practices. As a general framework, propositional attitude psychology is compatible with a number of different answers to these questions. The important point is simply that we can provide a systematic explanation of a large part of what we think, say, and do by relating subjects to propositions in appropriate ways.

What, then, must propositions be like in order to play the object of attitude role in propositional attitude psychology? Despite the many differences between different conceptions of propositions, there are a number of features of propositions that almost all theories agree on. First, propositions are taken to be abstract objects, in the sense that they do not inhabit a spatial or temporal location. Second, propositions are taken to be mind-independent, in the sense that a proposition can exist even if no one has ever believed it or taken any other attitude towards it. Third, propositions are taken to be language-independent, in the sense that two speakers might assert the same proposition using two different languages. And fourth, propositions have their truth-conditions essentially, in the sense that it is necessary that the proposition [p] is true just in case p and false just in case not-p. I will assume throughout that any theory of propositions will accommodate these features.

A widespread assumption about what propositions must be like that I do wish to challenge is what might be called the *representationalist assumption* about propositions. This terminology is offered as part and parcel of the

terminology described in the previous subsection. We can give a more precise characterisation of what it is for a proposition to be representational as follows: a proposition is or determines a way that reality must be when that proposition is true.[8] As before, I will use 'representational', 'descriptive', and 'factual' interchangeably, treating them as synonymous. And again, it is important to distinguish these uses from other uses of these terms as applied to propositions. For example, in certain debates about the metaphysics of propositions, the question of whether a proposition is 'representational' concerns whether it has its truth-conditions in virtue of its intrinsic representational properties, or whether it simply is or determines its truth-conditions without inherently representing anything.[9] This is not the sense of 'representational' in play here. As I will use the term, *any* proposition that determines world-characterising truth-conditions is representational, regardless of how it does so.

If the representationalist assumption is correct, then one might argue for the incompatibility of expressivism and normative propositions as follows. If a normative proposition determines a way reality must be when that proposition is true, and believing a proposition involves (inter alia) taking that proposition to be true, it follows that believing a normative proposition involves taking reality to be some particular way. But this contradicts the expressivist's negative thesis that normative thought does not represent or describe normative reality. So, the argument concludes, expressivists cannot appeal to normative propositions. Arguably, historic and widespread acceptance of the representationalist assumption explains why expressivists and their opponents have taken expressivism to entail a rejection of normative propositions (other than in a deflationary and hence non-explanatory sense).[10] However, this argument is sound only if we accept the representationalist assumption. If we reject the assumption, then it is open to expressivists to embrace normative

[8] Schroeder (2015: 2) frames in the issue in terms of whether propositions determine their own truth-conditions. If by 'truth-conditions' we mean something like *substantive* or *world-characterising* truth-conditions, then I think this is equivalent to the present framing. However, one might also mean something like *minimal* or *deflationary* truth-conditions, which all propositions plausibly do determine. Given a deflationary reading, an expressivist theory of propositions can respect the claim that propositions have their truth-conditions essentially.

[9] Examples of what might be called the *inherent representation* view include Hanks and Soames' respective views of propositions, discussed in the next chapter. For an example of the denial of the inherent representation view that is still 'representational' in the sense employed throughout this thesis, see Stalnaker (2012).

[10] Schroeder (2013: 86ff) offers some further explanations for the pervasiveness of the representationalist assumption.

propositions. Of course, if propositions are not representational, then we are owed an account of what they *are* like. More specifically, we are owed an account of what they are like such that they can realise the object of attitude role. Nonetheless, in principle, there is no reason why the expressivist cannot reject the assumption.

Some philosophers might object here. Isn't it a conceptual truth about propositions that they represent reality? One way to respond to this worry would be to provide counterexamples, i.e. to provide candidate examples of propositions that do not represent reality. Providing such examples will be a main objective of this thesis. However, we can also quickly say something more general in response to the worry. As I am using the term, 'proposition' is a theoretical term that is defined functionally.[11] Specifically, it is defined as whatever realises the object-of-attitude role in propositional attitude psychology. The constraints placed on the kind of entity that propositions can be will therefore be determined by the particular role that propositions play in one's overall theory. The significance of this is that if the expressivist denies that certain attitudes are representational in the relevant sense, but accepts that there are good grounds to individuate these attitudes in terms of their distinctive contents (more on this below), then we would not expect the appeal to propositions in and of itself to commit her to the representationalist assumption.

In fact, the above argument against the compatibility of expressivism and normative propositions can be resisted without rejecting the representationalist assumption about propositions. In other words, it is perfectly coherent to hold a nonrepresentationalist view of normative thought and discourse and a representationalist view of normative propositions. This is because even if one accepts that normative propositions correspond to distinctions in reality, one might deny believing a normative proposition consists in taking that proposition to be true. This is, in essence, the view of Kalderon (2007), who argues that accepting a moral proposition consists in having a noncognitive attitude towards that proposition. I will discuss some more specific reasons for rejecting this view in the next chapter, but I think that rejecting the representational assumption is a generally more attractive approach. So although a representationalist cum expressivist theory of normative propositions is possible, throughout this thesis I will take 'an expressivist theory of propositions' to refer only to nonrepresentationalist theory of propositions.

In this thesis, I explore, develop, critique, and defend a number of different frameworks that expressivists might utilise to develop a theory of normative

[11] Compare Yalcin (2014: 19).

propositional content. I will argue that certain approaches have serious limitations and that other approaches are more promising. In both cases, however, the inquiry will show that the range of options available to expressivists to develop a theory of normative propositions is far wider than one might expect. The approach of this thesis will not be to comprehensively survey all possible approaches, but to focus on a few promising approaches and see how far we can develop these theories. So my discussion of propositions is also meant to contribute to theorising in the philosophy of mind and language about what propositions are. In the end, I will argue that developing an expressivist theory of propositions requires us to rethink not just the nature of propositional content but also the nature of propositional attitudes more generally. So while the project undertaken in this thesis is primarily in the service of expressivism, if correct the view defended here has wide-reaching consequences for how we think about what we think, say, and do in general. Before examining the motivations for this approach, I want to highlight some further issues that will shape the discussion to come.

### *1.2.3 Further issues*

Now that I've provided an initial characterisation of expressivism and the theoretical role of propositions in propositional attitude psychology, I want to highlight a number of further issues that arise from these characterisations that will shape the subsequent discussion.

#### *Thought and language*

The first issue concerns the relation between thought and language. The current investigation is intended primarily as a contribution to theorising about thought and mental content. It is not intended, at least in the first instance, as a contribution to theorising about language and linguistic meaning. This has implications for how to think about propositions as well as how to think about expressivism.

Recall that propositional attitude psychology posits propositions as the objects of propositional attitude and acts, including speech acts. Many theorists believe that the same entities should also play the role of the meanings or semantic values of sentences (in context). However, although I say nothing to rule out this possibility, I will not make any assumptions about the relation between the meaning of a sentence and the content of an assertion that deploys that sentence. For in any case, there are a number of reasons for thinking that

the *object of attitude role* and the *semantic value of sentences in context role* serve different theoretical purposes and as such are subject to different constraints.

To name just one, it is not clear that propositional contents are compositional in the same way as semantic values are compositional (see, for example, Dummett 1973; Lewis 1980; Stanley 1997; Ninan 2010; Rabern 2012; Yalcin 2014). For example, consider the two sentences (from Rabern 2012):

(1) "It is raining."
(2) "It is raining now."

Plausibly, asserting these two sentences say the same thing relative to the same context, i.e., they have the same propositional content. Similarly, if I believe [that it is raining], and I believe [that it is raining now], plausibly these come to one and the same belief. However, embedding (1) and (2) in more complex sentences shows that they have different meanings relative to the same context. For example:

(3) "It has been the case that it is raining."
(4) "It has been the case that it is raining now."

Relative to the same context of utterance, these sentences might have different truth-values. Whereas the truth-value of (3) depends on what has happened before the time of utterance, the truth-value of (4) depends on what is happening at the time of utterance. However, if two sentences have the same meaning, then substituting each sentence should preserve truth-conditions. Thus, while (1) and (2) express the same proposition relative to the same context, they have different semantic values. Similar examples abound.[12]

The point I want to make here is simply that a theory of propositions need not be a theory of sentence meanings (in context). These are two distinct notions answering to different theoretical demands. Of course, a comprehensive theory of normative thought and discourse will include a theory of the meanings of normative terms and sentences as well as the objects of normative beliefs and assertions. It will also provide a 'postsemantics' that explains the connection between the theory of meaning and the theory of content. For example, a straightforward view might be that the meaning of a sentence in context determines the proposition expressed by its utterance, where 'determines' expresses some relation other than identity (e.g. Lewis

[12] See especially Yalcin (2014: 23ff) for further conceptual and empirical reasons against the identification of belief content and compositional semantic value.

1980). However, answering such questions is beyond the purview of this thesis.[13]

Given my focus on expressivism, it might initially seem surprising that I will not focus on language, since expressivism is often characterised primarily as a thesis about normative language. For example, it is often characterised as the view that the meaning of a normative term or sentence is explained in terms of the nonrepresentational state of mind it conventionally expresses (the canonical statement is Gibbard [2003: 7]; see also the passages quoted above from Kalderon and Wedgwood). However, I think that it is a mistake to see metaethical expressivism primarily as a view about normative language. Although expressivists often motivate their theory from observations about normative language, the explanations that expressivists aim to provide are fundamentally *psychological* rather than linguistic. When expressivists offer explanations of the practical significance of normative discourse in our lives, or of how to dissolve metaphysical and epistemological puzzles about normativity, they do so by offering explanations of what it is to *think* normative thoughts and to *have* normative attitudes. Thus, expressivists ultimately address metaethical problems with psychological explanations, not linguistic ones.

This is not to deny that we live much of our normative lives through language. Indeed, I find it plausible that language possession is a necessary condition for the capacity for normative thought. But the psychological explanations that expressivists appeal to in order to solve the problems of metaethics are compatible with a range of different approaches to theorising about the semantics of normative language. Traditionally, expressivism has been characterised in terms of its commitment to a *psychologistic semantics* in which the meanings of sentences are explained in terms of the mental states they conventionally express (e.g. Rosen 1998: 387; Kalderon 2007: 54; Wedgwood 2007: 35f; Schroeder 2008: 33). More recently, however, a number of philosophers have sought to reconcile expressivism with orthodox truth-conditional semantics (e.g. Yalcin 2012, 2018a; Silk 2013, 2015; Charlow 2014; Ridge 2014; Chrisman 2016; Köhler 2018). And a number of others think that expressivists should endorse an increasingly popular 'dynamic semantic' approach (e.g. Charlow 2015; Starr 2016; Willer 2017).

So while expressivists will need some satisfactory semantics for normative expressions at the end of the day, I don't take this to be the central task for expressivism. At the very least, for the reasons suggested above, we should

[13] Sentential theories of belief according to which beliefs are relations to sentences (whether in a public language or a language of thought) might collapse this distinction. I will briefly say something about such theories below.

expect there to be a division of labour in providing an expressivist theory of normative thought and an expressivist theory of normative language. More generally, I think a focus on psychological rather than linguistic explanations is supported by the plausible idea that thought is *explanatorily prior* to language.[14] This is in contrast to the view that language is explanatorily prior to thought (e.g. Dummett 1973) and the view that neither has priority (e.g. Davidson 1975). Nonetheless, throughout the thesis it will at times be convenient to formulate certain arguments or views in terms of what is expressed by certain sentences or statements. However, such formulations are primarily heuristic and should not be understood as semantic claims that specify the meanings of sentences.

*Propositional attitude psychology as a theoretical enterprise*

As I have explicated the term, 'propositional attitude psychology' describes a *theoretical* approach to explaining our cognitive lives. Another important preliminary, therefore, concerns the relation between propositional attitude psychology on the one hand and ordinary language attitude attributions and folk psychological practice more generally on the other. For there are a number of different ways one might understand the relation between our theoretical and folk views. Ultimately, I think my approach in this thesis is compatible with any number of views, so I won't here take a stand on how we should conceive of this relation. Nevertheless, it is worth highlighting some of the differences to get a sense of the different ways in which one might understand propositional attitude psychology.

The most straightforward view of the relation between propositional attitude psychology and folk psychology would be that propositional attitude psychology is just an explicit articulation of (some subset of) folk psychology. On such a view, the 'that'-clauses in ordinary language attitude attributions would denote propositions, and a theory of propositions provides an account of what we refer to in ordinary language. A distinct though nearby view is that propositional attitude psychology provides *reforming definitions* for folk psychological terms and concepts. Perhaps our ordinary practices are too messy, vague, indeterminate, and context-dependent to be straightforwardly identified with propositional attitude psychology. But perhaps the latter captures in a more precise and determinate way what we need from folk

[14] This is compatible with language possession being a necessary condition for the possession of certain kinds of thoughts — see Peacocke (1996). Additionally, I take it that 'language' here means *public* language, as presumably one can accept the priority of thought while also accepting the language of thought of hypothesis.

psychology in order to explain our attitudes in a systematic fashion and perhaps vindicate our folk psychological practices.

Another view of the relation between folk psychology and propositional attitude psychology might be that the latter provides an explanatory *model* of (certain aspects of) folk psychology in the following sense (Godfrey-Smith 2005: 2f; see also Yalcin 2018c: 353):

> A model-builder's usual goal is to construct and describe various hypothetical structures. These structures are used to help us understand some actual target system or systems. Generally, the understanding is supposed to be achieved via a *resemblance* relation between hypothetical and real systems. But both the degree and kind of resemblance that is sought are adjustable. So a model itself does not contain commitments about what the target system is like, or even which system is the target. A model and its application are two different things.

Thought of in this way, the structures described by propositional attitude psychological provide indirect representations or idealised structures that shed light on our attitudes, behaviour, or whatever exactly we are appealing to propositional attitude psychology to explain. On this view, we can provide an explanation of our thought and discourse in terms of relating subjects to propositions without necessarily saying that the attitudes attributed in folk psychology are (approximately) constituted by such relations. There is a difficult philosophical question of how exactly to understand the explanatory relation between models and the phenomena they target (for discussion, see the essays collected in Morrison and Morgan 1999). However, explanatory models are commonplace and widely accepted across a wide variety of scientific disciplines. So the difficulty in explaining precisely how such a model could be explanatory is not itself an objection to such a view.

*Propositional attitude psychology and the metaphysics of mind*

I will argue later in the thesis that an adequate metaphysics of propositional content needs to be grounded in a suitable metaphysics of propositional attitudes. But it's worth highlighting that the framework of propositional attitude psychology itself is relatively noncommittal with respect to the nature of propositional attitudes. This is so for a number of reasons. Most obviously, if we understand propositional attitude psychology as providing a model, as was suggested above, then strictly speaking propositional attitude psychology itself makes no commitments about the nature of what it is used to model. It is in the application of the model that it says something about the target

phenomena, but this application may be more or less precise in what it says about the nature of the target phenomena. Further, even if we take propositional attitude psychology to be an explicit articulation of folk psychology, certain meta-theoretical views of folk psychology will give rise to similar positions. For example, Godfrey-Smith (2005) argues that our folk psychological capacities themselves should be understood as a kind of model-based understanding.

Putting aside the idea of propositional attitude psychology as a modelling enterprise, at the very least it might seem that the approach is committed to a *relational* view of attitudes, according to which propositional attitudes are essentially relations between agents and propositions. Strictly speaking, however, the relational view is not entailed by propositional attitude psychology. This is because propositional attitude psychology first and foremost provides a relational view of attitude *attributions*. To see how these come apart, consider the following analogy with temperature. For example, consider:

(5) The room is 20°C.

This claim attributes a certain temperature to the room. A natural reading of this claim is that it involves relating to the room to a certain abstract object. The logical form of this attribution might be thought to be one of the following:

(5′) $R$(room, 20)
(5′′) $R^*$(room, 20°C)

According (5′), the attribution states that the *temperature in* °C relation obtains between the room and the real number 20. According to (5′′) the attribution states that the *temperature* relation obtains between the room and the number 20 on the degrees Celsius scale. The differences between (5′) and (5′′) do not matter here. The relevant point is that in both cases, temperature attributions seem to relational in much the same way as attitude attributions. However, it is highly plausible that temperature is a *non-relational* intrinsic property. That the room is 20°C is plausibly a fact about the physical state of the room itself, and not a fact about the room's relation to something else. So from the fact that certain kinds of attributions are relational, it does not necessarily follow that what is attributed is itself relational.

Instead, the idea is that relating the room to an abstract object is a way of *measuring* an intrinsic state of the room. This is possible because the empirical structure of the room's temperature bears an isomorphism or homomorphism

to the structure of the real numbers. This structural similarity allows us to "surrogatively reason" about certain physical magnitudes by using an abstract representational domain (Matthews 2007: 133ff). Returning to propositional attitudes, a number of philosophers have suggested that propositional attitude attributions might be thought of as a tool for measuring intrinsic properties of subjects, such as behavioural aptitudes. This suggestion has been developed in detail by Robert Matthews (2007), but has also been suggested by Churchland (1981: 70f), Stalnaker (1987: 9f), and Dennett (1987e: 123ff), among others. The idea is that the empirical structure of our attitudes bears a homomorphism to the logical structure of the domain of propositional contents, and so we can use the latter to "surrogatively reason" about the former. On this view, there is a sense in which we literally measure each other's attitudes in our attitude ascriptions.

I mention this view not to endorse it but rather to illustrate the way in which propositional attitude psychology is compatible with a range of views about the metaphysics of attitudes. If we accept a measurement-theoretic view of attitude attributions, then we still need a theory of the abstract representational domain by which we measure our attitudes. Among other things, this will involve providing a theory of the abstract relata of attributions. In other words, a measurement-theoretic account needs a theory of propositions no less than a view according to which propositional attitudes are essentially relations to propositions. The measurement-theoretic account will simply say that standing in relations to propositions is not part of the metaphysical essence of our attitudes.[15]

The modelling and measurement views notwithstanding, the most straightforward explanation of the relational form of attitude attributions is simply that the attitudes attributed are themselves relational. Even here, however, it's worth highlighting that there remains considerable scope about how best to understand the metaphysical nature of such relational attitudes. One question is whether the relevant relations are conceptually or metaphysically reducible to naturalistic relations (e.g. Loar 1981; Stalnaker 1987) or whether propositional attitude psychology provides non-reductive autonomous intentional explanations (e.g. Dennett 1987a). In Chapter 5, I will

[15] See also Hanks (2015: 2f). Crane (1990) argues that the measurement analogy breaks down because intentional states are *essentially* characterised by their content, whereas it is arbitrary what scale we use to measure physical magnitudes like temperature. However, against this, one might question whether there is a uniquely correct notion of belief or belief content rather than different explications serving different theoretical purposes (compare Moore 1999).

endorse a non-reductive approach, but either approach is compatible with the general project undertaken in this thesis. It's also worth pointing out that all of the views discussed above are compatible with a wide range of views about how propositional attitudes are empirically realised in human beings or other creatures.

*Sententialism about propositional attitudes*

Propositional attitude psychology explains our mental states by relating subjects to abstract entities. This propositionalist approach is often contrasted to a sententialist approach according to which beliefs and other attitudes are essentially relations to sentences. Historically, the sententialist approach has been motivated by scepticism about propositions (e.g. because of a thoroughgoing physicalist ontology or because of worries about identity conditions for propositions). As such, it is often presented as a rival to the propositionalist approach. As I am simply assuming a propositionalist approach to explaining intentional thought and discourse, then insofar as the sententialist approach is a genuine rival to the propositionalist approach, it would be beyond the scope of this thesis to discuss the former in any depth. Moreover, I take the arguments discussed in the next section to most straightforwardly motivate a propositionalist approach rather than a sententialist approach. However, putting aside general scepticism about propositions, I think that the sententialist and propositionalist approaches are in fact compatible and might even be complementary.

Indeed, a natural way to develop the sententialist approach appeals to propositions, or at least something closely analogous such as meanings. To see how, consider that it is highly plausible that two subjects who speak different languages can believe the same thing. A straightforward way to explain this given a sentential approach would be to provide the following kind of analysis: a subject *S* stands in the *believing* relation to a proposition *p* in virtue of standing in the *believing** relation to a sentence $\sigma$ that means *p* (compare Field 1978), where *p* is the proposition expressed in that context.

This kind of analysis might be motivated on the grounds that we need to posit a relation to mental particular (i.e. a tokened mentalese sentence) in order to explain the causal efficacy of thought (compare Fodor 1987). So when sentential theories are not motivated by general scepticism about propositions, a theory of propositions might play a role in a sentential theory of belief. However, it seems to me that sentential theories are better viewed as empirical theories about how beliefs are empirically realised in human beings rather than as metaphysical theories about the essential nature of belief. So while it might

be true that human beings have beliefs in virtue of tokening mentalese sentences inside their 'belief boxes', this does not explain *what it is* to have beliefs.

This concludes my discussion of some of the issues surrounding expressivism and propositional attitude psychology that frame and constrain the discussion of this thesis. The main constraint is that providing an expressivist theory of normative propositional content is primarily understood to be a task for the expressivist's *theory of thought*. However, while I will in the end endorse a particular conception of propositional attitude psychology, the main lesson from this discussion is that the general project undertaken in this thesis is compatible with a wide range of views within the philosophy of mind about how to understand the nature of our attitudes and the relation of our theories to our folk psychological conceptions.

## 1.3 Motivations

In this section, I examine in more detail some of the motivations for positing propositions in general, as well as for positing normative propositions within an expressivist theory of normative thought and discourse. The arguments sketched here are not meant to decisively show that a nonpropositional theory of normative thought and discourse will fail. Rather, they make a prima facie case for a propositional approach, both in general terms and for expressivists. I doubt there are any general arguments that would decisively favour one approach over the other. The plausibility of either will ultimately depend on the plausibility of the particular theories that each approach provides. The main focus of this thesis will be to make the case that expressivists should, pace most of the expressivist tradition, adopt a propositional approach by showing how far we can develop that approach into detailed theories of propositional attitudes consistent with expressivism. Nonetheless, I think the arguments put forward in this section suggest that embracing propositions provides a simple and powerful way of explaining intentional thought and talk and that rejecting propositions for the normative domain is a heavy theoretical cost.

Stephen Schiffer (2003) argues that propositions are implicated in the 'face value theory' of belief and that this theory enjoys a prima facie status. The basic idea is that they are implicated in the most straightforward account of the logical form of belief attributions and the inferences we can make using them. For example, consider the following inferences (taken from Schiffer 2003: 12):

(6) Harold believes that there is life on Venus, and so does Fiona.
So, there is something that they both believe — to wit, that there is life on Venus.

(7) Harold believes everything that Fiona says.
Fiona says that there is life on Venus.
So, Harold believes that there is life on Venus.

(8) Harold believes that there is life on Venus.
That there is life on Venus is Fiona's theory.
So, Harold believes Fiona's theory.

(9) Harold believes that there is life on Venus.
That there is life on Venus is implausible.
So, Harold believes something implausible — to wit, that there is life on Venus.

Schiffer argues that the most straightforward way to account for the validity of these inferences is by attributing the following logical structure to each argument:

(6*) $Fab$ & $Fcb$
$\therefore \exists x(Fax \,\&\, Fcx)$

(7*) $\forall x(Fax \rightarrow Gbx)$
$Fac$
$\therefore Gbc$

(8*) $Fab$
$b = d$
$\therefore Fad$

(9*) $Fab$
$Gb$
$\therefore \exists x(Gx \,\&\, Fax)$

The idea is that we should interpret 'that'-clauses and expressions like 'Fiona's theory' as singular terms denoting propositions, and interpret 'something' and 'everything' respectively as existential and universal quantification over these propositions. If we reject propositions, then we need to find some other way to make (6)-(9) come out valid. Importantly, it seems that we can replace the

descriptive sentence "that there is life on mars" with any arbitrary normative sentence, e.g. "that one ought to give to charity" without affecting the validity of the inferences. So the expressivist who rejects normative propositions needs some alternative explanation of the logical form of the normative equivalents of (6)-(9).

(A natural candidate here would be to say that the singular terms refer to *sentences*. However, these inferences seem to be just as valid if we suppose that Harold and Fiona speak different languages. So the singular terms cannot simply refer to sentences. Perhaps one could maintain that they refer to classes of sentences, namely those that have the same meaning. But this would be to posit a particular kind of abstract entity that is the object of beliefs, which is tantamount to positing propositions, given the minimal characterisation of them above. If we accepted this approach, one question that would need to be answered is this: in virtue of what does a singular term refer to one set of sentences rather than another set? A natural candidate is that the sentences share truth-conditions. But if we are typing sentences by their truth-conditions, why not simply type those sentences by the proposition they express and in virtue of which they inherit their truth conditions?)

Additionally, Schroeder (2013: 76ff) argues that by positing propositions as the objects of our attitudes, we provide a straightforward account of what attitudes are expressed by logically complex sentences. Roughly, logically complex attitudes can be explained in terms of the complexity of the object of those attitudes. (This isn't quite right, as unstructured views of propositions do not consider them to be complexes; but the point remains that an attitude expressed by a logically complex sentence can be explained in terms of its propositional object.) That something like propositional structure is needed can be illustrated by examining a non-propositional analysis of complex attitudes. For instance, consider the following example from Schroeder (2008: 45, 2013: 77; the example is derived from Unwin 1999, 2001):

(10) Jon thinks that stealing is wrong.
(11) Jon does not think that stealing is wrong.
(12) Jon thinks that stealing is not wrong.
(13) Jon thinks that not stealing is wrong.

What account should the expressivist give of these states? Imagine an expressivist who thinks that claims about what is wrong express attitudes of *disapproval* towards what 'wrong' is applied to. On this view, we arrive at the following account of the above states:

(10*) Jon disapproves of stealing.
(11*) Jon does not disapprove of stealing.
(12*) ???
(13*) Jon disapproves of not stealing.

As the question marks indicate, it is unclear what attitude (12) consists in. This is because the attitude of disapproval does not possess the right kind of structure to explain what attitude is expressed by negating the sentence "stealing is wrong".

One possible strategy would be to posit a distinct attitude, say *toleration*, which is expressed by externally negated atomic 'wrong'-sentences. However, this move appears ad hoc, and in any case it no longer explains what (12) consists in. Rather, it simply posits that there is an attitude with the right properties to explain the relations between (12) and the other attitudes. By contrast, if (10)-(13) are analysed in terms of Jon standing in a single relation to different propositions, then we face no such problem. On this approach, the wide-scope 'not' expresses the truth-functional operation of negation, which provides a way of computing one proposition from another. Specifically, it is a function that takes a proposition [p] to the proposition [not-p]. If the contribution of 'not' in (12) is explained in terms of the object of Jon's thought, there is no need to posit a distinct attitude expressed by 'not wrong'.

I have only so far talked about negation, but the same point applies to other logically complex attitudes, including alethic modal attitudes about what is possibly the case and necessarily the case (Schroeder 2013: 82ff). As with logical connectives, the orthodox view of modals is to interpret them as propositional operators (e.g. Kratzer 1977). Given that normative sentences can fill the complement place of alethic modal terms, the most straightforward way to interpret normative modal claims would be to posit normative propositions as filling the argument position of a modal operator. More generally, if logical complexity can be explained in terms of the objects of our attitudes, then we have a straightforward way of accounting for the attitudes ascribed by logically complex attributions. By contrast, if logical complexity is to be accounted for in the attitude type rather than its content, as with (10*)-(13*) above, then we will face an ever-expanding taxonomy of different attitude types, the logical relations between all of which need to be explained. And it is very hard to see what a systematic explanation of all these relations would look like on this picture. This is a problem because all of our beliefs are systematically related in terms of their compatibility and inferential relations. By contrast, a propositional approach can provide a systematic explanation (partly) in terms

of the properties of the objects of belief, whereas a nonpropositional approach seems to have to proceed on a piecemeal basis.

Moreover, we not only want a systematic account of how our beliefs relate to one another, but also of how propositional attitudes in general relate to one another. Expressivists need to explain not just what it is for Jon to think [that stealing is wrong], but what it is for Jon to hope, fear, doubt, etc., [that stealing is wrong]. However, if we follow the disapproval model sketched above, then it hard to see how to generalise it to other attitudes in a systematic manner. This problem, originating from Rosen (1998: 393ff) and also put forward by Schroeder (2010: 83f; 2013: 81f), is known as the many attitudes problem. It is a problem because it seems that the expressivist must proceed in a piecemeal fashion to explain what each attitude consists in and how each attitude relates to other kinds of attitudes with the same content. By contrast, the propositional approach provides a systematic and straightforward explanation in terms of a subject's bearing a different relation for each kind of attitude towards the same propositional object.

While most expressivists have rejected explanatory appeals to normative propositions in their theories of normative thought and discourse, many are happy to accept their use in theories of descriptive domains of thought and discourse. However, this creates a number of additional problems for expressivism. The core problem is that taken at face value, much of the phenomena that justify positing propositions for descriptive domains are just as present in the normative domain. So even assuming that expressivists manage to provide some nonpropositional explanation for these phenomena in the normative domain, there remains a residual question about why it is that these different domains have completely different explanations for what look to be the same phenomena.

For example, all of the general relations that hold between different kinds of attitudes (e.g. the way belief and desires interact in the explanation of action) in the descriptive domain seem to hold just as true in the normative domain. Not only that, but these relations seem to hold across domains as well as within them. So whatever explanation we give of the underlying phenomena, we would expect it to explain how these relations hold at a fully general level. But given the expressivist's commitment to a bifurcation in explanation, it's difficult to see how this might be achieved. (As we will see in later chapters, considerations like these place important constraints on an expressivist theory of propositional attitudes.)

One option here would be for the expressivist to reject the use propositions in explaining descriptive domains of thought and discourse as well, resulting in a kind of global expressivism (e.g. Horwich 1993; Gibbard 2012; Price 2013).

But if we accept the utility of propositions in explaining descriptive domains, it is hard to resist their application to the normative domain. In any case, the main aim of this thesis is not to show why nonpropositional views, whether local or global, are doomed to failure (though I take the above to show that such approaches face significant explanatory burdens). Rather, the main aim is to explore some of the strategies that expressivists might employ in order to develop a theory of normative propositions as the objects of normative attitudes.

However, the aim is not only exploratory. It is also to evaluate and assess these different strategies with a view to arriving at the most promising expressivist view of normative propositions. I will argue that three different existing frameworks for theorising about content all provide an attractive basis for an expressivist theory of normative propositions. These will be the frameworks of cognitive act theories, conceptual role theories, and modal theories (to be explained below). However, I will also argue that all three approaches face difficult challenges in accommodating expressivism. In the end, I will argue that the last framework is the most promising and I will offer a solution to the challenges I raise for it in the form of a foundational theory of belief. The end result is not only an expressivist theory of propositions, but an expressivist theory of belief that incorporates both normative and descriptive beliefs. While this conclusion is reached through an examination of the particular theories, this result is exactly what we should expect, given the claim that the object-of-attitude role must be defined relative to one's overall theory of propositional attitudes.

In the final section of this introduction, I provide a roadmap to aid the reader through the remainder of the thesis.

## 1.4 Roadmap

Although the chapters of this thesis can be read as standalone chapters, the conclusions of each chapter are intended to lead naturally onto the next, and each chapter builds upon the conclusions of the previous ones. As I mentioned at the end of the previous section, the aims of this thesis are exploratory, constructive, critical, and prescriptive. Each chapter aims to explore how existing frameworks for theorising about mental content and the mind more generally can be adapted for an expressivist view of normative propositions and belief. Each chapter makes novel suggestions about how expressivists might develop such views. Each chapter aims to critically evaluate in the most general terms the application of each framework in an expressivist setting. And each chapter aims to assess whether expressivists should or should not adopt

the framework under consideration, or at least to make clear what challenges would need to be met if an expressivist were to adopt the framework. Moreover, the thesis as a whole is prescriptive in that it aims to recommend a particular approach to theorising about mental content and the mind more generally as the best framework for an expressivist theory of normative thought and discourse.

With the *in principle* case made for an expressivist theory of normative propositions in this chapter, Chapter 2 begins with a brief survey of orthodox conceptions of propositions and why these look unattractive from an expressivist perspective. I then suggest that cognitive act theories of propositions might provide an attractive alternative. According to such theories, propositions just are types of cognitive acts. I argue that expressivists who accept this view owe us an explanation of the identity of conditions of the class of cognitive acts that are propositions. Moreover, I argue that this explanation should apply uniformly to normative and representational propositions. I show how Michael Ridge's expressivist cognitive act theory fails to do this, and I propose a novel understanding of cognitive propositions aimed to solve the problem. Specifically, I suggest that we understand the relevant cognitive acts as acts of conceptual categorisation, which are individuated by the conceptual roles of the concepts involved. Although I do not defend the conceptual role view, I argue that this provides some licence for optimism, as this is a view that many expressivists are independently sympathetic to.

In Chapter 3, I take up the task of evaluating the relationship between expressivism and conceptual role accounts of content more generally. Although many expressivists endorse some kind of 'conceptual role expressivism', there is no general characterisation that unites these otherwise disparate views. I provide such a characterisation and explain why three recent versions of conceptual role expressivism due to Paul Horwich, Sebastian Köhler, and Neil Sinclair all fall under it. I then raise a challenge to conceptual role expressivism based on James Dreier's 'hiyo' predicate. In a nutshell, conceptual role expressivists seem to lack any story for why normative concepts are not defective if superficially similar 'hiyo'-like concepts are defective. After examining and rejecting some responses, I suggest that conceptual role expressivists might solve the problem by supplementing their theory with a suitable account of what content is *determined* by normative conceptual roles. For example, perhaps normative concepts roles determine some kind of planning content, in Allan Gibbard's sense. However, I argue that this makes conceptual role expressivism less distinctive and interesting than its proponents make out.

In Chapter 4, I examine Allan Gibbard's notion of planning contents and some of its subsequent developments. I first locate the view within a broader framework for thinking about content that I call the modal conception of content. This is the idea that content can be explained in terms of possibilities. The simplest version of this view is the possible worlds conception of content, and I argue that Gibbard's proposal that normative propositions encode planning information can be seen as a natural extension of the possible worlds view. Although this view has a number of features that should be attractive to expressivists, I argue that such a view of content needs to be grounded in a foundational philosophy of mind that is compatible with expressivism. After arguing against one such proposal due to Seth Yalcin, I propose that expressivists should ground this view of content in a general theory of belief that subsumes both normative and non-normative belief.

In Chapter 5, I develop a theory of belief within an interpretationist framework that aims to provide a suitable foundational philosophy of mind for the view of normative propositions examined in the previous chapter. I argue that interpretationism fits well with this view of content and should be attractive for expressivists wanting to develop a general theory of belief. I then show how classic arguments for expressivism can be reframed as arguments for this particular view of belief. Finally, I argue that this version of expressivism has a number of advantages over versions of expressivism that claim that descriptive and normative beliefs are fundamentally different kinds of attitudes.

In Chapter 6, the concluding chapter, I summarise the main conclusions of the previous chapters and discuss some of the further implications and directions of study suggested by the thesis. Additionally, I critically examine Mark Schroeder's nondescriptivist theory of belief and propositions. While Schroeder's view does not neatly fit into the frameworks discussed in the other chapters, it is one of the few nondescriptivist theories of propositions and belief aimed to solve the problems outlined in this chapter. It therefore constitutes a rival view to the theory of propositions and belief developed in Chapters 4 and 5. While the theory has a number of virtues, I will raise some objections to view that would need to be answered before we could accept the view. As the objections do not apply to the view endorsed in the previous chapters, I conclude that Schroeder's theory should be less attractive to expressivists.

# 2
# Cognitive Act Theories

This chapter explores the prospects for developing an expressivist theory of propositions within the framework of cognitive act theories of propositions. First, I motivate the framework as an attractive alternative to traditional theories of propositions (section 2.1). Second, I argue that the only extant expressivist theory of cognitive propositions — Michael Ridge's 'ecumenical expressivist' theory — fails to explain identity conditions for normative propositions (section 2.2). Third, I argue that this failure motivates a general constraint — 'the unity requirement' — that any expressivist theory of propositions must provide a unified nonrepresentational explanation of that in virtue of which propositional attitudes have the content that they have (section 2.3). Fourth, I begin to develop a novel account of the act type of predication that is both compatible with expressivism and designed to satisfy the unity requirement (section 2.4).

## 2.1 Cognitive propositions

In the previous chapter, I argued that expressivists would do well to incorporate a theory of normative propositions into their theory of normative thought and discourse. Suppose, then, that we are expressivists looking for a theory of normative propositions. Where should we look? We might choose to start from scratch and build our theory from the ground up. Alternatively, we might choose to survey some of the options already available. This chapter examines the prospects of one option in particular: cognitive act theories of propositions, which claim that propositions are types of cognitive acts. In part, this is because one of the few existing expressivist theories of propositions adopts the cognitive act framework (we will examine this theory in section 2.2). However, the cognitive act view provides a prima facie attractive framework for expressivism more generally. To see why, it is first helpful to consider why more orthodox conceptions of propositions appear less suitable for our purposes.

### 2.2.1 *Orthodox views rejected*

First, consider the Russellian view of propositions according to which propositions are structured entities consisting of objects, properties, and relations. Thus the proposition [that Socrates is human] is the complex abstract entity whose components include the individual Socrates and the property *being human*. These components are then combined in such a way that the proposition represents Socrates instantiating the property *being human*. Or the proposition [that Othello loves Desdemona] is the abstract entity whose components include the individuals Othello and Desdemona and the *loving* relation structured such that the proposition represents the *loving* relation obtaining between Othello and Desdemona (in that order).

It should be fairly clear that the Russellian view of propositions is unsuitable for expressivists to exploit. Russellian propositions are in part composed of the properties and relations that such propositions represent or are otherwise about. However, expressivists deny that there are any normative properties or relations. Or at least, expressivists deny that there are such things as normative properties or relations that could play an explanatory role in their theory of normative thought and discourse. Since we are after a conception of normative propositions that can play an explanatory role in an expressivist theory of normative thought and discourse, the expressivist cannot appeal to Russellian normative propositions.

Next, consider the Fregean view of propositions according to which propositions are structured entities composed of abstract senses or modes of presentation, where senses are cognitively individuated entities that determine a unique referent. Senses compose functionally and have differing levels of adicity. Thus the proposition [that Bob Dylan is a musician] is the structured entity composed of the "saturated" sense BOB DYLAN which has adicity 0 and the "unsaturated" sense MUSICIAN which has adicity 1, where the level of adicity is the number of argument places required to be filled in order to express a complete thought.

In contrast to the Russellian view, the proposition [that Bob Dylan is a musician] is distinct from proposition that [Robert Zimmerman is a musician] which is composed of the senses ROBERT ZIMMERMAN and MUSICIAN. This is because although BOB DYLAN and ROBERT ZIMMERMAN refer to the same individual, each sense has a distinct "cognitive significance". The reason each sense has a distinct cognitive significance is that substituting each sense in a complete thought can be potentially informative for a subject who grasps those senses. In the present example, one might know that Bob Dylan is a musician

without knowing that Robert Zimmerman is a musician, because one does not know that Bob Dylan is Robert Zimmerman.

The orthodox Fregean view is also unsuitable for the expressivist to exploit, although perhaps this is less obvious than with the Russellian view. The reason is because Fregean senses are ultimately individuated in terms of their worldly referents. Thus, while the senses of the expressions 'Bob Dylan' and 'Robert Zimmerman' do not consist in the individual these expressions refer to, grasping each sense does consist in knowing the condition for something to be its reference. This is because the sense of an expression determines what that expression refers to. However, expressivists deny that the referent of a normative expression is determined by its sense or meaning, other than perhaps in a deflationary sense. They must therefore deny that normative concepts express Fregean senses, at least standardly construed, as although Fregean senses are distinct from their denotation, they are nonetheless individuated in terms of what they denote.

Finally, consider an orthodox possible worlds view according to which propositions are sets of possible worlds, or alternatively functions from worlds to truth-values. Thus the proposition [that all grass is green] is the set of possible worlds in which all grass is green, or alternatively a function from worlds to truth values, specifically from worlds where all grass is green to the value True and worlds where not all grass is green to the value False. Possible worlds propositions are thought to represent, individuate, or perhaps simply be possible ways the world might be. In contrast to the Russellian and Fregean views, the possible worlds view provides an unstructured account of propositions that denies that propositions are complexes of simpler entities. Thus, on the possible worlds view, the proposition [that all grass is green] is identical to the propositions [that there is no grass that is not green] and [that it's false that it's false that all grass is green] because each proposition is true in all the same worlds.

As it stands, the possible worlds view is unsuitable for the expressivist to exploit. This is because expressivists deny that there are any normative "ways the world might be", where this is not a deflationary shorthand for a very general first-order normative claim. As such, the content of a normative belief cannot be represented or otherwise picked out by any set of possible worlds if this content is to play an explanatory role in characterising normative belief. We will see in Chapter 4 that there are ways of expanding the possible worlds model to incorporate certain kinds of nonfactual content, allowing the expressivist to adopt an enriched possible worlds model. However, I postpone discussion of this approach here. The important point for now is that as it

stands the orthodox possible worlds view of propositions is unsuitable for theorising about normative propositions from an expressivist standpoint.

What these orthodox views have in common is that they explain the nature of propositions partly in terms of what those propositions represent or refer to. Thus, in the terminology of the previous chapter, each view is committed to the representationalist assumption. Because expressivists deny that normative expressions or concepts refer to normative properties or relations, other than in a deflationary sense, the expressivist cannot appeal to orthodox conceptions of normative propositions. This means that expressivists require a conception of normative propositions that is not explained in terms of the normative properties, relations, or states of affairs that such propositions represent or individuate.

### *2.1.2 Expressivism and cognitive propositions*

A more promising candidate is found in cognitive act theories of propositions (examples include Carruthers 1989; Dummett 1991; Davis 2003; Soames 2010, 2014, 2015; Hanks 2015). According to such theories, propositions are ways of thinking and speaking. More specifically, propositions are certain kinds of cognitive acts or events. Call such entities 'cognitive propositions' for short. This immediately raises the question of what kind of cognitive acts could be suitable to play this role. But the basic idea should look attractive to expressivists. After all, normative cognitive acts are clearly the kind of thing that an expressivist can allow in her theoretical ontology. Indeed, expressivists do not simply allow for normative cognitive acts. Providing an account of such things is part of the central task for an expressivist theory of normative thought. Moreover, in the same way that expressivists aim to elucidate normativity by providing a psychological account of what it is to think normative thoughts, cognitive act theories aim to elucidate meaning and intentionality by providing a psychological account of what it is to have thoughts with propositional content. So expressivism and cognitive act theories appear to make natural bedfellows.

Further, the cognitive act view provides an attractive alternative to orthodox views more generally. Recent exponents of the cognitive act view see their view as answering a host of difficulties that face orthodox conceptions (see especially Soames 2015 and Hanks 2015). For example, according to orthodox views, propositions are inherently representational and are the fundamental bearers of truth and falsity. However, if this is the case, then a proposition cannot be identified with any set-theoretic object, such as an n-tuple of objects and properties or a set of possible worlds. For such objects themselves do not

represent anything; rather, they are representational only insofar as we interpret them as such. In light of this difficulty, it might seem that we would do better to treat (say) an n-tuple of objects and properties as *representations* of propositions (Soames 2015: 13). However, this leads us to posit a sui generis class of abstract objects with the necessary properties to play this role rather than providing any real explanation of what propositions are like such that they can play that role. Moreover, cognitive act theorists argue that if propositions are the primary bearers of truth and falsity, then our beliefs and assertions have their truth-conditions only derivatively in virtue of their being related to propositions. But it seems mysterious how concrete thought and talk is meaningful by being related to some mysterious Platonic realm of sui generis abstract objects. By contrast, the cognitive act view reverses the order of explanation by explaining the nature of propositions as derivative from concrete intentional activity, which is taken as fundamental. This is because the nature of a type is derived from its concrete instances. Among other things, the hope is that this provides a naturalistically plausible metaphysics and epistemology of propositions.

While the basic idea of cognitive propositions is a more promising starting place for an expressivist theory of propositions, the idea is typically developed in a way that is unsuitable for expressivists. This is because cognitive act theories generally still give a central role to the notion of *representation* in explaining the relevant cognitive acts. In other words, extant cognitive act theories are committed to the representationalist assumption about propositions. For example, Scott Soames (2015) argues that the proposition [p] is identical to the act of *entertaining p*, where to entertain p is to represent some state of affairs. Thus, the proposition [that the sea is blue] is identical to the act type of representing the sea as being blue (we will examine Soames' theory in more detail throughout this chapter). And Peter Hanks (2015) argues that the proposition [p] is identical to the act of *predicating properties or relations to things*. Thus, the proposition [that the sea is blue] is identical to the act type of *predicating blueness of the sea*. Such views are clearly unsuitable for expressivists.

Nonetheless, the basic idea that propositions are cognitive acts does not entail that the relevant acts are representational in nature. So it is open to the expressivist to characterize normative propositions as cognitive act types that are nonrepresentational in the relevant sense. This means explaining the relevant acts other than in terms of their representing normative states of affairs or attributing normative properties or relations to things. However, this is something that expressivists aim to do with respect to normative thought and discourse anyway. So the fact that cognitive act theories typically develop their theory by appealing to the representational properties of the relevant acts

should not deter us from the possibility of an expressivist cognitive act theory of normative propositions, at least from the outset.

The challenge for an expressivist variant of the cognitive act theory is to provide an account of cognitive propositions that is suitably nonrepresentational. In the remainder of this chapter, I examine two answers to the challenge. The first comes from Michael Ridge's 'ecumenical expressivist' theory of normative thought and discourse (section 2.2). Within this theory, Ridge adopts Soames' cognitive act theory as it applies to representational thought and talk, and then extends the account to also cover normative thought and talk as Ridge conceives it under ecumenical expressivism. I argue that Ridge's theory fails to explain the identity conditions for normative propositions and is thereby either extensionally or explanatorily inadequate. I then argue that that this failure motivates a general constraint — the 'unity requirement' — on any expressivist theory of propositions to the effect that any such theory must provide a unified nonrepresentational explanation of that in virtue of which propositional attitudes have the content that they have (section 2.3). Finally, in a constructive spirit, I propose a novel conception of *conceptual predication* that is suitably neutral with respect to whether normative predication is representational in nature (section 2.4). I suggest that conceptual role accounts of content might provide a suitably nonrepresentational explanation of the identity conditions of conceptual predication. Whether such a view is ultimately viable is then taken up in the next chapter.

### *2.1.3 Two objections*

Before proceeding, I want to briefly consider two possible objections concerning my discussion about expressivism and orthodox views of propositions. For the purposes of this thesis, I have understood expressivism in as broad terms as possible. I took this to be the conjunction of (i) the negative thesis that normative thought and discourse is nonrepresentational together with (ii) some positive thesis about its practical role. Taken this way, this characterization would seem to classify Mark Kalderon's (2007) moral fictionalist position as expressivist. (Kalderon rejects the label but that is merely terminological.) In relation to my above discussion, this is a problem because Kalderon argues that moral sentences express propositions that represent the attribution of moral properties to things. However, Kalderon also denies that accepting a moral sentence consists in believing the proposition it expresses. Instead, it consists in some kind of noncognitive attitude. But this apparently gives us an account of moral thought and discourse that respects the

conjunction of (i) and (ii) yet appeals to orthodox normative propositions. If this is a theoretical possibility, then what motivation is there for constructing a new theory of propositions rather than opting for Kalderon's?

By claiming that moral sentences express propositions, Kalderon aims to provide a straightforward solution to the Frege-Geach problem by showing how there can be straightforward semantic entailment between moral sentences. However, moral propositions play no role for Kalderon in relation to explaining the noncognitive attitudes involved in accepting moral sentences. In the previous chapter, I argued that there are good reasons for positing normative propositions in one's theory of normative *thought*. If this is right, then Kalderon's account of normative attitudes will face many of the same problems that more traditional expressivists face. For example, although his account might be able to explain how a set of *sentences* semantically entail their conclusion, normative propositions can play no role in explaining why *accepting* those sentences logically commits one to also accept the conclusion (see Eklund 2009).

It is also unclear how Kalderon proposes to explain thoughts with mixed contents. Further, because only moral sentences and not moral attitudes have moral propositions as their content, it is unclear to what extent Kalderon's approach is compatible with the priority of thought thesis, which we assumed in the previous chapter. Finally, supposing that there *are* orthodox moral propositions, as Kalderon claims, it is unclear why we can't *believe* such propositions. Kalderon argues that *in fact* we don't believe them because, very roughly, moral acceptance and belief are governed by different epistemic norms. This itself is contestable (see Joyce 2012; Chrisman 2007). But even granting that moral acceptance is governed by difference epistemic norms, it's unclear why this should entail that moral acceptance is noncognitive. In short, Kalderon's view faces many problems of its own. While a full discussion of Kalderon's view would take us too far afield, I think that the above considerations provide sufficient reason to set this kind of approach aside.

The next objection to my argument against combining expressivism with an orthodox view of propositions derives from a discussion of Michael Ridge's (2014: 223f). Although Ridge proposes a novel cognitive act theory of normative propositions, which we will discuss in the next section, he hedges his bets and argues that if the cognitive act view is wrong, his general expressivist theory can be combined with orthodox views of propositions. The suggestion comes at the end of a discussion about how his expressivist theory can accept a correspondence theory of truth, so long as talk of truth is construed as normative. Without going into details, the basic idea is that if talk of normative propositions is also normative, then such talk can be construed as

talk about which factual propositions constitute the normative proposition in question: "Different speakers can disagree about which descriptive/factual proposition does constitute a given proposition, and this can be a form of practical disagreement — disagreement in prescription." (2014: 223) Here, a 'factual' proposition describes any kind of orthodox proposition. If this is right, it might therefore seem that expressivists do not need any new or novel conception of normative propositions.

However, if we take Ridge's discussion to be an objection to the arguments against combining expressivism with orthodox propositions (which is not necessarily what Ridge intends), it arguably fails due to an equivocation concerning 'talk about propositions'. As I understand him, Ridge is primarily concerned to explain and vindicate our ordinary talk of normative propositions. Thus, for example, consider the sentence "Philip believes that talking in bed ought to be easiest." The idea is that it is a normative matter what factual proposition we refer to when we ascribe this belief in ordinary belief attributions. However, our current investigation was not simply to explain our ordinary talk of propositions but to provide a conception of normative propositions that can play an explanatory role in a theory of normative attitudes. By treating talk about propositions as normative in this way, it's unclear that we arrive at any such conception. From a theoretical perspective, I should be able to characterise Philip's belief without taking a stand on any first-order normative issue. But it's hard to see how I could provide a suitably neutral characterisation if proposition talk is normative in this way. Ridge compares his approach to Gibbard's (2003) proposal that normative properties are constituted by natural properties, even if normative concepts are distinct to natural concepts. However, Gibbard stresses that although normative properties are constituted by natural properties, the meaning of a normative concept cannot be explained in terms of its denotation. So while Ridge's approach may in a sense explain what normative propositions are, it fails to provide a conception of normative propositions that could play an explanatory role in the expressivist's theory of normative thought and discourse.

More generally, while there might be approaches that combine expressivism and orthodox normative propositions, these views fail to respect the considerations set out in the previous chapter that motivated our search for an expressivist theory of normative propositions. With these two objections to rejecting orthodox conceptions of propositions considered, I now turn to Ridge's theory of cognitive propositions.

## 2.2 Normative propositions in ecumenical expressivism

In this section, I briefly explain Ridge's (2014) ecumenical expressivism and the theory of normative propositions therein. I then raise two objections to the view. First, I argue that Ridge's theory fails to plausibly identify any specific cognitive act type to play the role of normative propositions. Second, I argue that even if the first objection can be answered, Ridge's theory fails to explain the identity conditions for normative propositions. This second objection gives rise to a general desideratum on any expressivist theory of propositions, discussed in section 2.3.

### *2.2.1 Ridge's theory*

The central claim of ecumenical expressivism is that normative claims express hybrid, relational states of mind (see also Schroeder 2015: chapters 6 & 7 and Toppinen 2013 for other 'relational' versions of expressivism). These are complex states comprised of a representational and nonrepresentational component. The nonrepresentational component is a kind of noncognitive practical stance (a 'normative perspective') that provides the agent with a set of policies about which standards of practical reasoning to reject and accept (Ridge 2014: 115). A standard is a rule that can be used in an action-guiding way as the basis of a practical judgment or decision (ibid: 40). The standards that make up an agent's normative perspective are 'ultimate', in that they are not derived from other standards and provide a complete guide to action (ibid: 116f.). However, standards are ultimately explained in terms of what it is for an agent to accept a standard, where this is for that agent to be disposed to issue the relevant prescriptions and intuitively endorse them (ibid: 111f.).

On this view, the representational component of the state expressed by a normative claim is a robustly representational belief. The belief is indexed to the agent's normative perspective such that, for any normative judgment, the object of evaluation is evaluated or ranked by the standards not ruled out by the agent's normative perspective (Ridge 2014: 119). The exact nature of the ranking will depend on the normative predicate employed in the claim (e.g., very roughly, 'good' will mean 'ranked high', while 'bad' will mean 'ranked low'). While the content of the representational component is robustly representational, the overall content of the normative claim itself is not identified with any representational content. Rather, it is irreducibly normative.

To illustrate the above, consider the following example:

(1) "Giving to charity is good."

According to Ecumenical Expressivism, (1) conventionally expresses the relational state of mind comprised of:

($1_a$) A normative perspective.
($1_b$) The belief that giving to charity is highly ranked by any admissible ultimate standard of practical reasoning.

The state expressed is relational in virtue of the concept of an *admissible standard*, which refers to the standards not rejected by the normative perspective of the speaker (Ridge 2014: 119). As different speakers will have different normative perspectives, ($1_a$) and ($1_b$) together comprise the relational state type that is multiply realized by any agent that tokens the state expressed by (1). Logical complexity is then "off-loaded" to the representational component of the state, which sets the stage for Ridge's attempt to solve the Frege-Geach problem (ibid: 144ff). For example, the claim "If giving to charity is good, then Socrates would approve of it" would express the multiply realizable relational state comprised of ($1_a$) together with the belief [that if giving to charity is highly ranked by any admissible ultimate standard of practical reasoning, then Socrates would approve of it].

Thus far, this account has explained normative thought and discourse without recourse to normative propositions. However, as I argued in the previous chapter, there are lots of good reasons for everyone to recognize normative thought and discourse as propositional. To accommodate normative propositions within ecumenical expressivism, Ridge appeals to Soames' (2010, 2014, 2015) cognitive act theory of propositions. For Soames, the nature of propositional content is explained by concrete cognitive activity. On his view, propositions are certain types of cognitive acts or events.[1] More specifically, propositions are acts that are tokened in representational activity: "*Propositions are repeatable, purely representational, cognitive acts or operations the performance of which results in concrete cognitive events; to entertain a proposition is to perform it*" (Soames 2015: 16, original emphasis). Identity conditions for propositions are specified in terms of the essential representational properties of such acts.

For example, consider the proposition [that the sea is blue]. According to Soames, this is identified with the cognitive act type of *representing the sea as being blue*. Judging, asserting, imagining, and so on [that the sea is blue] are all

[1] I'll follow Soames in assuming that acts are events rather than, say, processes. For the latter view, see Hornsby (2013).

*ways* of entertaining this proposition. To perform any of these acts is to token the act type of *representing the sea as being blue*. Thus, the identity conditions of the proposition [that the sea is blue] are explained in terms of representational properties of concrete acts of *representing the sea as being blue*.

With this basic notion of entertaining a proposition in place, other propositional attitudes and acts can be defined in terms of it. For example, to judge [that the sea is blue] is to perform that predication in "the affirmative manner", where this is cashed out in terms of its role in one's cognitive economy (Soames 2015: 18). To believe [that the sea is blue] is to be disposed to judge [that the sea is blue], and so on (Soames 2014: 97; 2015: 18f.). Soames states that all propositional acts and attitudes are ways of entertaining propositions. However, the case of belief shows that although this might be true of propositional *acts*, propositional *attitudes* like beliefs are best thought of as having content in virtue of their standing in some appropriate relation to propositional acts.

For Soames, as for most others, propositional thought is eo ipso representational. Hence, in order to accommodate normative propositions, expressivists need to broaden the relevant class of cognitive event types that constitute propositions. This is analogous to the more familiar expressivist move of broadening the relevant class of mental states that are 'beliefs' to include not only robustly representational beliefs, but any mental state conventionally expressed by sincere utterances of declarative sentences (Ridge 2014: 128). Whereas Soames appeals to an intuitive antecedent understanding of 'representing' out of which propositional acts and attitudes can be defined, Ridge reverses the order of explanation. First, he provides an account of what it is to believe, desire, assert, fear, and so on, the normative claim 'p' without any appeal to the proposition [p]. Second, he abstracts away from each case to that cognitive event type that is tokened in all and only those acts and attitudes. The idea is that each such act is a way of entertaining the normative proposition [p], and each such attitude stands in an appropriate relation to entertaining the normative proposition [p].

Because ($1_a$) and ($1_b$) provide an account of what is involved in normative belief qua mental *state*, this cannot itself be a way of entertaining a proposition for the reasons given above. What, then, is the cognitive act that it is appropriately related to? Ridge (2014: 128) proposes the following:

($1_a$) A normative perspective.
($1_c$) The judgment that giving to charity is highly ranked by any admissible ultimate standard of practical reasoning.

Here, the belief component is replaced with a judgment component, which we saw above was plausibly belief's active counterpart. In the next subsection, I'll argue that this move faces a number of problems. But let's run with it for now. The suggestion, then, is that <($1_a$) and ($1_c$)> constitute a way of entertaining the proposition [that giving to charity is good], and <($1_a$) and ($1_b$)> has the content [that giving to charity is good] by being appropriately related to the judgment constituted by <($1_a$) and ($1_c$)>.

Judgment is not the only way of entertaining a proposition, however. The proposition [that giving to charity is good] can also be entertained in other propositional attitudes, such as desire, doubt, hope, and so on. Glossing over complications that arise from the difference between states and events, ecumenical expressivism is well placed to generalise its account of belief to other kinds of attitude (Ridge 2014: 148). For example, the doubt [that giving to charity is good] can be understood as the complex state of mind comprised of:

($2_a$) A normative perspective.
($2_b$) The doubt that giving to charity is highly ranked by any admissible standard of practical reasoning.

Here, the 'that'-clause in ($2_b$) denotes a representational content indexed to the agent's normative perspective in the same way as ($1_b$). This move can then by applied *mutatis mutandis* to other attitude types. Ridge further notes that it is possible to entertain a normative proposition whereby one "simulates" a normative perspective, where this is run "off-line", such as when one "merely entertains" that *p* (ibid: 128).

The proposition that giving to charity is good is then identified as the minimal cognitive event type tokened in cognitive acts such as <($1_a$),($1_c$)> and is appropriately related to relational states such as <($1_a$),($1_b$)>, <($2_a$),($2_b$)>, and so on. In contrast to the representational case, it is somewhat harder to intuitively grasp what this act type this consists in. However, it should be clear that it is the act type all instances of which have the following features: (i) a component with a stable representational content[2]; (ii) an actual or simulated normative perspective; and (iii) the structural relation between (i) and (ii) encoded by the concept of *being ranked highly by any admissible ultimate standard of practical reasoning*. It is in virtue of this concept that the component with stable representational content is necessarily tied to the agent's normative perspective

[2] Given that this content will involve the indexical concept of 'admissible standard', the stability for Ridge will have to be at the level of *character* rather than *content*, in Kaplan's terminology.

in every possible instance of entertaining the proposition that giving to charity is good. With Ridge's theory of normative propositions now explained, I will raise two objections to it.

### *2.2.2 A problem concerning states and events*

The first objection to Ridge's theory of propositions is that by not paying sufficient attention to the distinction between *states* and *events*, Ridge's theory fails to identify *any* mental event that could be the object of normative attitudes. It is highly plausible that states and events are distinct metaphysical categories and should not be identified (see, for example, Kenny 1963; Vendler 1967; Mourelatos 1978; Steward 1997; Chrisman 2012). Events are dynamic in that they possess a temporal structure that states lack. They have a different 'temporal shape', in that they occupy time in different ways (the term is Steward's 1997: 72). Whereas events *happen*, states *obtain*. This distinction can be illustrated by comparing the different temporal properties of judgments and beliefs. Understood as a mental event, a judgment is an achievement. To make a judgment is something that one does. It is active in a way that believing is not. Belief, understood as a mental state, is better understood as a disposition that obtains across time. Thus, to say that Jones believes such and such at time *t* is not to say that Jones *does* something at *t*, in the same way that to say that Jones judges such and such at *t* does imply that Jones does something at *t*.

With this distinction in mind, the problem for Ridge's theory is that the general account of normative attitudes from which normative propositions are abstracted only provides an account of normative mental *states*. And if states are different in kind to events, then we cannot identify *any* mental event by abstracting from mental states. Thus, Ridge's theory fails to identify any candidate cognitive act type that could play the role of normative propositions.

As I noted in the previous section, Ridge is aware of the problem. I'll argue, however, that his response is unsatisfying. Consider again the judgment [that giving to charity is good]. According to Ridge, this is constituted by:

($1_a$) A normative perspective.
($1_c$) The judgment that giving to charity is highly ranked by any admissible ultimate standard of practical reasoning.

Here, ($1_c$) replaced the belief state ($1_b$) in the initial analysis. Because judgment seems to be a clear example of a mental event, it might seem that ($1_a$) and ($1_c$) together constitute a normative mental event. The obvious problem with this, however, is that although we have substituted an event for a state for the

representational component, we have failed to do so for the nonrepresentational component. If both components were events, then the occurrence of both events could constitute a complex event, which could be identified with the judgment [that giving to charity is good]. However, a normative perspective is not an event but a state. And it is far from clear that an event could be constituted by a distinct event together with a state.

To be fair, Ridge (2014: 128n) acknowledges this problem as well. In a footnote, he offers two different answers to this problem. The first strategy maintains that normative perspectives qua states can either be or else partly constitute events. The second strategy maintains that the manifestations of normative perspectives can provide the right sort of event. I consider each strategy in turn.

The first strategy can be maintained by arguing either that (i) normative perspectives are themselves events, or (ii) normative perspectives, while not themselves events, can partly constitute events. Taking (i) first, Ridge argues that we might understand a normative perspective as an event in virtue of the fact that one's *having* of a normative perspective is something that can happen at a moment in time. Depending on one's view of events, it is not obvious that any such event actually exists. However, granting that this is a real event, we have been given no reason to identify this event with the state itself. Simply because the durational properties of the event match the state, it does not follow that the event can be identified with the state. As we saw above, the fact that the durational properties of an event match some state, it does not follow that they have the same temporal properties, for they plausibly have a different temporal shape. By comparison, suppose we can pick out the event of my believing [that 2 + 2 = 4], which has been happening since my early childhood to the present day. It seems counterintuitive and misleading, however, to identify that event with the belief state itself. The same goes for normative perspectives.

Even if normative perspectives are not themselves events, perhaps they can partly constitute events. This was suggestion (ii) of the first strategy. However, this is to merely state what needs to be shown. We haven't yet been given any reason to think that this is true. Given their different temporal shape, one might be sceptical of the general claim that events can be partly constituted by states (compare Steward 1997: 72ff). But even putting such general scepticism aside, it is totally unclear why a descriptive judgment such as ($1_c$) together with a normative perspective could together constitute a distinct event. Note that it is not enough for Ridge to say that it is necessary that certain states obtain for a certain event to occur. What needs to be shown is something stronger, namely that the relevant state partly constitutes the event. For example, it seems

plausible to suppose that in order to judge [that p] it is necessary to believe [that p]. However, there is no pressure whatsoever to think that the judgment is partly constituted by the belief. Without any details as to how states generally could be partly constitutive of events, let alone any arguments for why normative perspectives in particular could be, the onus is surely on the ecumenical expressivist to argue this point.

So I think we have no reason to accept the claim that states can be or else partly constitute events. What about the suggestion of the second strategy, that is it the *manifestations* of normative perspectives that partly constitute the relevant events? Ridge suggests that these manifestations might be "intentions to act and reason in certain ways" (2014: 128n). This can't quite be right, because intentions are states and not events. But perhaps *deciding* to act and reason in certain ways or *forming* intentions to act and reason in certain ways might be better candidates. However, this can't be right either because it is possible to judge some normative claim as true without deciding or forming an intention to act in accordance with that judgment. For instance, perhaps I have already made the decision or formed the intention from a previous judgment. Perhaps instead one might invoke some broad notion of *affirming* to act or reason in certain ways, where this applies to deciding and forming an intention, as well as re-affirming where one has already decided or formed the intention. However, this is just to give a label to type of mental event that we are looking for and gives us no reason to believe that such a type exists.

So the second strategy owes us an account of the relevant manifestations of normative perspectives that could partly constitute normative judgments. Moreover, it owes us an argument that some such manifestation does indeed exist in every instance of entertaining a normative proposition, not just in judgment but across all possible ways of entertaining it. Although I am here primarily concerned with Ridge's theory of propositions, it's worth noting that this difficulty applies to his theory of normative thought more generally. Whatever the fate of normative cognitive propositions, ecumenical expressivism needs to explain what it is to judge normative claims. But without any account of the active counterpart of normative perspectives, it fails to do this. Nonetheless, I haven't said anything to *rule out* the possibility of such an account, so the argument here is not decisive. In any case, the objection raised in the next section will apply to Ridge's theory of normative propositions however this issue is resolved. So let's turn to that now.

### *2.2.3 A problem concerning identity conditions*

A more serious problem for Ridge's theory is that it presupposes rather than explains identity conditions for normative propositions. Whatever a theory of propositions is for, presumably it should explain their identity conditions, i.e., the conditions under which propositions are individuated and distinguished from one another. (Hence the *Stanford Encyclopedia of Philosophy* entry on propositions states: "[A]ny good theory of propositions ought to have something to say about when propositions are identical and when they are distinct" [McGrath and Frank 2018].) For example, on the possible worlds view, propositional identity is explained in terms of set identity: *p* and *q* are identical just in case *p* and *q* have the same worlds as elements. Or on the Russellian view, propositional identity is explained in terms of constituents and structure: roughly, *p* and *q* are identical just in case *p* and *q* have the same constituents in the same ordering. The ecumenical expressivist view, I will argue, fails to provide principled identity conditions for normative propositions. As such, it either makes incorrect predictions about which cognitive act types are propositions or makes correct predictions but only by presupposing which cognitive acts are normative propositions.

To see why, consider the following comparison. We saw above that for Soames, propositions are not just representational cognitive acts, but *purely representational* cognitive acts. Soames needs this additional constraint because there are far more representational cognitive acts than there are propositions, and this allows him to identify the subset of these acts that are propositions. Thus, for example, "the acts of *predicating humanity of Plato* (i) *on Thursday,* (ii) *in Peru,* (iii) *while dancing,* (iv) *in giving a lecture,* or (v) *when speaking in a whisper* all represent Plato as being human, even though we would not be happy thinking of them as propositions" (Soames 2015: 70, original emphasis). If just any representational cognitive act were a proposition, then *predicating humanity of Plato on Thursday* would be a distinct proposition from *predicating humanity of Plato*. But it is implausible that there is any such proposition. However, if cognitive propositions are purely representational, then the act type of *predicating humanity of Plato on Thursday* is not a proposition. This is because the condition of the act being on Thursday does not contribute in any way to representing Plato as human.

It is important to note that Soames' constraint is not an ad hoc fix to a problem of deviant or gerrymandered cases. Rather, it is a consequence of his commitment to explaining the nature of propositions in terms of concrete representational activity. We will explore the significance of this in the next

section. For now, however, we can ask the following question of Ridge's theory. For any normative claim 'p' and for the cognitive act types

($A_1$) entertaining p
($A_2$) entertaining p on Thursday,

in virtue of what is ($A_1$) a proposition but ($A_2$) not a proposition? As far as I can see, Ridge's theory offers no answer to this question. (It's worth highlighting that the act type of ($A_2$) has all the same properties as ($A_1$) that were needed to solve the Frege-Geach problem for normative attitudes.) This means either that Ridge's theory makes incorrect predictions about what cognitive act types are propositions, or it rules out cognitive acts like ($A_2$) as the objects of attitudes simply by fiat. There are many more normative cognitive act types than normative propositions, and Ridge's theory provides no principled way of delineating the latter class from the former.

One possible response would be to highlight that because the theory of propositions is explanatorily downstream from the theory of attitudes, there is no real worry here. The idea is that since we begin with an account of what it is to judge, doubt, and so on [that p] and only then go on to abstract the proposition [p] from this account, there is no possibility of deriving acts like ($A_2$) as propositions. However, in response to this response, there is no reason why we cannot provide an account of what it is to judge, doubt, and so on [that p] *on Thursday*. Indeed, the account is the same other than the condition that the acts occur on a Thursday. It is true that we probably have no interest in giving an account of this kind of attitude. But that is a pragmatic question about our interests. Just as with Ridge's original account, we have a theory of a set of concrete acts or attitudes from which we can abstract the act type ($A_2$) that is tokened in all and only those acts or attitudes. So even if we might never in fact have use to refer to the act type ($A_2$), the theory still predicts that it is a proposition.

Another response would be to find some property that ($A_1$) possesses and ($A_2$) lacks in virtue of which ($A_1$) but not ($A_2$) is a proposition. Given the hybrid nature of normative thought, perhaps we might say that the act of entertaining a normative proposition must possess the conjunctive property of having some *purely representational* component, some *purely practical* component, and no other component, where 'purely practical' identifies the relevant action-guiding property in virtue of which normative judgments settle the thing to do. To make good on this suggestion, one would need to cash out the notion of 'purely practical' without begging the question about what sort of acts can be purely practical. The notion must be related to guiding action. However, it

cannot be related in just any way. For example, suppose that while entertaining the proposition [that I ought to Φ], I simultaneously experience an unconnected occurrent desire to Φ (or manifestation thereof). Insofar as this desire might determine my actions or intentions to act, on what grounds can we exclude this desire from being 'practical' in the relevant sense? Further, suppose I entertain the proposition [that I ought to Φ] by imagining [that I ought to Φ]. Such an act bears no direct link to action or intention, yet must still be a realization of an act type that has a 'purely practical' component. So 'purely practical' must somehow include these cases as well.

These considerations are not decisive. For example, we might first appeal to the purely representational component and then derivatively identify the practical components necessarily implicated by the representational component. Specifically, we would identify the normative perspective necessarily implicated by the concept of an admissible standard. However, even assuming a suitable notion of 'purely practical' can be provided, it is not clear that this response does anything other than re-invite the original worry. Why is it that only act types with this conjunctive property are normative propositions? The appeal to the property of being purely practical is not simply meant to provide an account of what is distinctive about the nonrepresentational component of normative propositions. Rather, the property was meant to explain (in part) that in virtue of which cognitive act types that possess the conjunctive property are propositions. But again, we seem to have simply stipulated that cognitive act types that have this conjunctive property are propositions.

In the next section, I argue that the very *form* of this kind of explanation is problematic. This is because the approach posits an implausible explanatory bifurcation regarding why representational and normative cognitive acts can play the role of propositions. As we will see, this motivates a desideratum on *any* expressivist theory of propositions.

## 2.3 The unity requirement

In light of the considerations above, I propose that any expressivist theory of propositions must conform to the following desideratum:

> **Unity requirement**. Expressivists need a unified explanation of that in virtue of which our cognitive acts and attitudes have propositional content.

By 'unified', I mean that this explanation should hold for all domains of propositional thought. Because expressivists must hold that some propositions are not purely representational, this general story must appeal to some other, nonrepresentational (or at least not purely representational) unifying property of propositional thought. By 'in virtue of', I mean a grounding explanation for why some act or attitude has propositional content.

Using Ridge's theory as an illustrative example, suppose that we fully cash out the notion of 'purely practical' in such a way as to successfully delineate the subclass of normative propositions from the broader class of normative cognitive act types. I want to suggest that it is unclear whether this could adequately *explain* the individuation conditions for normative propositions. To be clear, the problem is not that normative propositions are distinct in kind from representational propositions — since both fall under the broader heading of cognitive act types, normative and representational propositions can be understood as different species of a unified genus. And the problem is not a lack of uniformity as such — after all, expressivism is premised on the idea that there is an important disunity between normative and representational thought. Rather, the problem is that in each case we have a completely different explanation regarding *that in virtue of which* each respective domain of thought is propositional.

This is a problem because the account implies a conjunction of the following form: (a) event type *R* can play the role of propositions in virtue of the *F*-properties of representational thought; and (b) event type *N* can play the role of propositions in virtue of the *G*-properties of normative thought. This seems strange. Given that *R* and *N* play the same role (objects of attitudes), is it plausible that they could both do this for completely different reasons? Should we not expect some unifying feature of each domain to explain how *R* and *N* could play the same role and stand in the right sorts of relations to each other? For example, any arbitrary proposition must be able to stand in the right sort of inconsistency, entailment, and independence relations to any other arbitrary proposition. This is true regardless of whether the propositions in question are normative or representational. In the absence of any unifying properties common to each domain of thought, it would be a coincidence or unexplained fact that the propositions of each domain are apt to play the role that they do. And if there is some further underlying common feature that does explain this, then (a) and (b) do not provide the full explanatory story.

Perhaps one might respond that while it would be *desirable* to have a unified explanation, there is no reason to think that this is a *requirement*. After all, the thought goes, there is something special about normative thought and discourse, and so there is no reason to rule out an explanatory bifurcation along

more traditional expressivist lines. Although nothing I have argued strictly rules out this approach, I think there are reasons to believe that a unified explanation is indeed a requirement. First, expressivists about the normative domain are often sympathetic to expressivism in other domains of discourse. On the assumption that there would be strong reasons to posit propositions in these other domains, we would no longer have an explanatory bifurcation, but a distinct explanation for each domain. Second, and relatedly, one of the motivations for positing normative propositions was to avoid unacceptably ad hoc explanations about unified phenomena (e.g., quantification over attitude contents, treatment of modals, etc.). However, by providing distinct explanations of the shared properties of different kinds of propositions, we simply introduce new ad hoc explanations at a different explanatory level.

If we accept the unity requirement, a diagnosis of the failure of Ridge's theory is that it retains too much of the representationalist paradigm of Soames' theory while trying to break away from it in a select case. Ridge is committed to explaining the logical properties of normative propositions as derivative from their representational properties (2014: 144ff). However, act types such as *predicating humanity of Plato on Thursday* possess exactly the same kind of representational properties, but they do not plausibly stand in consistency and entailment relations. By explaining the identity conditions of propositions in terms of what and how they represent, we provide no principled grounds for accepting Ridge's normative propositions while rejecting Soames' deviant examples. Both are cases of impurely representational acts of predication. In light of this, the representationalist paradigm for explaining propositional content should be rejected by expressivists.

(As an aside, other commentators also seem to miss the central explanatory role that Soames assigns to representation. For example, Davis (2019: 7) complains: "To judge that water is wet, [Soames] said, is to entertain the proposition in a specific way. To imagine that water is wet is to entertain it in another way. Judging and imagining are specific ways of entertaining a proposition. This does not provide a reason to identify a proposition with entertaining rather than judging or imagining, however. For if entertaining P is a proposition, and judging and imagining P are specific ways of entertaining P, then judging and imagining P would have to be a [sic] more specific propositions." However, given that Soames thinks that propositions are purely representational acts, this clearly rules out judging or imagining P as being more specific propositions. For the fact that P is judged or imagined does not contribute to the representational content of P. Moreover, the purely representational constraint *explains* why entertaining P but not judging P or

imagining P are propositions. As a purely representational act, entertaining P captures what judging P and imagining P have in common. So Davis' objection fails because he fails to take into account the central explanatory role of representation in Soames's theory.)

Although I have used Ridge's theory as an illustrative example, the unity requirement supplies a general constraint on any expressivist theory of propositions. But what implications does this have for the prospects for an expressivist theory of cognitive propositions? Extant theories of cognitive propositions assume a broadly representationalist explanatory framework. However, it is not obvious that a representationalist framework is essential to the basic idea that propositions are types of cognitive acts or events. In the next section, I outline a novel way of conceiving cognitive propositions that does not presuppose a representationalist framework that might provide the basis for an expressivist theory of cognitive propositions.

## 2.4 Rethinking predication

In the remainder of this chapter, I suggest that expressivists might embrace the idea that propositions are cognitive acts of *predication* by conceiving of predication as a kind of *categorisation by concepts*. I'll argue that this conception can apply both to representational and nonrepresentational thought. As such, it provides a suitable starting point for an expressivist theory of cognitive propositions that respects the unity requirement. It is only a starting point, however, because conceiving of predication in this way does not entail any particular grounding explanation of the identity conditions of such acts. It is, rather, a way of conceptualising the target phenomenon that is compatible with expressivism and suitably unified across domains. I then explain how conceptual role accounts of content determination might be utilised by expressivists to provide a suitable grounding explanation of the identity conditions of acts of predication. This then leads to the discussion of the next chapter, which examines the prospects for conceptual role expressivism more generally.

### *2.4.1 Predication as categorisation*

We saw that Soames began his account with an antecedent understanding of entertaining a proposition and then built his theory around this. Because entertaining a proposition is taken by Soames to be a single unified act type (representing states of affairs), this is apt for providing a unified explanation

of that in virtue of which our cognitive acts have propositional content. Namely, the concrete representational activity of agents explains why acts of entertaining propositions have propositional content. As a first step towards an expressivist theory of cognitive propositions, I propose that we follow suit and find some alternative general understanding of entertaining a proposition and then build our expressivist theory around that. My proposal is that the act type of predication should remain central, but that if we characterise predication in terms of *categorisation by concepts*, as opposed to the attribution of properties and relations, then we have an antecedent understanding of entertaining a proposition that is compatible with expressivism. While the proposed account of predication is compatible with expressivism, it does not entail expressivism either. For that, we need some additional nonrepresentational story about normative concepts that meets the unity requirement.

Let's begin first with the idea that the cognitive act of predication is an act of categorisation. This seems both plausible and intuitive. When we predicate something of something else in thought or in speech, we group the object of predication with other things we so predicate. If propositions are acts of predication, as the cognitive act view suggests, then we can think of propositions as ways of categorising. I have left this description deliberately vague as there are at least two distinct ways we can conceive of categorisation in thought and speech. I call these the properties-first view and the concepts-first view.

According to the properties-first view, categorisation is something fundamentally explained in terms of properties and relations. This seems to be the view endorsed by contemporary cognitive act theorists. For example: "Acts of predication are acts of sorting things into groups. When you predicate a property of an object, you sort that object with other objects in virtue of their similarity with respect to the property" (Hanks 2015: 64). If propositions are acts of predication, it follows from this that propositions are ways of categorising. Thus, in predicating (say) yellow of my scarf, I perform a mental act of grouping my scarf together with other objects, such as rubber ducks, daffodils, and Rothko's *No 14 No 10*. According to Hanks, this involves sorting these objects in virtue of their similarity to the property of *yellowness*. Thus, according to the properties-first view, properties and relations feature in the individuation conditions for acts of categorisation.

Theorists like Hanks and Soames often write as if notions like predication and categorisation simply presuppose a properties-first view. If this were correct, then thinking of predication in terms of categorisation would do little to help the expressivist. However, I think it should be clear that the notion of

categorisation does not presuppose the properties-first view. After all, it is presumably coherent to be a nominalist who denies the existence of any properties without thereby denying that there are any true acts of categorisation. However, this would seem to follow if the notion of categorisation *presupposed* a properties-first view.

Moreover, it seems to me that there is a clear sense in which acts of normative predication involve sorting things into groups (the good, the right, etc.) that expressivists can and should accept. For example, when I judge [that it is wrong to skip the queue], there is a sense in which I am sorting the act of skipping the queue with other acts, such as murder and lying. Moreover, we can think of this kind of sorting as independent from *endorsing* that sorting or judging that sorting to be true. Thus, when I imagine [that it is wrong to skip the queue], we can think of this as involving the same act of sorting as my judging [that it is wrong to skip the queue]. It is just that the categorisation involved in this act is not endorsed or affirmed as true.[3] Again, it seems to me that there is a theoretically neutral sense of categorisation on which the expressivist can and should accept the above characterisations of normative thought.

Hence, I want to I propose that expressivists make sense of this by adopting the concepts-first view, according to which categorisation is something fundamentally explained in terms of concepts. That is, we sort things into groups according to our concepts, which provide us with certain rules or principles of categorisation. Given that expressivists can and commonly do accept normative concepts in their theoretical ontology, I see no reason why they should not accept these claims. Moreover, thinking of predication this way allows us to identify acts of categorisation without appeal to properties or relations. We therefore have a way of conceiving of predication that is in

[3] This issue mirrors the debate between Hanks (2015: 35ff) and Soames (2015: 219ff) concerning whether predication is essentially *forceful* or *committal*. Whereas Soames argues that it is non-committal, Hanks argues that the idea of non-committal predication is incoherent. He then has to introduce the notion of a "cancellation context" (2015: 90ff) where a predication is "cancelled" or "overridden" when it is embedded in a logically complex act of predication. Although I'm inclined to accept Soames' view, it would take us too far afield to assess these arguments here. However, the notion of a cancellation context might be of interest to expressivists more generally in dealing with the Frege-Geach problem. One of Geach's (1965) initial complaints was that nondescriptivist views failed to respect the distinction between force and content. However, if Hanks is right, there is no such distinction. Although I cannot pursue the question here, it would be interesting to see whether cancellation contexts could explain how expressive contents might embed in logically complex constructions without subjects expressing the attitudes such contents would express in atomic assertoric contexts.

principle compatible with expressivism. However, this is just to give a characterisation of the target phenomenon. Expressivists also need to provide a suitably nonrepresentational explanation of what individuates one act of categorisation from another.

Moreover, while the concepts-first view of categorisation is compatible with expressivism, it is also compatible with a representationalist view of normative thought. For all that has been said, categorising according to concepts might be explained in terms of the properties and relations denoted by such concepts (compare this to the worry about orthodox Fregean views of propositions discussed in section 2.1.1). If this is true of normative concepts, then the concepts-first view will be of no help to expressivists. However, this possibility should not deter expressivists from developing a concepts-first approach to predication. This is because quite independently of providing a theory of cognitive propositions, expressivists already deny that normative concepts are explained in terms of what those concepts represent or refer to. What is needed is a theory of normative concepts that can explain how acts of categorisation are (i) suitably nonrepresentational and (ii) apt to play the object-of-attitude role for normative attitudes.

In the remainder of the chapter, I explore the idea that conceptual role accounts of content might provide expressivists with some helpful resources for explaining predication. I argue that there are a number of features of conceptual role accounts that appear prima facie promising for explaining (i) and (ii). In the next chapter, I will take a more critical approach to expressivist theories that embrace conceptual role accounts of content. But given that a number of prominent expressivists are independently sympathetic to conceptual role accounts of content (e.g. Blackburn 2006; Horwich 2010; Gibbard 2012; Båve 2013; Köhler 2017; Sinclair 2018), this might seem an attractive approach for expressivists.

### *2.4.2 Toward a conceptual role account*

In the next chapter, we will take a more detailed look at conceptual role accounts of content. Here, I briefly sketch one particular approach to conceptual role theory to illustrate how it might be combined with the picture of conceptual predication suggested above. According to this approach, concepts are abstract objects individuated by their possession conditions (see Peacocke 1992; Wedgwood 2007; other approaches to conceptual role theory include Harman 1973; Field 1977; Block 1986; Horwich 1998). The possession conditions for a concept specify the transitions to and from mental states involving that concept that an agent generally makes, is disposed to make, or

is rationally committed to making. The idea is clearest when applied to logical concepts. For example, Peacocke (1992: 6) proposes that the concept of conjunction is that concept *C* to possess which an agent must find the following transitions primitively compelling (i.e. not derived from or answerable to anything else):

$A(C)$ $\quad$ $p, q \rightarrow p\ C\ q$ $\quad$ $p\ C\ q \rightarrow p$ $\quad$ $p\ C\ q \rightarrow q$

At first glance, this might look promising to the expressivist because $A(C)$ individuates the concept of conjunction not in terms of what it represents, but in terms of its relational role in cognition, namely the inferences involving conjunction that we find primitively compelling. In the next chapter, I argue that things are in fact more complicated for the expressivist. But I will put this complication aside for now. What is important here is that the above formula for identity conditions for concepts makes no appeal to their representational or referential properties.

How do concepts fit into the cognitive act view? Concepts are components of thoughts. If concepts are abstract entities, as we are presently assuming, then they are components of propositions. If propositions are cognitive acts, it follows that concepts are components of cognitive acts. A natural way to understand this thought is that propositions are complex acts comprised of sub-acts, and these sub-acts are what concepts are. (The idea that concepts are event types is developed in detail by Davis [2003]; Hanks [2015] and Soames [2015] also endorse a structured view of cognitive propositions as complex acts, though because they accept a properties-first view they do not characterise these sub-acts in terms of concepts). An attractive feature of combining these accounts is that it seems to demystify how concepts as abstract entities could be involved in concrete thought. Namely, concepts are involved in concrete thought by being *tokened* in concrete thought (though these tokens are not themselves concepts).

In the simplest case, we can think of predication as a complex act type comprised of: (i) identification of the predication target; and (ii) application of a rule of categorisation to the predication target. What makes it the case that an agent applies some concept *C* rather than *C*′ and thus applies one rule of categorisation rather than another is the transitions she is disposed or ought to make to and from judgments involving the concept. (Hanks [2015: 23] and Soames [2015: 23] suggest a tripartite structure for predication which involves a distinct sub-act of identifying or "expressing" a property in addition to the sub-act of predicating that property of the target; given that the present account aims to give nonrepresentational individuation conditions for propositions, it

has no need to invoke a distinct act over and above the application of the rule of categorisation). Identity conditions for cognitive propositions can then be given in terms of the constitutive conceptual roles of their component concepts. Where [p] and [q] are cognitive propositions: [p] and [q] are identical just in case the constituent concepts in each act of predication license the same transitions to and from mental states with those contents.

One possible objection here is that the application of the rule of categorisation is not itself a distinct act in the way that identifying the predication target seems to be. So it is not clear in what sense concepts are act types according to this view. In fact, this is a problem not just for this view, but for Soames' and Hanks' respective views as well, which also appeal to sub-acts of predicating properties of the predication target that are distinct from the propositions they partly constitute (see Davis 2019: 13f for this objection to Soames and Hanks).

One possible response would be as follows. It is true that there is no act of applying a concept that can be performed independently of performing a whole act of predication, in contrast to the plausible (though not uncontroversial) claim that one can identify a predication target without performing an act of predication. However, whole acts of predication are nonetheless complex structured acts. And the way to arrive at the concept is to take the entire act of predication and abstract away the sub-act of identifying the predication target. The remainder will give us something like an "unsaturated" cognitive act type *component* that is common across all predications performed using that concept, forming a constitutive structural part of the cognitive proposition. The thought then is that concepts are act type *components*, tokens of which are necessarily parts of more complex acts.

Recall that Ridge's theory was rejected because it failed to explain why cognitive act types like *entertaining p on a Thursday* are not propositions. By contrast, the conceptual predication view does seem to explain this. According to this view, the reason why cognitive act types like *entertaining p on a Thursday* are not propositions is because the sub-act of Φ-*ing on a Thursday* has no constitutive conceptual role. That is, the act type of Φ-*ing on a Thursday* is not plausibly individuated in terms of any transitions agents are disposed to make between mental states. So the act type is not a concept and by definition therefore not a propositional constituent. Thus, *entertaining p on a Thursday* is not wholly individuated by the constitutive conceptual roles of its components and so is not a proposition.

Further, it's worth re-emphasising that concepts are here individuated as abstracta with an essential or core conceptual role. If they were individuated as mental representations, in the sense discussed in the previous chapter

(section 1.2.1), conceptual roles arguably would not be suitable for determining the propositional content of a mental state. This might be because such roles differ across times and persons or because of their holistic individuation (see, for instance, Fodor and Lepore 1991). The present account faces neither problem.

I have not defended the conceptual role approach nor offered any conceptual roles for normative concepts in particular. Rather, the approach sketched above is offered as a general explanatory framework in which the expressivist might develop an account of predication as categorisation according to concepts that is suitably nonrepresentational. Whether some version of conceptual role expressivism is ultimately viable is a question I take up in the next chapter. The point here is that *if* conceptual roles for normative concepts can be given an adequate nonrepresentational explanation, then expressivists have a suitable framework in which to explain cognitive propositions. Likewise: *if* the cognitive act view is a plausible view of propositions, then expressivists have a suitable framework for thinking about normative propositions. Given that a number of expressivists do in fact endorse some sort of conceptual role account of normative concepts, this provides an attractive framework in which to situate an expressivist theory of cognitive propositions.

Moreover, just as conceptual role expressivism can provide support for an expressivist theory of cognitive propositions, I think that the converse might hold as well. A number of contemporary philosophers prefer to think of conceptual role expressivism as a *metasemantic* view about that in virtue of which certain mental states have the content that they have, as opposed to a *semantic* view that specifies the contents of particular mental states (e.g. Chrisman 2016, 2017; Köhler 2017). However, this view leaves a residual question about the nature of the content that is determined by the metasemantic theory. It is precisely here that a conceptual role metasemantics might be supplemented with a theory of cognitive propositions to explain what the contents of mental states are. Indeed, without some such account, we have not ruled out the possibility that one's nonrepresentational metasemantics for some concept determines a robustly representational content for that concept. (I discuss this possibility at length in the next chapter). So the two approaches to thinking about content can be thought of as complementary.

This point would also apply to ecumenical expressivism, which Ridge endorses as a metasemantic thesis. Indeed, Ridge (2014: 222) also suggests that ecumenical expressivism would fit well with conceptual role accounts of content. So the general framework should be compatible with various ways of implementing expressivism. However, while the view of predication as an act

of conceptual categorisation goes some way to identifying what normative propositions could be for the expressivist, it seems fair to say that the explanatory weight of explaining how such acts could play this role will be carried by the conceptual role account. This leads us to the next chapter, which discusses the prospects of conceptual role expressivism more generally.

## 2.5 Conclusion

Cognitive propositions provide a prima facie attractive framework for an expressivist theory of propositions because normative cognitive act types are exactly the kind of thing expressivists already appeal to in their theoretical ontology. I argued that Ridge's theory of cognitive propositions failed to adequately explain identity conditions for normative propositions, and diagnosed this failure as a result of failing to respect a more general desideratum. This desideratum was the unity requirement, which states that any expressivist theory of propositions must provide a unified nonrepresentational explanation of that in virtue of which our attitudes have propositional content. I then proposed a novel conception of predication that expressivists might use to explain normative propositions and argued that conceptual role accounts of content might be well placed to provide an explanatory framework for this view of predication. This move shifts the explanatory burden to the conceptual role account, which will be topic of the next chapter.

# 3
# Conceptual Role Expressivism

Conceptual role expressivism is the label for expressivist theories that endorse a conceptual role approach to thinking about content and meaning. At the end of the previous chapter, I suggested that the fate of an expressivist theory of cognitive propositions might be dependent on the fate of conceptual role expressivism. This chapter examines conceptual role expressivism more generally. Despite a number of recent endorsements of conceptual role expressivism, there has not been a general characterisation of the view. First, then, I provide such a characterisation, explaining how extant versions of conceptual role expressivism fit this characterisation (section 3.1). Next, I raise a challenge for conceptual role expressivism in relation to defective concepts (section 3.2). More specifically, I argue that conceptual role expressivists owe us an account of why normative concepts are nondefective. I then go on to discuss deflationist responses to the challenge and argue that they are lacking (section 3.3). Finally, I propose an alternative solution that appeals to something like Gibbard's (2003) planning contents (section 3.4). However, I argue that doing so leaves conceptual role expressivism a less interesting and distinctive position than it might otherwise seem. Nonetheless, I argue that this is exactly what we should expect. I conclude by considering the implications this has for the cognitive act view suggested at the end of the previous chapter, as well as examining some similarities to Gibbard's (2012) more recent account of normative concepts.

## 3.1 Conceptual role expressivism: an overview

In this section, I will first explain conceptual role theory, and then provide a fully general characterisation of conceptual role expressivism. I will then briefly outline three competing conceptual role expressivist views in the literature and explain how they all fall under the general characterisation.

### *3.1.1 Conceptual role theory*

As I will be using the label, conceptual role theory is a general approach to explaining the contents of *thoughts* and *concepts*. Programmatically, the idea is that concepts should be explained fundamentally by their role in reasoning, belief formation, and practical deliberation. We saw in the previous chapter one particular way in which to develop conceptual role theory. Due to the wide variety of approaches that fall under the label, however, it can be difficult to say informatively at a general level what the overall approach consists in. Indeed, as we will see below, there are different conceptions of the explananda as well as the explanantia of conceptual role theory. But what all such theories have in common is a commitment to explaining mental content *relationally* in terms of certain connections between a subject's mental states involving that content and certain other features of the subject.[1] More specifically, conceptual role theories explain mental contents in terms of certain conditions under which a subject *accepts* those contents, where these conditions specify connections between states of acceptance and other states of the subject. In addition to other states of acceptance, these might also include connections from *input-states*, which might include perceptual states or worldly states in which the subject finds herself, and connections to *output-states*, which might include actions or intentions to act.

As we saw in the previous chapter, the acceptance conditions provided by conceptual role theorists are a form of *possession conditions*, which explain under what conditions a subject possesses some concept or other. So the idea is that the content of a concept is explained fundamentally by its possession conditions, where these specify certain transitions in thought that a subject is apt to make. In the previous chapter, we saw how this might work for logical concepts like conjunction. The basic idea was that to possess the concept CONJUNCTION is be disposed to accept contents on the basis of the introduction and elimination rules for conjunction. Another example might be colour concepts. Very roughly, to possess the concept RED might be to form beliefs about what is red in the presence of red things (e.g. Horwich 1998) or on the

[1] This characterisation already rules out certain views that might go by the name 'conceptual role theory', such as Brandom's (1994) inferentialism. Here the relevant "connections" are not between a subject's mental states, but rather between certain *commitments* and *entitlements* that an agent can possess within a linguistic community. Although Brandom's view is a somewhat neglected framework for developing a nondescriptive view of normative thought and discourse, it would take us too far afield to examine it here (see Chrisman 2016 for discussion). In any case, such a view does not respect the priority of thought thesis, which I am assuming throughout (see section 1.2.3).

basis of certain perceptual experiences (e.g. Peacocke 1992). The transition or set of transitions that individuates a thought or concept is its *conceptual role*.

In developing a conceptual role theory for some concept or concepts, there are a number of choice points that one must make with respect to cashing out the conceptual roles in question. For example (compare Whiting 2006; Chrisman 2017; Sinclair 2018):

- *Concept-first* vs *thought-first*. Do conceptual roles primarily apply to *concepts*, the constituents of thoughts, or to entire *thoughts*?
- *Particulars* vs *abstracta*. If conceptual roles primarily apply to concepts, are concepts *concrete mental representations* or *abstract components of propositions*?[2]
- *Naturalism* vs *normativism*. Do conceptual roles specify transitions that subjects *in fact* make or are *disposed* to make, or that subjects *ought* to make or are *rationally committed* to making?
- *Atomism* vs *holism*. Do concepts have *core* or *essential* conceptual roles, or are conceptual roles locally or globally *holistic*?
- *Solipsistic* vs *communitarian*. Are *individuals* the locus of conceptual roles, or *communities*?
- *Narrow* vs *wide*. Do conceptual roles specify transitions *only* between mental states, or also between mental states and worldly objects, properties, events, states of affairs, etc.
- *Doxastic* vs *non-doxastic*. Do narrow conceptual roles specify transitions only between doxastic states (beliefs, suppositions, perceptual experiences, etc.), or also between non-doxastic states (desires, intentions, etc.)?

Assessing the prospects for conceptual role expressivism will require taking a stand on some of these issues with respect to characterising conceptual roles for normative concepts and thoughts. Where either choice point is open to the expressivist, I will remain noncommittal. For ease of exposition, I may at times talk in more particular terms than is required. However, where some choice is required, I will always make this explicit.

### *3.1.2 Conceptual role expressivism characterised*

I argued in the previous chapter that conceptual role accounts of content should look prima facie attractive to expressivists. According to expressivism, normative concepts should not be explained fundamentally in terms of their

[2] Note that this sense of 'representation' is distinct from the sense in which expressivists deny that normative thoughts are 'representations' of reality — see section 1.2.1.

representational properties, but rather by their role in practical deliberation. According to conceptual role theory, concepts in general should not be explained fundamentally in terms of their representational properties, but by their role in reasoning, deliberation, and belief formation. Insofar as conceptual role views provide an independently attractive framework for explaining meaning and content, it would seem that expressivists would do well to embrace them.

The basic idea of conceptual role expressivism, then, is to explain normative concepts in terms of their conceptual roles in a way that respects the central commitments of expressivism. Although specific accounts will vary, both in terms of the choice points above as well the details of the conceptual roles for different normative concepts, I propose that conceptual role expressivism is characterised by its commitment to the following three theses:

(A) Conceptual roles for normative concepts do not specify any particular worldly or perceptual input conditions.
(B) Conceptual roles for normative concepts specify some connection to conative or affective states.
(C) Conceptual roles for normative concepts do not determine robustly representational content for normative thoughts.

(I focus my attention here on thin normative concepts; thick normative concepts would plausibly have some particular input conditions.) Below, I will show how existing versions of conceptual role expressivism go about meeting these constraints. However, first some general comments on each.

Commitment (A) is a consequence of the expressivist's claim that in making a normative judgment, one is not responding to normative reality. In order to possess some normative concept, expressivists hold that it is sufficient to know how to deploy that concept in practical reasoning. There are no constraints concerning the worldly or perceptual conditions one must satisfy in accepting some normative content in order to possess some normative concept. This is not to deny that subjects do in fact respond to features of reality when deploying normative concepts. But there is no distinctive domain of reality that all subjects must respond to in order to *possess* normative concepts. Among other things, this is meant to explain how different subjects can systematically respond to different features of reality in applying a normative concept without thereby employing distinct concepts. Thus, for example, two subjects might have radically divergent conceptions of what makes up the subject matter of moral wrongness, for example a Kantian who thinks it is wrong to violate the categorical imperative and a moral egoist who thinks it is wrong for anyone to not pursue their own self-interest. These two subjects will respond

to distinct aspects of reality in applying the concept MORALLY WRONG, but they still seem to disagree about what is wrong. Consequently, expressivist conceptual roles for normative terms will be *narrow* in scope.

Commitment (B) is meant to capture whatever it is that expressivists want to claim is the distinctive function or purpose of normative concepts in our cognitive economy. Traditionally, expressivists have endorsed some form of motivational internalism, which claims that there is some kind of necessary connection between normative judgments and being motivated to act in accordance with those judgments. On the present approach, this can be cashed out in terms of normative concepts involving conceptual connections between normative judgments and desires, intentions, or other action-guiding states. Expressivists might also claim that normative concepts are conceptually linked to affective states as well, such as certain reactive attitudes. No doubt different normative concepts will be linked to different kinds of attitude, and the exact details of these connections will be controversial. What the expressivist will not take as controversial, however, is that normative concepts are distinctive in their connections to *non-doxastic* attitudes. This also suggests a more *atomistic* or at least *locally holistic* approach to characterising normative conceptual roles, though (B) remains compatible with a more global holism. However, I will assume for simplicity that normative thoughts and concepts have a core conceptual role.[3]

As we will see below, existing conceptual role expressivists seem to apply their account only to *practical* normative concepts. One might worry that commitment (B) fails to apply to *epistemically* normative concepts. This is because it is does not seem that judging (say) [that I ought to believe p] can be connected to conation in the same way as action, because I cannot decide or form an intention to believe anything. Because expressivists typically aim to explain normative thought in general, it is a problem if their theory only applies to practical normativity. Epistemic normativity has been somewhat neglected by expressivists, at least in terms of providing detailed accounts of what epistemic normative judgments are like, and it would be too big an issue to adequately address this here (see Chrisman 2012 for an overview of the issues). However, it can be pointed out that commitment (B) specifies only that there be *some* connection to conative or affective states. It does not say anything about what those states are. So perhaps the expressivist can say that the acceptance conditions for [I ought to believe p] do not involve a connection to an *intention to believe p*, but rather something more complex, such as a

[3] As we saw in the previous chapter, a holistic conceptual role theory might have a hard time individuating propositional contents, as there is bound to be variation interpersonally and intrapersonally across time — see section 2.42.

disposition to feel a corrective response and to intend to take self-corrective measures when I do not believe p (compare Gibbard 2012: 173), or to criticise myself and others in discussion about what to believe, and so on. In sum, commitment (B) is at least in principle compatible with expressivism about epistemic normativity as well as practical normativity.

Moving on to commitment (C), conceptual role theorists typically hold that the conceptual role of a concept or thought in some sense determines its content. A more orthodox reading of this claim is that the conceptual role of a mental state determines its *representational content*, i.e. the worldly relata that the mental state is in some sense *about*. Clearly, conceptual role expressivism must deny that this is true of at least normative concepts, and this is why commitment (C) is also a requirement of the view. As we will see below, this has usually been taken to involve a commitment to some kind of minimalist or deflationary notion of content. But as we saw in the previous chapter, it is also open to the expressivist to adopt some kind of substantive notion of content, so long as it is suitably nonrepresentational. Insofar as there is good reason to posit normative propositions, as I argued in Chapter 1, then conceptual role expressivism should adopt a theory of propositions akin to that developed in Chapter 2. However, because the issues discussed in this chapter apply to conceptual role expressivism more broadly, I will not emphasise this point in this chapter but instead focus on commitment (C) more generally.

(It's worth emphasising that commitment (C) is itself a substantive and independent commitment that does not simply follow from (A) and (B). Kalderon [2007: 82ff] examines and rejects a number of arguments that attempt to derive something like (C) from something like (A) and (B). However, if we take (C) to be an independent commitment motivated on independent grounds rather than derived from (A) and (B), then Kalderon's arguments do nothing to undermine it.)

### *3.1.3 Three theories*

In the remainder of this section, I examine three different versions of conceptual role expressivism that can be found in the literature, and explain how each version adheres to commitments (A), (B), and (C). As my aim here is simply to show how existing versions of the view respect these commitments, for now I will suppress critical discussion of these views. For ease of exposition, I have regimented the terminology to that used above. However, these terminological alterations do not change the substance of the views.

The first version comes Paul Horwich (2010), who holds that to believe a proposition [p] is to accept a sentence that expresses [p], where acceptance

involves relying on that sentence in theoretical and practical reasoning. Horwich argues that the meaning-constituting property of a word is its property of having some core conceptual role. This core conceptual role, together with other factors such as the environment and the meaning-constituting roles of other words, explain when sentences containing the word are accepted and rejected (hence for Horwich, conceptual roles are *solipsistic*, *naturalist*, *concept-first*, and *subjectivist*). Thus, for Horwich, the propositional content of a belief is determined by the meaning-constituting properties of the components of its sentential object, together with its syntactic structure. Because Horwich denies that the meaning-constituting properties of words are relations between those words and what they stand for, including semantic vocabulary such as truth and reference, he denies that *any* meaning-constituting properties determine a robustly representational content. Hence, his view respects commitment (C) by being globally nonrepresentationalist.

He then goes on to propose the following example of a conceptual role for the ought of practical rationality (2010: 188):

(1) (i) S believes [that he ought to do X] ↔ S is strongly inclined to do X
(ii) S denies [that he ought to do X] ↔ S is not inclined to do X

(Throughout this chapter, I use arrow notation to specify the transitions between mental states constitutive of the conceptual role in question.) Hence, Horwich's view respects both (A) and (B), as the core conceptual role for the ought of practical rationality does not specify any worldly or perceptual input conditions and is constitutively connected to an action-guiding attitude. It's worth noting that Horwich (2010: 189ff) does not place too much credence in the specific proposal, accepting that an adequate characterisation will no doubt be more complicated. However, he nonetheless endorses the general approach, and it seems that any complication will still respect (A), (B), and (C).

The next example comes from Sebastian Köhler (2017), whose primary aim is to provide a deflationary account of propositional content. However, he does not arrive at deflationism by the sort of sententialist view of belief endorsed by Horwich. Instead, Köhler aims to provide an account of the nonrepresentational function of *content attributions* that does not invoke any entities as the referents of 'that'-clauses. In a nutshell, Köhler's idea is that the sentence mentioned in a 'that'-clause serves as an *illustrative example* of certain "basic explanatory properties" of the belief being attributed (2017: 198). Specifically, it serves as an illustrative example of the belief's *conceptual role*, where Köhler understands this as the subset of its total causal-functional role that relates to perceptual, inferential, and action-producing processes (hence

for Köhler, conceptual roles are *thought-first* and *naturalist*). Declarative sentences are then understood as codifying information about the conceptual role of the belief state they express.

By explaining the role of content attributions without appealing to contents, Köhler's account is committed to (C). He then proposes that representational states are those whose conceptual roles include sensory input conditions but no direct relation to action production, and that conative states are those whose conceptual roles include no sensory input conditions but directly relate to action production. Although Köhler offers no specific conceptual roles for normative mental states, he suggests that normative judgments are conative states in the above sense. However, these conative states are nonetheless belief states in virtue of their having sufficiently rich inferential transitions to and from other belief states. Hence, the account respects commitments (A) and (B).

Finally, Sinclair (2018: 109) proposes that expressivists follow Wedgwood's (2001) conceptual role for the concept X IS (ALL THINGS CONSIDERED) BETTER FOR ME TO DO AT T THAN Y as that concept *B* such that:

(2) S accepts $B(x,y,me,t) \rightarrow$ S has conditional intention to do x rather than y at t.

If (2) is the conceptual role for the normative concept *B*, it should be clear that it respects commitments (A) and (B). However, although Sinclair endorses (2), it's unclear what other parts of Wedgwood's view he endorses. For example, because Wedgwood is a normativist about conceptual roles, Sinclair claims that (2) derives from the following norm: If acceptance of $B(x,y,me,t)$ is rational then a conditional intention to do x rather than y is rational. However, Wedgwood would actually claim that (2) derives from the distinct norm: If acceptance of $B(x,y,me,t)$ is *not irrational* then *not* having a conditional intention to do x rather than y at t is *irrational* (compare Wedgwood 2007: 84). Further, whereas Wedgwood is explicitly committed to an abstract Fregean view of concepts, Sinclair does not acknowledge the distinction between concepts as concrete representations or abstracta and seems to waver between the two.

A more pressing difference between Sinclair and Wedgwood is that Wedgwood is a robust realist about normativity. As we will see in more detail below, Wedgwood thinks that (2) determines a unique worldly referent for *B*. So Wedgwood's own view does not respect commitment (C). By contrast, Sinclair (2018: 110) believes that he can respect (C) by endorsing a deflationary view of truth and reference. I will return to this, and to important features of Horwich's and Köhler's views, in more detail below. However, I hope it should be clear how the otherwise disparate expressivist views advanced by these

theorists respect the three core commitments of conceptual role expressivism, thereby supporting its claim to full generality.

As we have seen, conceptual role theory provides an attractive framework for expressivism insofar as it individuates mental states with distinct contents relationally in terms of their role within one's cognitive economy rather than in terms of what those mental states represent. According to conceptual role expressivism, normative thoughts and concepts are individuated by their nonrepresentational, practical, core conceptual roles. Although these three examples by no means exhaust all versions of conceptual role expressivism, they nonetheless provide examples of prominent expressivist views that make explicit appeal to the resources of conceptual role theory. (Horwich does not actually use the label 'conceptual role theory', but that is terminological —see Whiting [2006]; Block [1998].) However, they nicely illustrate how conceptual role expressivists might cash out commitments (A), (B), and (C). Moreover, given the wide variety of differing choice points, they show how different versions of conceptual role expressivism can respect these commitments in theoretically diverse ways. In the next section I will raise a challenge to conceptual role expressivism. Because I take the challenge to be fully general, I will articulate the challenge in abstract terms.

## 3.2 A challenge for conceptual role expressivism

In this section, I first introduce the notion of a *defective* concept and explain why this notion poses a challenge for conceptual role expressivism. I then argue that the challenge arises from expressivism's distinctive commitments by showing how the challenge can be solved by descriptivist conceptual role theories.

### *3.2.1 Defective concepts*

Any theory of concepts must explain their semantic and logical properties. Conceptual role expressivism must therefore explain the semantic and logical properties of normative concepts. Specifically, it needs to explain how normative concepts can meaningfully embed in complex thoughts and participate in genuine inference, given commitments (A), (B), and (C). With this in mind, let a candidate concept be *defective* just in case it cannot meaningfully embed in complex thoughts or participate in genuine inference. The challenge for conceptual role expressivism is to explain why normative concepts are not a species of defective concepts.

The challenge arises because normative concepts as conceived by conceptual role expressivism bear a number of similarities to other concepts that are defective. I will argue that without any explanation of why normative concepts are not defective, the expressivist must accept either (i) that normative concepts are also defective, or (ii) that certain clearly defective concepts are in fact not defective. While perhaps early emotivists might happily accept (i) (Ayer sometimes describes moral concepts as 'proto-concepts'), both claims are highly implausible.[4] So expressivists need to explain why normative concepts are not defective.

In what follows, I provide two examples of defective concepts. Before this, however, let me say something more about what it is for a candidate concept to be defective. First, it is worth stressing that I am only concerned here with defectiveness in the narrow sense defined above. There are, of course, a number of other ways in which a concept might be described as "defective", but my concern here is only with a concept's capacity for participating in embedding and inference. Second, it is important to distinguish between the property of being defective and the explanation for why a concept is or is not defective. In targeting the phenomenon, I do not wish to presuppose any particular explanation of what makes concepts defective.

Third, where I have been careful, I have been using the qualification of whether a *candidate* concept is defective or not. This is because depending on one's view of concepts, one might maintain that a "defective concept" is in fact no concept at all. That is, one might think that the class of concepts is not divided into those which are defective and those which are not. Rather, the capacity to participate in inference and embedding is a necessary condition for something's being a concept, and thus a "concept" that lacks this capacity is a *spurious concept*. On this view, the challenge for conceptual role expressivism

[4] Derek Shiller (2018) makes an interesting suggestion that, in effect, expressivists should claim that moral concepts are defective. First, he supposes the representational theory of mind according to which concepts are mental particulars. Next, he argues that moral concepts are characterised by "*non-representational proper functions that operate in the context of straightforwardly predicative moral judgments*." (2018: 432 original emphasis) However, he claims that although moral concepts are *syntactically* compositional and so can embed and occur in processes of reasoning and so on, such combinations are in fact *semantically* incoherent. To support this idea, he introduces the notion of a "spandrel context" which explains why evolution may have given us concepts with properties that outstrip the requirements of their designed function. As Schiller admits, however, the view is highly revisionary of our ordinary conception of morality and does not provide the sort of vindication of morality that quasi-realist expressivism aims for. As I take this to be one of expressivism's most attractive features, I leave this sort of view aside.

is to explain why normative conceptual roles successfully individuate any concept whatsoever. Although it is a substantive issue as to which is the correct way of framing the challenge, the arguments discussed below apply mutatis mutandis to either formulation. For ease of exposition, I will therefore continue to talk about defective concepts without taking a stand on whether defective concepts are genuine concepts that have the property of being defective, or whether they are spurious concepts, descriptions of which are merely a façon de parler.

With these qualifications out the way, I will now provide two examples of defective concepts. The main focus of the challenge for expressivism will be the second example, but it will be helpful to first examine a more familiar example to get a grip on the notion of a defective concept. The classic example comes from Arthur Prior's (1960) 'tonk' connective. Transposing Prior's example from the linguistic to the conceptual mode, suppose we define TONK as that concept $T$ to possess which a subject $S$ is apt to accept the following transitions:

$A(T)$ S accepts $p \rightarrow$ S accepts $pTq$
S accepts $pTq \rightarrow$ S accepts $q$

TONK therefore functions as a logical connective, with $A(T)$ functioning much like Peacocke's possession conditions $A(C)$ for CONJUNCTION. It might therefore seem that like $A(C)$, $A(T)$ explains the rules of inference governing TONK, as well as how TONK can meaningfully embed. However, a moment's reflection shows that any such impression would be an illusion.

To see why, first note that given $A(T)$, it apparently follows that anyone who possesses the concept TONK can derive any arbitrary proposition from any other arbitrary proposition. So much follows simply from the meaning of TONK. But it is surely not possible to genuinely infer the truth of any arbitrary proposition from another simply by possessing TONK. Perhaps there is some loose sense of 'inference' such that one could 'derive' $q$ from $p$ via TONK in that one can recognise the rules that individuate TONK and consciously follow the rules that lead from $p$ to $q$, perhaps drawing 'truth tables' for TONK and writing out TONK 'derivations'. But this weak sense of recognising and following the rules specified by $A(T)$ is not plausibly sufficient for one to genuinely infer anything. The lesson that Prior thought we should draw from this example is that stipulating the meaning of an expression in terms of its role in inference is not sufficient to bestow genuine meaning. In the present context, the lesson is that stipulating the meaning-constituting conceptual role of a concept in terms of its role in inference is not sufficient to explain that concept's capacity to participate in inference and embedding.

A number of responses have been given to Prior's example. For instance, the conceptual role theorist might explain TONK as defective in virtue of the fact that it provides a *non-conservative extension* of our conceptual scheme (Belnap 1962). The idea is that the rules specified by $A(T)$ are inconsistent with the rules of our existing concepts, allowing TONK to lead to contradictions. Alternatively, one might deny that TONK is a genuine concept at all because it is impossible for a subject to be disposed to accept TONK-thoughts (Horwich 1998: 133ff). This is because forming beliefs using TONK would lead to radical instability in our beliefs. So it's unclear that the example is as fatal to conceptual role theories as Prior took it to be (see Whiting 2006 for an overview).

In any case, I do not wish to examine these responses here for two reasons. First, insofar as TONK does present a problem for conceptual role expressivism, it is only because it presents a problem for conceptual role theory in general, which I'm not here calling into question. Second, as a candidate *logical* connective, the conceptual role for TONK differs in important ways to conceptual roles for normative concepts as conceived by expressivism. Specifically, TONK does not respect commitment (B). This means, I think, that the kind of responses given to Prior's example will not apply in the normative case. Nonetheless, I hope the discussion of TONK helps to clarify the notion of a defective concept.

A more worrying example for conceptual role expressivism derives from James Dreier's (1996: 44) 'is hiyo' predicate. I will first introduce the example as Drier presents it and then go on to explain its relevance to conceptual role expressivism. Suppose we use the word 'hiyo' to perform the speech act of *accosting*. For example, by uttering 'Hiyo, Bob!' in Bob's presence, I thereby accost Bob. Clearly, this isn't the kind of word that could meaningfully embed in logically complex sentences or participate in genuine inference. Most obviously, 'hiyo' does not have the right syntactic properties — 'Hiyo, Bob!' is not a declarative sentence. However, suppose we introduce the predicate 'is hiyo' whose constitutive meaning is to perform the speech act of accosting what it is predicated of. Thus, by uttering 'Bob is hiyo' in Bob's presence, I thereby accost Bob in virtue of the meaning of the sentence uttered. Because 'is hiyo' is syntactically a predicate and 'Bob is hiyo' is a well-formed declarative sentence, superficially it might appear that 'is hiyo' can meaningfully embed and participate in genuine inference. However, any such appearance would be false. Even if 'hiyo'-sentences are syntactically well-formed, we surely have no conception of what a sentence like 'If Bob is hiyo, then a Dingo is near' means. And if we have no conception of what this sentence means, then surely it is because 'is hiyo' is not able to meaningfully embed in complex sentences. In other words, 'is hiyo' is a defective predicate.

Dreier's original target in constructing his example was to show that it is not enough for expressivists to appeal to a minimalist or deflationary theory of embedding and inference to explain how a predicate with an expressive function can meaningfully embed and figure in inference. The challenge for expressivists, as Dreier sees it, is to explain the meaning of expressive predicates in such a way as to explain their aptness for embedding and figuring in inference. How does this relate to conceptual role expressivism? It might initially seem that conceptual role expressivism is isolated from Dreier's example. This is because 'is hiyo' is predicate rather than a concept. Moreover, it does not seem to correspond to any concept in the standard sense in that its meaning is specified in terms of its role in performing certain *speech acts*. By contrast, conceptual role theory individuates concepts in terms of transitions to and from *mental state*s involving that concept. Because the meaning-constituting rules for 'is hiyo' fail to specify any such transitions, they fail to individuate any concept whatsoever.

Granting this, however, it is easy to construct an example that mirrors Dreier's but is formulated at the level of concepts rather than predicates. For example, consider the following. Let HIYO be that concept *H* to possess which a subject *S* is apt to accept the following transitions:

*A*(*H*) S wants to accost x → S accepts *H*x
S does not want to accost x → S rejects *H*x

Here, we have clearly defined acceptance rules for when to form HIYO-beliefs. However, just as we have no conception of what the sentence 'If Bob is hiyo, then a dingo is near' means, we surely have no conception of what the content of the thought IF BOB IS HIYO THEN A DINGO IS NEAR is. Thus, although superficially it might appear that HIYO can embed in complex thoughts and figure in genuine inference, this appearance would be false. So HIYO is a defective concept.

This is a problem for conceptual role expressivism because the meaning-constituting possession conditions for HIYO bear a strong resemblance to the conceptual role expressivist's meaning-constituting possession conditions for OUGHT and BETTER THAN. After all, the conceptual role for HIYO (A) does not specify any worldly or perceptual input conditions, (B) specifies an essential connection to conative states, and (C) is not plausibly interpreted as denoting any particular property (at least simply in virtue of its particular conceptual role). So HIYO respects the three core commitments for normative concepts constitutive of conceptual role expressivism. However, it is highly plausible that HIYO is a defective concept. But it is also highly plausible that OUGHT and

BETTER THAN are not defective concepts. At least, insofar as expressivists aim to vindicate our normative practices rather than debunk them, then expressivists will not accept that normative concepts are defective. Conceptual role expressivists therefore owe us an explanation of why our normative concepts are not defective when similar candidate concepts like HIYO are.

Before examining some responses on behalf of conceptual role expressivism, I first want to sharpen the challenge by showing how descriptivist variants of conceptual role theory can explain why normative concepts are not defective in terms of their representational properties.

### *3.2.2 Conceptual role and representational content*

One response to the challenge might be that insofar as HIYO presents a challenge to conceptual role expressivism, it is because it presents a challenge to conceptual role theory in general. In other words, HIYO is everyone's problem. More generally, it seems, any version of conceptual role theory should provide some account of the difference between conceptual roles that bestow genuine meaning and conceptual roles that characterise defective concepts. So one might think that the conceptual role expressivist is no worse off than any other conceptual role theorist. And although this would certainly count against conceptual role expressivism, my aim in this chapter was not to assess the overall viability of conceptual role theory on its own terms. Rather, it was to examine whether expressivists would do well to adopt some form of conceptual role theory insofar as it is an independently attractive framework.

It certainly seems right that any conceptual role theory should have something to say about defective concepts. However, classical conceptual role theories typically *do* have something to say about defective concepts. But unfortunately for the expressivist, the standard explanation of defectiveness results from combining conceptual role theory with a representationalist picture of conceptual thought. Although conceptual role theory fundamentally explains concepts in terms of their conceptual roles, many of its proponents also hold that concepts have representational properties. Specifically, they hold that a concept's conceptual role *determines* it's representational properties. Moreover, we will see that this provides a straightforward explanation of why concepts like HIYO are defective. However, the representationalist explanation also predicts that normative concepts as conceived by expressivism are defective. As such, explaining defective concepts provides an acute challenge for conceptual role expressivism in particular.

This can be easy to overlook because the representationalist explanation of defectiveness is not a response to a challenge, but is rather a consequence of

already held assumptions about the representational nature of conceptual thought. The dialectic here should be familiar from the previous chapter. There, we saw how Soames' already held assumption that propositions are *purely representational* cognitive acts predicted the correct identity conditions for propositions. By denying this assumption, Ridge's theory robbed itself of the resources needed to capture the identity conditions for normative propositions in a principled way. Similarly, by denying the assumption that conceptual roles for normative concepts determine any kind of robustly representational content for the thoughts in which they figure, conceptual role expressivists rob themselves of the resources needed to explain defective concepts. To see this, however, I will first need to set out exactly what these resources are.

I will begin by providing some evidence that many prominent conceptual role theorists do in fact accept a representationalist picture of conceptual thought. In the previous chapter, I discussed Christopher Peacocke's (1992) conceptual role account of logical constants. Although Peacocke individuates concepts in terms of their possession conditions and not in terms of what they denote, he does not deny that concepts are denotational or that reference plays an explanatory role in his theory of concepts. For example (ibid: 17):

> As a matter of principle, the level of reference is inextricably involved with concepts, as understood here. Concepts are individuated by their possession conditions; the possession conditions mention judgments of certain contents containing the concepts; judgment necessarily has truth as one of its aims; and the truth of a content depends on the references of its conceptual constituents. It would be wrong, then, to regard the referential relations in which concepts stand as grafted onto a structure of concepts that can be elucidated without any reference to reference. Referential relations are implicated in the very nature of judgment.

So, for Peacocke, that concepts are representational is entailed by a representationalist conception of judgment.

Another conceptual role theorist who endorses a representationalist picture is Ralph Wedgwood (2007). This will be of particular interest in what follows, because, as we saw above, Wedgwood provides conceptual roles for normative concepts that respect commitments (A) and (B) while denying (C). As such, this will provide an acute presentation of exactly how commitment (C) robs expressivists of the resources available to representationalist views for explaining defective concepts. For now, however, I simply want to emphasise that Wedgwood's commitment to the representationalist picture of concepts is presupposed from the outset (ibid: 81, original emphasis):

> What is essential to concepts is just that they play a *representational* role: the nature of a concept consists purely in the contribution that it makes to the nature of the thoughts in which it appears; and such thoughts are nothing more than ways of representing some possible state of affairs. […] [T]he essential features of a concept must all be relevant to determining the concept's reference or semantic value.

Finally, consider the sort of conceptual role picture endorsed by Ned Block (1986). Block's picture is in many ways very different to that of Peacocke and Wedgwood. Whereas the latter two endorse a broadly Fregean view of concepts and propositions, Block is more concerned to adopt a psychological view of content as part of a more empirically informed explanation of cognition and the mind. However, he takes the initial desideratum on any such account to explain "the relation between meaning and truth/reference" (ibid: 616) and argues that (ibid: 643f):

> [W]hat theory of reference is true is a fact about how referring terms function in our thought processes. This is an aspect of conceptual role. So it is the conceptual role of referring expressions that determines what theory reference is true. Conclusion: conceptual role factor determines the nature of the referential factor.

Although Block does not engage with the issue of defective concepts, these comments make it clear that embracing conceptual role theory does not rule assigning an explanatory role to the level of reference in a theory of concepts. However, I'll continue to focus on the kind of framework employed by Peacocke and Wedgwood, as this brings the issue more sharply into focus.

We see, then, that conceptual role theorists like Peacocke and Wedgwood are still committed to a representationalist picture of conceptual thought. I will now describe how this picture explains defective concepts. Peacocke and Wedgwood argue that any theory of concepts needs not just a theory of possession conditions but also a *determination theory* that explains how the reference of a concept is fixed by its possession conditions. As well as explaining how the conceptual role aspect of their view squares with the representationalist aspect, a determination theory also aims to explain why it is *appropriate* or *correct* for subjects to make the transitions specified by the theory of possession conditions. If there is no reference that can explain the appropriateness of correctness of the transitions specified by some concept's conceptual role, then we can explain that concept as defective.

There are some minor differences between how Peacocke and Wedgwood develop their determination theories, so I will take each in turn. For Peacocke, correctness is ultimately cashed out in terms of *truth* and *reference*. For logical concepts, a determination theory assigns to the concept whatever reference would make the transitions specified in its conceptual role *truth-preserving*. Thus, recall Peacocke's theory of possession conditions for CONJUNCTION, according to which CONJUNCTION is that concept *C* to possess which a subject S is apt to make the following transitions:

$A(C)$ S accepts $p$ and S accepts $q \rightarrow$ S accepts $pCq$
S accepts $pCq \rightarrow$ S accepts $p$
S accepts $pCq \rightarrow$ S accepts $q$

The determination theory then assigns as the reference of *C* whatever would make the transitions specified by $A(C)$ truth-preserving, here the classical truth function of conjunction (Peacocke 1992: 18). Because one preserves truth in one's beliefs in accepting $pCq$ on the basis of accepting $p$ and accepting $q$, we can say that it is correct or appropriate to transition from accepting $p$ and accepting $q$ to accepting $pCq$. The determination theory for empirical concepts (e.g. RED) will take a slightly different form. This is because the possession conditions for empirical concepts will involve beliefs based on perceptual experience. The determination theories for these concepts will then assign references to concepts such that these belief-forming practices result in truth beliefs. In both cases, however, correctness or appropriateness in conceptual roles is explained in terms of truth and reference.

Importantly for our purposes, a determination theory such as Peacocke's can explain why a concept like TONK is defective whereas CONJUNCTION is not. Like CONJUNCTION, TONK functions as a logical connective with its conceptual role specifying mind-to-mind transitions about what combinations of contents to accept. However, unlike CONJUNCTION, there is no semantic assignment for TONK that would make the transitions specified in $A(T)$ truth-preserving and hence correct or appropriate for subjects to follow. Accordingly, Peacocke's theory correctly predicts that TONK is a defective concept.

By contrast, the central notion for explaining correctness for Wedgwood is not truth but *rationality*. This is driven by his theory of normative concepts. I will here focus on Wedgwood's theory of the practical all-things-considered ought. He proposes that OUGHT is the unique concept *O* to possess which a subject is rationally committed to the following (2007: 97):

$A(O)$ S accepts the first-person proposition $[O_{<me,\ t>}(p)] \rightarrow$ S makes $p$ part of $S$'s ideal plan about what to do at $t$.

Here, one's *ideal plan* is what would be one's plan if it were not affected by ignorance and uncertainty about what to do in the situation one is in at $t$; to make $p$ part of one's plan is to adopt a plan that entails $p$. It should already be clear why $A(O)$ forces Wedgwood to reject a Peacocke-style determination theory for OUGHT. This is because $A(O)$ specifies a transition from a judgment to a planning attitude. Because plans are not truth-apt, the correctness of the transition cannot be explained in terms of truth-preservation or as a true belief-forming practice. (Some might argue that plans are in fact truth-apt; however, even if they are, it seems implausible that their being so would explain why it is appropriate for subjects to follow $A(O)$.)

Instead of truth, Wedgwood (2007: 100) appeals to the "point" or "purpose" of the attitudes in question and to the corresponding rational norms that govern them. Thus, while "getting things right" in believing is having a true belief, getting things right in planning consists in something else, such as having "*genuinely choiceworthy*" plans (ibid: 101). Wedgwood then argues that if being genuinely choiceworthy is what governs our plans, then the determination theory for $A(O)$ should assign to $O$ the logically weakest property of a proposition in virtue of which it is correct to make that proposition a part of one's ideal plan. So although Wedgwood appeals to rationality rather than truth in providing a determination theory for OUGHT, his determination theory still explains OUGHT in terms of its reference.

Importantly for our purposes, a determination theory like Wedgwood's can explain why HIYO is defective whereas OUGHT is not. Although it might be rational to accost someone if one has a desire to accost that person, this is not captured by $A(H)$. Rather, on Wedgwood's framework, what is captured by $A(H)$ is a connection between wanting to accost some individual and the belief that that individual possesses the property denoted by $H$. However, it is unclear that there are any rational norms governing our desires to accost people such that there is some property that could be assigned to $H$ that would explain the correctness of the transition specified by $A(H)$.[5] If this is correct, then Wegdwood's account explains why HIYO is defective while OUGHT is not.

[5] Perhaps one could assign to $H$ the property of *being wanted to be accosted by S*. If this is correct, then HIYO might in fact be nondefective. Two comments. First, if this is correct, then it is not obviously problematic that HIYO is nondefective, initial intuitions notwithstanding. Second, this would be of no help to expressivists, because saying something similar about normative concepts would entail some form of *subjectivism* rather than expressivism.

My aim here has not been to show that these explanations are correct or without problems of their own.[6] Rather, my aim has been to show that existing strategies for dealing with defective concepts and for explaining the properties of conceptual roles more generally make reference to the representational properties of concepts. Because expressivists deny that normative concepts are representational in this sense, they cannot pursue this strategy in any form. So the challenge to explain defective concepts is particularly pressing for expressivists. They must look elsewhere.

## 3.3 Deflationist responses

In this section, I examine whether any of the expressivist theories discussed earlier in the chapter (section 3.1.3) have the resources to explain defective concepts. I argue that they do not. Although only Horwich explicitly addresses the challenge, it is nonetheless instructive to see how the challenge applies in each case. Moreover, despite the differences of these theories, they all take on board some kind of deflationism about propositional content in order to respect commitment (C). As such, I have brought these responses together under the heading 'deflationist responses'. In the next section, and in keeping with the thesis of this thesis, I suggest that conceptual role expressivists need some *non*-deflationary notion of normative propositional content to answer the challenge. But first, I examine responses on behalf of the expressivist theories introduced above in reverse order.

### *3.3.1 Sinclair*

Although Sinclair does not acknowledge the problem of defective concepts, he does acknowledge that a theory of possession conditions for a concept should be supplemented with a determination theory of that concept. This is because Sinclair (2018: 110) considers it a platitude that the truth conditions of thoughts are determined by the semantic values of their constituents. If concepts are the constituents of thoughts, and we accept a conceptual role account of concepts, then it follows that conceptual roles determine the contents of thoughts. And explaining how conceptual role determines the contents of thoughts is the job of a determination theory. However, Sinclair argues that a *deflationary*

[6] They do have problems of their own. For example, it's not obvious that Wedgwood's conceptual role for OUGHT really provides enough constraints to specify a unique reference across all uses.

determination theory can do just as well as Peacocke or Wedgwood's realist versions.

Sinclair's idea is that all of the relevant platitudes about conceptual roles determining semantic values and truth conditions can be satisfied by a deflationary interpretation of 'truth conditions' and 'semantic values'. Specifically, his idea is that "if truth-conditions are understood in deflationary terms and semantic value is just contribution to truth conditions, then semantic value is also deflationary." (2018: 110) Thus, the truth conditions of a simple predicative thoughts can be given by the deflationary schema: the thought [that x is *F*] is true iff x is F. If we then understand semantic value in terms of contribution to truth conditions, we can then say that the deflationary semantic value of a predicative concept *F* is the set of F-things (ibid). Sinclair then argues that his deflationary determination theory respects the platitudes about how conceptual role determines content without supposing that normative concepts denote worldly properties. In this way, Sinclair believes that he can adopt Wedgwood's theory of possession conditions for concepts like OUGHT and BETTER THAN while respecting commitment (C).

I will grant for the sake of argument that Sinclair's deflationary determination theory respects all of the platitudes about conceptual roles determining the contents of thoughts. The problem, as we have already seen, is that a determination theory is meant to do more than this. Among other things, a determination theory is meant to explain defective concepts. However, the deflationary determination theory looks ill equipped to do this. First, and most generally, deflationary semantic notions cannot play *any* explanatory role in a theory of meaning or content. A fortiori, they cannot play a part in explaining defective concepts. Although semantic deflationists do not deny that deflationary truth or reference or whatever is genuine truth or reference or whatever, they claim that these notions can only play an expressive role in theoretical discourse. According to deflationists, there is nothing it is to be true, and because truth has no (substantive) nature, it isn't such that it can play an explanatory role.

Second, suppose we put this general worry about the explanatory potential of deflated notions aside and look Sinclair's determination theory for normative concepts in particular. Recall that Sinclair endorses the following conceptual role for BETTER THAN as that concept *B* to possess which a subject must be apt to make the following transitions:

*A*(*B*) S accepts *B*(*x*,*y*,*me*,*t*) → S has conditional intention to do x rather than y at t.

Could Sinclair's deflationary determination theory explain why subjects are correct to follow *A*(*B*)? It is hard to see how it could. For Sinclair, the deflationary semantic value of a predicative concept is its extension. In this case, we can think of the extension of *B* as a set of tuples corresponding to <*x*,*y*,*me*,*t*> — though this set will not be determined by some metaphysical property or relation of *better-than-ness* instantiated or obtaining between the members of the tuple, which would amount to a representationalist determination theory. But if the semantic value of *B* is merely a set, it is unclear that this explains why one's accepting that <*x*,*y*,*me*,*t*> falls within the extension of *B* makes it appropriate for one to intend to do x rather than y at t.

Third, not only do we lack an explanation of the correctness of the transitions specified by *A*(*B*), but we also lack an explanation of why we cannot provide a deflationary determination theory for concepts like HIYO. Recall the possession conditions of HIYO were as follows:

*A*(*H*) S wants to accost x → S accepts *H*x
S does not want to accost x → S rejects *H*x

If a deflationary determination theory says that the semantic value of a predicative concept is just its extension, then it would seem that there is nothing to stop us from providing a deflationary determination theory for *H* in terms of its extension. Presumably, Sinclair would want to deny that HIYO is a genuinely predicative concept and thus deny that it has any extension. But this is exactly what needs to be explained. Unlike the representationalist determination theory, the deflationary theory alone lacks the resources to explain this. So Sinclair's deflationary determination theory neither explains why normative concepts are nondefective nor why HIYO is defective.

### *3.3.2 Köhler*

Because Köhler endorses a thought-first version of conceptual role theory, according to which whole thoughts are primary rather than their constituents, the challenge will take a slightly different form. More specifically, the challenge will not so much be to explain defective concepts, but rather to explain a defective class of mental states. So instead of individuating a concept by its core conceptual role, on Köhler's picture we would need to individuate the relevant class of mental states in terms of the normative predicate used to conventionally express such states. Roughly, we can individuate (say) ought-beliefs as those mental states conventionally expressed by sentences of the form

'S ought to do x'. According to Köhler, this sentence provides an illustrative example of a shared functional profile of ought-beliefs.

Like Sinclair, Köhler does not discuss the possibility of defective concepts or the challenge it might pose to expressivism. In any case, it does not seem that his account has the resources to explain defective concepts. First, note how on Köhler's account, logical and semantic relations between declarative sentences earn their keep in codifying functional relations between mental states that are expressed by such sentences. Next, note that it is surely *possible* that there could be a functionally defined state such that a subject comes to accept [that Bob is hiyo] on the basis of the subject's desire to accost Bob, and which also causally interacts with other mental states that preserves the isomorphism of the 'logical' relations between 'hiyo'-sentences and other sentences. If this is a genuine possibility, then 'hiyo' can meaningfully embed and participate in genuine inference just as much as any normative predicate. So Köhler's account fails to explain why HIYO is defective or why normative concepts are not.

### *3.3.3 Horwich*

Recall that for Horwich, to believe [that I ought to do x] is to accept the sentence 'I ought to do x'. If the meaning-constituting property of 'ought' is that subjects accept sentences of the form 'I ought to do x' when they are strongly inclined to do x, the challenge for Horwich is to explain why a subject cannot genuinely accept 'hiyo'-sentences. As far as I'm aware, Horwich is the only conceptual role expressivist to acknowledge that something like HIYO might pose a particular challenge for his view. Like Sinclair and Köhler, Horwich embraces a thoroughgoing deflationism about semantic notions. And Horwich, perhaps more than anyone, is keen to stress that deflationary notions cannot be put to explanatory work. So Horwich's solution is to introduce additional constraints on the theory of possession conditions to explain why HIYO is defective but normative concepts are not. (Again, it's worth stressing that Horwich doesn't frame his theory or arguments in these terms, but that's merely a matter of terminology.) Given Horwich's sententialism about belief, we have already seen one such additional constraint. This is that to believe some proposition, the sentence one accepts in believing it must be syntactically well-formed. However, because 'hiyo'-sentences are syntactically well-formed, we will need some further constraint.

To identify the relevant constraint, Horwich begins with the observation that when one person genuinely accepts something and someone else accepts its negation, there is a *disagreement*: "there is conflict, a clash, a feeling that the other person is somehow in bad shape." (2010: 182) This seems to be something

lacking in the hiyo-case. Suppose I want to accost Bob and come to accept [that Bob is hiyo]. Next, suppose you do not want to accost Bob and come to accept [that Bob is not hiyo]. Given *A*(*H*), we are both competent users of HIYO who have 'correctly' come to form our beliefs according to the meaning-constituting rules governing 'hiyo'. However, it seems implausible to suppose that there is any real sense in which we disagree about whether Bob is hiyo. From this, Horwich argues that it is a necessary condition on genuine acceptance that accepting something and its negation must have the potential to engender practical conflict. This is grounded in the way that belief interacts with other attitudes in producing action. Specifically, Horwich claims "the conflict associated with contradictory beliefs consists in their potential, through inference, to engender conflicting desires and decisions." (2010: 183) I will call this the disagreement constraint.

In many ways, this notion of practical conflict echoes Stevenson's (1944) notion of disagreement in attitude. An example Stevenson gives concerns a disagreement between a curator of a museum who wants to buy contemporary art and his advisors who prefer to buy traditional art. In this example, there is a practical disagreement about what to do (i.e. which kind of art to buy for the museum). While we can assume that such practical disagreement does not consist in any particular disagreement in belief, it is easy to see how such disagreement might derive from contradictory beliefs. For example, perhaps each party has contradictory beliefs about how many visitors each kind of art would attract to the museum. Or perhaps they have contradictory beliefs about the value of contemporary art in relation to traditional art. In each case, we can imagine a practical conflict that is engendered in part from each party holding contradictory beliefs. Moreover, as the example illustrates, this seems to be true of contradictory evaluative beliefs as well as empirical beliefs. For we can imagine that the practical disagreement stems from a disagreement over the proposition [that traditional art is more popular than contemporary art] or over the proposition [that traditional art is more valuable than contemporary art]. Because no such practical disagreement seems to stem from contradictory HIYO-beliefs, Horwich claims that we have an explanation of why HIYO is defective but normative concepts are not.

However, Horwich's proposal is unsatisfactory and should be rejected. On first appearances, it might look like Horwich's disagreement constraint is too strong. First, there are examples where it is not obvious that believing contradictory propositions engenders disagreement. For example, suppose I believe [that gin is tasty] and you believe [that gin is not tasty]. Second, there are examples of beliefs that are far removed from our actions or practical attitudes. Suppose Kurt believes [that numbers are Platonic objects] and David

believes [that numbers are symbols to which we give an interpretation]. This is a case where there does seem to be disagreement, but it is not one that is likely to engender any practical conflict. Indeed, it seems plausible that the majority of our beliefs are not directly tied to action in any particular way. So one might think that the disagreement constraint is too strong. In fact, however, given Horwich's characterisation of what it is to engender practical conflict, the constraint is far too weak. This is because *any* contradictory beliefs can engender practical conflict *given other suitable premises to reason from*. And this is no less true of HIYO-beliefs.

Consider again Kurt and David. As a matter of fact, it seems unlikely that Kurt and David's beliefs would engender the kind of practical conflict that Horwich has in mind. However, it is easy to imagine examples in which their beliefs would engender a practical conflict. For example, they might disagree over which of two research applications to grant funding to based on which application they believe is most worthy, which in turn might be based on what philosophical views they take to be true. Or perhaps Kurt has the bizarre belief [that if number are abstract objects they demand respect where demanding respect involves frustrating the desires of nonbelievers]. Together with Kurt's belief [that numbers are abstract objects], this spells trouble for David. This shows that even beliefs that are typically far removed from practical considerations can engender practical conflict given suitable premises to reason from. The problem for Horwich is not that this is false, but that it cannot explain why HIYO is defective.

Suppose Kurt and David both learn of a new faculty member called Bob who will be joining the department next semester. Upon learning that Bob is a hardcore Platonist, Kurt forms a desire to accost Bob and David does not form any desire to accost Bob. In virtue of their respective desires, Kurt comes to accept [that Bob is hiyo] and David comes to reject [that Bob is hiyo]. Now suppose that Kurt has the bizarre belief [that people that are hiyo demand respect where this involves frustrating the desires of hiyo-nonbelievers]. Taken together, Kurt's beliefs spell trouble for David. So again, we have a practical conflict engendered by contradictory beliefs. Hence, Horwich's disagreement constraint is too weak because it falsely predicts that we can genuinely accept 'hiyo'-sentences.

What's going on here? Horwich's proposal claims that a sentence can participate in genuine inference in virtue of its capacity to engender practical conflict when it and its negation are accepted. I think the above examples show that this gets things the wrong way around. Rather, what the examples show is that accepting a sentence and its negation can engender practical conflict in virtue of their capacity to participate in genuine inference. So Horwich's

explanation for why OUGHT-beliefs but not HIYO-beliefs can participate in inference begs the question, as it already presupposes that the former but not the latter can be used in practical reasoning. As such, Horwich fails to explain why HIYO is defective but normative concepts are not.

### *3.3.4 Biting the bullet?*

I have assumed throughout this chapter that HIYO is a defective concept. All of the responses discussed above accept this assumption as a premise and try to explain the difference between HIYO and normative concepts. Could this assumption be challenged? In other words, could the conceptual role expressivist simply claim that HIYO, though weird, is not a defective concept? Although I'm not aware of anyone defending this position in print, it has been suggested to me that conceptual role expressivists would do well to respond this way.[7] If this is right, then it's at least not obvious that the expressivist faces any challenge explaining defective concepts, as the particular challenge for expressivism was motivated by the similarities between normative concepts and HIYO.

We might develop this thought in the following way. First, suppose we adopt an inferentialist view of the logical connectives like the material conditional that says something like the following: to believe [that if p then q] is to be committed to not accepting p while also rejecting q (compare Blackburn 1988). Given that we have an account of what it is to accept and reject HIYO-thoughts, then contrary to my above arguments, it might seem that we do have a conception of how HIYO can meaningfully embed and participate in genuine inference, at least for truth-functional connectives like 'if-then'. Specifically, to accept an embedded HIYO-thought is to be committed to holding certain combinations of HIYO-thoughts and other attitudes. This commits us to a certain picture of logical connectives, but it is one that conceptual role expressivists are likely to be sympathetic to anyway.

Further, we might provide a debunking explanation of the intuition that HIYO is a defective concept as follows. Perhaps HIYO seems weird simply because *we* are not HIYO-people. We do not possess any such concept and would have no use for one. Moreover, it is hard to imagine what use a community would have for such a concept. This is in comparison to normative concepts, which expressivists claim earn their keep in motivating actions and coordinating attitudes. However, just because it is hard to imagine a use for HIYO it does not follow that there is no possible use. Perhaps we can imagine a

[7] This was suggested to me by separately by Michael Ridge and Wolfgang Schwarz. One might also deny that normative concepts are non-defective—see footnote 2 of this chapter.

community which for some bizarre reason attaches the utmost significance to people's desires about who they accost. Surely such a community is possible, and it seems at least more plausible that they might have use for a HIYO-like concept. Arguably, therefore, HIYO is not defective.

I think that there are at least three things that can be said in response here. First, one might worry that embracing HIYO will overgeneralise the quasi-realist ambitions of most contemporary expressivists. Recall from the first chapter that expressivists typically employ deflationary interpretations of notions like normative truth, normative facts, normative properties, and so on. If we accept that HIYO is on all fours with normative concepts, then we should expect to be able to provide similar interpretations of hiyo truth, hiyo facts, hiyo properties, and so on. Of course, *we* might not care to vindicate these notions, but it does not follow that such notions *could* not be vindicated by or on behalf of those (possible) subjects who possess the concept. However, that there are vindicatory explanations to be had of these notions seems highly less plausible than for normative concepts. Perhaps one might reject expressivism's quasi-realist ambitions, but I think this is a significant cost to those sympathetic to expressivism.

Second, embracing HIYO in the way suggested above arguably begs the question. For recall that the above response relied on a certain view of logical connectives, namely that to believe [that if p then q] is to be committed to not accept [p] and reject [q]. It was then suggested that we simply substitute a 'hiyo'-sentence in place of 'p' or 'q' to see how HIYO can meaningfully embed. However, this already presupposes that we can genuinely accept and reject HIYO-thoughts, which is exactly what needs to be established (compare Woods 2017: 231f). However, whether or not this does in fact beg the question might depend on thorny issues concerning whether the burden of proof lies with showing whether HIYO is defective or not.

Regardless, surely not just *any* candidate constitutive acceptance conditions will individuate a concept. Surely there must be *some* general constraints on what kind of acceptance conditions can individuate nondefective concepts. So even if we are entitled to assume that HIYO is nondefective, it still seems that we are owed some account of what makes it the case that certain conceptual roles individuate meaningful concepts whereas others do not. One might respond that all that matters is whether we could in fact use a concept according to its constitutive acceptance conditions. However, this would seem to reject any distinction between defective and nondefective concepts. In effect, this seems to debunk the claim that conceptual role determines content. Rather, it is closer to a kind of meaning or content scepticism. I won't try to argue

against such a view here, but it seems a high price to pay for embracing conceptual role expressivism.

In sum, deflationist responses to the challenge do not adequately explain defective concepts. This is not to deny that a deflationary account of semantic notions should not be part of an overall expressivist theory. On the contrary, it seems plausible that expressivists need deflationism to explain normative truth, reference, and the like (though see Ridge 2014: ch.7 for an argument that expressivists can accept a correspondence theory of truth). But deflationism cannot help expressivists answer the challenge.

## 3.4 An alternative

Having rejected various deflationist responses to the challenge of explaining defective concepts, in this section I propose an alternative response. In effect, the response is to propose a nondeflationary but nonrepresentational determination theory for normative concepts. Although I think this response is more promising than its deflationist competitors, I will argue that it leaves conceptual role expressivism a less interesting and distinctive position than its proponents claim. However, I argue that this is in fact what we should expect. The alternative solution then provides a natural segue to the next chapter.

An instructive place to begin is with Schroeter and Schroeter's (2003) critical discussion of Wedgwood's conceptual role theory for normative concepts. Focusing for the moment on the linguistic case, Schroeter and Schroeter consider the following putative possession conditions for the expression 'x is right for me':

(3) S accepts 'x is right for me' → S intends to do x.

If, following Wedgwood, we want to provide a realist determination theory for (3), then we might suppose that 'x is right for me' denotes the property *being the right thing for me*. This property would then explain why subjects are correct to follow the transition specified by (3). Against this, Schroeter and Schroeter (2003: 202) argue that a determination theory for a predicate should not assign a property as its denotation if competence with that predicate does not exhibit any sensitivity to that property. Because (3) places no constraints on when a subject can form beliefs about what is right for her to do, they argue that the correct determination theory for (3) should not assign any property as the reference of 'right for me'.

Instead, they tentatively propose that the correct determination theory for (3) will assign the same semantic value associated with the sentence 'I hereby

intend to do x'. It is plausible to think that this tentative suggestion correctly rationalises the transition specified by (3). If one sincerely utters 'I hereby intend to do x' then it is highly plausible that it would be appropriate to thereby intend to do x. However, it is less plausible that this suggestion can correctly explain other semantic features of the sentence 'x is right for me'. Unfortunately, Schroeter and Schroeter do not say what they take the semantic value associated with the sentence 'I hereby intend to do x' to be. Whatever it is, however, it cannot be the same as 'x is right for me' because the two sentences cannot be substituted salva veritate.

To be fair to Schroeter and Schroeter, they also deny that (3) provides the correct possession conditions for 'x is right for me', so perhaps their suggestion should not be taken too seriously. Nonetheless, I think their proposal points in the right direction. Transposing the example from the level of language to the level of thought, we might think of the mental analogue of uttering 'I hereby intend to do x' as something like a *decision* or *commitment* to intend to do something in the relevant circumstances. As before, this way of thinking would rationalise the transition from thinking [that x is right for me] to *intending to do x*. What we need is some way of understanding the left hand-side of the transition that reflects the propositional structure expressed by the sentence 'x is right for me'. We need this because not only do we need to explain the correctness of (3), but we also need to explain how the content [that x is right for me] can meaningfully embed in other attitudes and participate in genuine inference.

In fact, such a view of normative thought and content is available in Allan Gibbard's (2003) plan expressivism. Very roughly, Gibbard thinks that normative attitudes are akin in important ways to planning attitudes, and that the contents of normative attitudes can be explained in terms of the sets of plans not ruled out by those attitudes. Thus, to think [that x is right for me] is something like *planning to do x* in the relevant circumstances, where the content of this thought can be specified in terms of a set of plans that prescribe doing x. Returning to conceptual role expressivism, we might then provide a nonrepresentational but nondeflationary determination theory in terms of the planning contents assigned to normative thoughts. In the next chapter, I will argue that this account has the right structure to explain the propositional structure of normative thought. The important point for now is that planning contents both rationalise the conceptual roles of the concepts they are assigned to and have the right sort of structure to explain embedding and inference.

If we adopt this strategy for answering the challenge, however, we see that conceptual role expressivism becomes a less interesting position that it might have first appeared. For what this approach does is to shift much of the

explanatory burden from the theory of possession conditions to the determination theory. In other words, it is no longer the conceptual role part of the theory that is doing the heavy lifting, but the theory of content proper, which is to say, the theory of normative propositions. Where we were looking to conceptual role theory to explain normative content, the bulk of the explanation has come from elsewhere. Nonetheless, I think this is exactly what we should expect. This is because conceptual role theories are often couched as *metasemantic* theories or theories of content *determination*. Because conceptual role expressivists are typically also deflationists about propositional content, it is easy to overlook this point. This is because if a deflationary determination theory cannot explain defective concepts, then the theory of possession conditions has to provide all of the explanation. But as we saw above, such a view is highly problematic.

As well as being less interesting, conceptual role expressivism also becomes less distinctive. Although expressivists might not appeal to a conceptual role framework to explain normative concepts, I doubt few would deny that normative concepts should be explained in terms of their distinctive role in practical reasoning. If we accept that expressivists need some further theory of meaning or content to explain the semantic and logical properties of normative concepts, then this might seem to be compatible with a wide variety of expressivist theories of normative concepts. This isn't to say that expressivists should not embrace conceptual role theory. If conceptual role theory is an independently attractive view of content determination, or if conceptual role theory provides useful resources to formulate expressivist claims about normative concepts, then expressivists have every reason to embrace it. However, in doing so, the expressivist does not discharge the usual explanatory burdens that come with developing an expressivist theory of normative thought and discourse, contrary to what some conceptual role expressivists suggest.

As a final note, it is worth considering the implications that our discussion has for the expressivist theory of cognitive propositions sketched at the end of the previous chapter. There, I proposed that normative propositions are acts of predication, where such acts were understood as acts of categorisation according to concepts. This notion was sufficiently theoretically neutral to be compatible with expressivism without entailing expressivism. I thus argued that we would need some suitably nonrepresentational account of the concepts involved in such acts in order to vindicate their use in an expressivist theory of normative thought and discourse. I then examined conceptual role expressivism as a candidate for such an account. However, I argued that the approach failed to fully explain the relevant properties of normative thought.

Bringing this back to the account of cognitive propositions, this means that the conceptual role approach fails to explain all the relevant properties of acts of normative predication such that they could play the role of propositions within an expressivist theory. I suggested that conceptual role expressivism might be vindicated taken in conjunction with something like Gibbard's notion of planning contents, where conceptual role determines such contents for normative attitudes. While this might seem like an alternative to the cognitive propositions view, it is in fact compatible. For normative categorisation itself might be understood in terms of planning. Very roughly, categorising something as right for me to do would consist in grouping that thing with other things that I plan on doing in the relevant circumstances. Abstract planning contents could then be understood as modelling properties of such cognitive acts. So the account of cognitive propositions might be elucidated in terms of planning contents. However, because the approach to content examined in the next chapter can be maintained independently of the cognitive act view of propositions, I leave such discussion aside.

## 3.5 Conclusion

In this chapter, I have examined expressivist theories that appeal to conceptual role accounts of content to explain normative thought. Although many contemporary expressivists express sympathies with conceptual role approaches to content, there is no consensus about what combining these two approaches amounts to and what commitments their conjunction entails. My first aim in this chapter was to provide a fully general characterisation of conceptual role expressivism that unites otherwise disparate views. Additionally, this characterisation provides a general framework in which expressivists sympathetic to conceptual role expressivism can develop their theory, making explicit recourse to the various choice points discussed above. This shows that the conceptual role framework has room for many different versions of conceptual role expressivism in addition to those theories discussed above. While I have expressed scepticism that the conceptual role framework alone has the resources to explain the semantic and logical properties of normative thought, I have not argued that this cannot be a fruitful framework for expressivists. So this chapter advances debates about conceptual role expressivism by making precise what the view is committed to and the theoretical options available to conceptual role expressivists.

However, I have also argued that conceptual role expressivism cannot be the whole story of normative thought. Conceptual role expressivists owe us an explanation of how normative conceptual roles can determine genuinely

propositional content given commitments (A), (B), and (C). This was highlighted by the comparison with defective concepts like HIYO, but the point is a more general one. I think this point is particularly significant because conceptual role expressivists often write as if embracing conceptual role theory automatically solves the Frege-Geach problem. If the arguments I have given here are correct, then it shows that such a thought is mistaken. I therefore think that the most important lesson of this chapter is this. The conceptual role framework might provide a fruitful framework for expressivists to develop theories of normative thought. But the framework itself should not be seen as the solution to any problem or set of problems. All the puzzling features of normative thought as conceived by expressivists stand in no less need of explanation once we adopt a conceptual role view.

## 3.6 Addendum: Gibbard on Meaning and Normativity

I concluded this chapter proposing that conceptual role expressivism needs some notion of content, such as Gibbard's notion of planning content, in order to answer the challenge of explaining defective concepts. As far as I know, Gibbard does not anywhere discuss this challenge. However, the view he puts forward in his *Meaning and Normativity* (2012) embraces both a conceptual role account of content and a plan-expressivist account of normative mental states. It is therefore worth asking whether the view I am proposing here is the same as Gibbard's. I will argue that despite this initial similarity, the appeal to planning states in Gibbard's theory plays a very different role to the one I am proposing here. However, it should be said at the outset that this will only be a very cursory examination of Gibbard's view in *Meaning and Normativity*. This is in part because the view he proposes is complex and raises many questions of its own which would take the current discussion too far afield. But it is also because, as I will discuss in the next chapter, Gibbard has an idiosyncratic conception of the role that contents play in a theory of attitudes. However, I will also argue in the next chapter that one can make use of his distinctive notion of content while accepting a more orthodox conception of the role such contents can play in one's theory of attitudes.

Gibbard endorses a normativist conceptual role approach that assumes that we think in a version of the public language we speak and understand (2012: 117). The meaning of a word is the concept it expresses, where a concept is individuated in terms of the rules of rationality governing it. Specifically: "The meanings of the words in a sentence combine to explain which inferential oughts apply to the sentence and the evidential conditions under which one ought to accept or reject the sentence. A word's meaning what it does consists

in the pattern of oughts that enters into such explanations." (2012: 114) So, for example, for a word to express the concept SOMETHING is for that word to be governed by inferential oughts such as: one ought to infer 'something is white' from 'snow is white' (2012: 113). Gibbard then goes on to develop notions of synonymy and analytic equivalence to flesh out the conditions under which the same oughts apply to the same word, and hence express the same concept. Roughly, the idea is that two sentences are analytically equivalent and express the same thought just in case one ought to give each thought the same credence under any intelligible supposition (2012: 121ff). The precise details of Gibbard's view need not concern us here (see Williamson 2018 for critique). What is important is that Gibbard explains concept identity in terms of the rational connections between accepting thoughts containing that concept and other mental states. Hence, Gibbard endorses a conceptual role approach as I am using the term.

Gibbard then goes on to provide an expressivist meta-theory of meaning claims. Because meaning claims are claims about what one *ought* to believe or infer, Gibbard proposes that they express a species of planning states. By 'planning', Gibbard has something different in mind to our ordinary notion of planning. I will talk about this in more detail in the next chapter. But very roughly, planning states are a kind of directive attitude that govern our will and beliefs (2012: 172). Specifically, ought claims place restrictions on what to do or what to think for any subjective circumstance one could conceivably be in (2012: 174ff). So Gibbard thinks that our thought and talk about what things mean is tied to claims about what we should believe and infer, which are normative claims. However, this creates a strange sort of dual account of normative thought. On the one hand, normative thoughts are individuated in terms of their conceptual roles. But the same attitude is also modelled directly in terms of the restrictions it places on plans. So a question arises as to how these two accounts are related, and Gibbard is not entirely clear in explaining their relation.

In line with my discussion above, however, I think that we can profitably think of this in the following way. The conceptual roles of normative thoughts *determine* the restrictions those thoughts place on plans, and Gibbard's planning contents make explicit the form these restrictions take. Moreover, as we will see in the next chapter, the structure of these restrictions is such that it explains how normative thoughts can participate in genuine inference. To be clear, I am not claiming that this is how Gibbard himself would characterise his view. But it seems that by taking meaning to be normative, Gibbard ends up with a view that is similar to what is needed in order to solve the challenge of explaining defectiveness (though again, this is not something Gibbard takes

his account to do). Bringing the discussion back to conceptual role expressivism more generally, this discussion reemphasises two important implications. First, we need to appeal to more than just conceptual role in a full account of conceptual content. Second, conceptual role expressivism will need to take on other substantive commitments about the nature of normative content. And these commitments will be controversial and take us further away from the general view characterised by (A), (B), and (C).

# 4
# The Modal Conception of Content

This chapter examines whether an extension of possible worlds models of content can provide expressivists with a suitable theory of normative propositions. First, I locate the possible worlds model of content within a broader explanatory framework that characterises propositions in terms of possibilities (section 4.1). I call this the modal conception of content. Second, I explain how expressivists can develop a model of normative propositions within the modal conception of content (section 4.2). For concreteness and familiarity, I illustrate this with Allan Gibbard's (2003) notion of planning contents, though this is simply one way of implementing a more general approach. Third, I raise a general challenge for expressivists who appeal to the modal conception (section 4.3). The challenge is to answer what I call the foundational question, is to explain how the expressivist's model of content is grounded in an adequate foundational philosophy of mind compatible with expressivism. Fourth, I examine an answer to the challenge due to Seth Yalcin (2018a) and argue that it fails (section 4.4). Finally, I propose that the way to answer the challenge is to develop a general theory of belief compatible with expressivism that subsumes representational and normative beliefs (section 4.5). The task of providing such a theory is then taken up in the next chapter.

## 4.1 The possible worlds model of content

Among the views of propositions discussed in Chapter 2, I rejected the standard possible worlds view as suitable for an expressivist theory of normative propositions. According to the standard possible worlds view, propositions are sets of possible worlds, or functions from possible worlds to truth values, which carve out possible states of the world (I hereafter omit the latter formulation for ease of exposition). Because expressivists deny that normative beliefs are fundamentally in the business of tracking how the world is, believing a normative proposition for expressivists cannot (only) involve taking some possible state of the world to obtain. As such, normative propositions cannot be sets of possible worlds. In this chapter, I argue that although the possible worlds view is unsuitable for an expressivist theory of

normative propositions, the view is an instance of a more general approach to explaining content that can provide an attractive framework for expressivists. This more general framework is what I will call the *modal conception of content*, which explains content in terms of possibilities.

The basic thought behind why the modal conception of content might be attractive for expressivists is this. According to the modal conception, to accept a proposition is to rule out the set of possibilities incompatible with that proposition. The possible worlds view of content assumes that the relevant possibilities are possible states of the world. However, this assumption is not entailed by the modal conception as such and may be rejected. Doing so provides space for a view according to which to accept a normative proposition is to rule out a set of possibilities, but where these possibilities characterise some kind of nonfactual or nonworldly domain. So long as we can characterise this domain in terms of possibilities, such a view would remain firmly within the modal conception. So this approach would seem well suited to respecting the unity requirement argued for in Chapter 2.

In the next section, I explore in more detail what such a view might look like. First, however, I explain and motivate the modal conception of content in more detail. Initially, I focus on the simplest version of the modal conception, the possible worlds model of content. I then show that quite independently of explaining normative belief, there are grounds for extending and enriching the possible worlds model to explain certain other distinctive domains of thought, specifically de se and de nunc thought. This provides support for the expressivist extension of the possible worlds model by showing that there are already pressures to enrich the model to account for other distinctive domains of thought. However, the support is only indirect, because while de se thought is not world-characterising content, it is factual in the broader sense defined in Chapter 1.

### *4.1.1 The possible worlds model: an overview*

At the most general level, the modal conception of content makes the following claim: contents distinguish between possibilities. To grasp the content of assertion is to know which possibilities are ruled out by that assertion. To believe a proposition is to rule out certain possibilities as true, namely those incompatible with that belief. This makes intuitive sense. For example, if I believe [that Edinburgh is north of London], then presumably I rule out various possibilities in doing so, such as Edinburgh being south of London or their having a shared latitude. In failing to know what possibilities the belief rules out, it would seem that I fail to grasp its content. If someone asserted [that

Edinburgh is north of London] and I did not take this to rule out these other alternatives, then plausibly I have not understood what was said.

More precisely, then, which possibilities do I rule out and which do I 'rule in' when I believe [that Edinburgh is north of London]? According to the simplest version of the modal conception, we can answer this question in terms of possible worlds. Specifically, where $W$ is the set of all possible worlds and the variable $w$ ranges over the members of this set, the content of my belief is the set of possible worlds in which Edinburgh is north of London, or $\{w \in W$: Edinburgh is north of London in $w\}$. (Equivalently, this can be characterised as the function $f$ such that $f(w)$ = true iff Edinburgh is north of London in $w$.) A possible world provides us with a complete specification of the way things might be, and the truth of the proposition [that Edinburgh is north of London] is compatible with many other ways the world could be besides how it actually is. The possible worlds view of content identifies this proposition with the set of those possibilities.

More precisely, we can understand the possible worlds model of content to be the following (Yalcin 2018b: 24):

> **The possible worlds model of content**. The content of a state of belief is representable as a set of possible worlds, intuitively the worlds "left open" by what is believed. Propositions are sets of possible worlds, and the propositions an agent believes are those true with respect to all of those worlds the state leaves open.

The possible worlds model of content is compatible with a number of different ways of understanding possible worlds. I will simply assume here that possible worlds are properties, namely possible states of the world or ways that the world might be (e.g. Stalnaker 1987; 2012). This contrasts with possible worlds being concrete (e.g. Lewis 1986), mind- or language-dependent (e.g. Carnap 1947), or maximally consistent sets of propositions (e.g. Adams 1974). It follows from this conception that propositions are unstructured, in the sense that the proposition [that p and q] is identical to the proposition [that q and p]. This is because both propositions leave open the same set of possibilities and so have the same worlds as elements.

Given this conception of possible worlds as properties, a further question arises as to how to characterise the set of all possible worlds $W$ and how to individuate the worlds within $W$. As before, the possible worlds model as such does not commit us to any particular answer. A straightforward answer would be that $W$ consists in the complete set of metaphysically possible ways the world can be. This is not the answer I will assume here, however. Rather, I will

assume that the relevant domain of possible alternatives is defined relative to the discriminatory abilities of the subject or subjects whose attitudes are in question (see Stalnaker 1987 for the classic statement of this view). This means that from the perspective of the theory of content, there is not a single domain of possibilities relative to which all propositions are subsets. Rather, the domain will vary depending on the theoretical context. So relative to subjects with differing cognitive capacities, or relative to different contexts of inquiry or conversation, the possible worlds in *W* might be cut up more or less finely. This feature, which I will call *discrimination relativity*, will prove important in the discussion ahead and will be returned to in more detail.[1]

### *4.1.2 The possible worlds model: attractions*

The possible worlds model provides a first approximation of belief content as conceived by the modal conception. It provides a perspicuous representation of what is involved in claiming that contents are ways of dividing up possibilities. However, not only does the possible worlds model explain our intuitions about what is involved in grasping and accepting propositions. It is also attractive from a theoretical perspective. First, the possible worlds model provides an elegant way of explaining standard propositional relations in set theoretic terms (this will be important in accommodating a nonfactualist view of normative content). For example, for any sets of possible worlds $\phi$ and $\psi$:

**inconsistency** = $\phi_1, \ldots, \phi_n$ are inconsistent iff $(\phi_1 \cap \ldots \cap \phi_n) = \emptyset$
**entailment** = $\phi_1, \ldots, \phi_n$ entails $\psi$ iff $(\phi_1 \cap \ldots \cap \phi_n) \subseteq \psi$

Logical notions can likewise be defined, for example:

**negation** = $\phi'$
**disjunction** = $\phi \cup \psi$
**conjunction** = $\phi \cap \psi$

As set-theoretic relations are already well understood, this provides an attractive framework for explaining propositional and logical relations.

[1] Stalnaker calls this feature "the relativity of content" (1987: 20) and Dennett calls it "agent relativity" (1987e: 207). Note also that defining the domain of possibilities relative to a subject does not entail that the existence of the domain is in any way dependent on the subject, so possible worlds can still be mind-independent on this view.

Second, another advantage of the possible worlds model is that it allows for non-linguistic creatures to possess beliefs and other propositional attitudes. This is because propositions are not linguistic entities but abstract possibilities individuated relative to the creature's discriminatory capacities. This fact also provides the resources to explain how we could have evolved to have propositional thought in terms of the development of our discriminatory responses to our environment. Discrimination relativity also helps to explain a number of other features of belief. For example, it is plausible that understanding a proposition is gradable. Consider a young child whose mother is a doctor (this example comes from Dennett 1969: 183; see also Stalnaker 1987: 64f). From a young age, it is true to ascribe to the child the belief that her mother is a doctor. However, although the child believes this proposition from a young age, we can expect her understanding of the proposition [that her mother is a doctor] to increase as she grows older. The possible worlds view can explain this nicely. As the child grows older, we may suppose that the range of possibilities she rules out in believing [that her mother is a doctor] grows as she comes to discriminate more finely between possibilities in which her mother is a doctor and possibilities in which she isn't.

Third, a number of other advantages follow from a feature of the possible worlds model not yet commented upon. This is the idea that it is a subject's *total* belief state that is primary as opposed to her individual beliefs. Thus, according to the possible worlds model, a subject S's beliefs are modelled by the set of worlds $W_S$ compatible with S's total belief state. Individual beliefs are then modelled as properties of the total belief state. Specifically, S believes [p] just in case $W_S$ is a subset of [p]. This holistic aspect of the model is well placed to explain phenomena such as how belief change seems to be holistic. For example, when I come to believe [that my keys are in the car], this seems to involve my coming to believe many other things, such as [that my keys are not in my pocket], [that they are in some car], and so on (Yalcin 2018b: 25). Given how extensive we could make this list, it seems implausible that acquiring each of these beliefs is a distinct cognitive achievement over and above the initial belief change. If changes in belief are primarily changes in one's total belief state, however, then we can understand coming to believe all of these propositions as a single cognitive achievement.

The holistic aspect of the model also explains how we can seemingly have beliefs of unbounded complexity (Yalcin 2018b: 25). For instance, consider an example in which A and B are said to have common knowledge of the proposition [p]. It is often thought that common knowledge entails that A knows [that B knows that p], A knows [that B knows that A knows that B knows that p], and so on. It seems implausible that with each rising level of

complexity, A grasps and accepts a distinct content. By contrast, the possible worlds model can characterise everything entailed by the common knowledge [that p] simply in terms of the possibilities ruled in and out by A and B's common knowledge.

To sum up, the possible worlds model provides a theoretically attractive model of belief content in that: (a) it reduces propositional and logical relations to well-understood set-theoretic relations; (b) it can capture the variability of content ascriptions across subjects in terms of discrimination relativity; and (c) it can capture holistic features of belief in virtue of the primacy of subjects' total belief states.[2] Of course, the possible worlds model is not without its problems either. I here mention three, the last of which will be important in setting the stage for an expressivist-friendly modal view of content.

### *4.1.3 The possible worlds model: three problems*

The first problem is that it seems highly plausible that we can have distinct beliefs about distinct necessarily true claims. However, the possible worlds model apparently entails that all necessarily true propositions are identical, because they are characterised by exactly the same set of possibilities. If beliefs are individuated by possible worlds propositions, it follows that all beliefs in necessarily true propositions are identical. But this seems absurd. For example, my belief [that 5 +7 = 12] is surely not identical to my belief [that modus ponens is valid]. Although this is a serious problem, I do not wish to dwell on it here, as this would take us too far from the present inquiry. But one response to this problem might be to treat necessarily true claims as metalinguistic claims about which sentences express the one necessary proposition (see for example Stalnaker 1987: 72ff). In this way, we can take the objects of our mathematical and logical beliefs to be about which statements express the one necessary proposition. (I will briefly mention another response below.)

A second problem is that it seems plausible that we sometimes acquire new beliefs by working out the consequences of what we already believe. However, the possible worlds model apparently entails that we already believe the logical consequences of what we already believe. As such, it fails to capture this kind of activity. Again, I do not wish to dwell for long on this problem. But one response might be to complicate the possible worlds model by attributing separate *systems of belief* to a subject as opposed to a single total belief state. This is the idea that a subjects' beliefs are 'fragmented' or 'compartmentalised' to separate systems of belief (see Stalnaker 1987; Lewis 1986; for a discussion of

[2] In the next chapter, we will examine some further benefits of the modal conception of content in providing a foundational philosophy of mind.

independent empirical support for the idea of fragmented belief systems, see Porot and Mandelbaum forthcoming). Deductive reasoning can then be thought of in terms of integrating separate systems of belief (see Yalcin 2018b for further discussion).

A third problem for the possible worlds model is its inability to explain de se and de nunc thought, or beliefs that involve indexical concepts such as I, HERE and NOW. For instance, it seems possible that (a) I believe [that my interview is at 1:00pm], (b) I believe [that my interview is not now], and (c) it is now 1:00pm. The possible worlds model struggles here because it apparently fails to differentiate between the propositions believed in (a) and (b), as given (c), both propositions are characterised by the same set of possible worlds. Unlike with the previous responses, I want to a spend a little more time exploring this problem, as it will set the stage for what's to come.

In general terms, the problem is this. I do not simply inhabit a possible world within logical space. I also inhabit a time and a place within that world. Indexical concepts such as I, HERE and NOW allow us to have what Perry (1977) calls *self-locating beliefs* — beliefs that allow me to 'locate' myself within the possible world I inhabit. We need indexical concepts in order to distinguish my belief [that my interview is at 1:00pm] and my belief [that my interview is now]. If I were to learn that it is now 1:00pm, I do not learn something new about what the world is like. Rather, I learn something new about the perspective I inhabit within that world, in this case something about my temporal location.

On the basis of examples like these, Lewis (1979a) proposes that the objects of beliefs are *centred* worlds. It is a matter of controversy exactly how to conceive of centred worlds (see Liao 2012 for discussion), but as a first approximation we can say that they consist of an ordered set of (i) a possible world and (ii) a perspective within that world. More precisely, we might understand the perspective as a space-time coordinate <*x*, *y*, *z*, *t*> or an agent-time pair <*i*, *t*>. The details here will not concern us, so I will assume that centred worlds are tuples of world, agent, and time <*w*, *i*, *t*>. Roughly, then, indexical content encodes information relating to the *i* and *t* parameters.[3]

Using centred worlds to model various kinds of indexical thought is fairly commonplace. This highlights that those who utilise the modal conception of content are happy to add further complexity and structure to the domain of alternative possibilities if this is needed to model certain aspects of belief not

[3] Lewis (1979a) argues that this shows that the objects of beliefs are not propositions, because Lewis assumes that propositions just are sets of possible worlds. Because I define 'propositions' as whatever entity plays the object of attitude role, sets of centred worlds (or the properties they consist in or represent) are understood here as a possible candidate for what propositions are, as opposed to something to be contrasted with propositions.

captured by the initial model. Indeed, this general strategy has also been used to answer the problem of distinguishing between the logically equivalent truths of mathematics by introducing a distinct parameter to model mathematical information (Pérez Carballo 2016). Similar strategies have also been employed to explain discourse about epistemic modality (Yalcin 2007; 2011) and know-how (Santorio 2016). So although the possible worlds model is perhaps the most standard implementation of the modal conception, the modal conception itself is flexible about what exactly will constitute the domain of alternative possibilities in any particular theoretical context. Moreover, if content is discrimination relative, as was suggested above, this is exactly what we should expect.

In the next section, I argue that a familiar expressivist strategy for explaining normative content is best understood as a particular instance of the more general strategy described in the previous paragraph. Specifically, I argue that Gibbard's notion of planning contents can be understood as introducing a distinctive nonfactual parameter into the possible worlds model with which to model normative belief content. The point to be stressed here is that this move can be motivated from within the possible worlds framework. To that extent, Gibbard's proposal is more conservative than is often thought. However, to the extant that normative belief content is nonfactual, Gibbard's proposal constitutes a more radical departure from orthodox views. This is because although there is a sense in which indexical content is not world-characterising, it is still factual in the broader sense that the truth or falsity of an indexical belief is still determined by reality.

## 4.2 The world-hyperplan model of content

Gibbard (2003) proposes that normative beliefs are a kind of planning attitude, which can be characterised in terms of the plans they rule in and out, much as descriptive beliefs can be characterised in terms of the worlds they rule in and out. In this section, I show how we can understand this as a proposal within the modal conception of content. As we will see, locating Gibbard's proposal within this broader explanatory framework will prove vital to providing an expressivist account of normative propositions that respects what in Chapter 2 I called the unity requirement:

> **Unity requirement**. Expressivists need a unified explanation of that in virtue of which our cognitive acts and attitudes have propositional content.

The basic idea will be that if belief content is discrimination relative, then in order to see how we might extend the possible worlds model to characterise normative belief, we should first look to the distinctive discriminatory abilities implicated in normative belief in virtue of which it is correct to assign nonfactual contents.

This way of developing Gibbard's proposal will differ in a number of important ways to Gibbard's own view, and I will indicate where I take there to be substantive rather than merely presentational differences between the two. However, it will not matter too much whether the resulting view is one that Gibbard himself would or would not accept. What matters is the view itself. The remainder of this section will be divided into two parts. First, I explain the world-hyperplan model of content. Second, I will further clarify the view by considering and responding to four possible objections.

### *4.2.1 The world-hyperplan model: an overview*

Our ordinary factual beliefs aim to capture how the world is and where we stand in it. The centred worlds model of content cashes this out in terms of dividing up a domain of centred worlds and locating the actual world as among those worlds in which the content of one's belief is true. Normative beliefs, the expressivist maintains, aim to settle not how things are but what to do, think, and feel in various circumstances. If the object of a prosaically factual belief is a way that things might be, we might therefore think of the object of a normative belief to be something like a possible way of acting, thinking, or feeling. If the object of a normative belief is something like a possible way of acting, thinking, or feeling, then how exactly should we characterise the relevant domain of possible alternatives that normative contents carve up?

To begin to answer this question, let us first follow Gibbard (2003; 2012) in supposing that normative judgments are *plan-laden* attitudes. For instance, imagine that I judge [that all things considered I ought to give comments on my friend's manuscript]. We might think of this judgment as akin to forming a plan to comment on my friend's manuscript in the relevant circumstances. After all, if I make this judgment, then in normal circumstances and other things being equal, we would expect me to act as if I had made such a plan. So the suggestion seems intuitively plausible.

However, the sense in which normative judgments are like plans is broader than our ordinary notion of planning in a number of ways. For example, consider the judgment [that Caesar ought to have crossed the Rubicon when Rome was lightly defended]. In some ways, this judgment is akin to a contingency plan to cross the Rubicon if I were in circumstances subjectively

indistinguishable from Caesar's. However, it seems implausible that in making this judgment, I form any such plan in the normal sense. But it is nonetheless plan-like in that it prescribes a particular action based on a contingency. Further, consider the judgment [that Caesar ought to have believed that Rome was lightly defended, given the evidence]. Clearly, this judgment cannot involve a plan in the ordinary sense to have a certain belief in Caesar's circumstances since I cannot literally plan to believe anything. Believing is simply outside the scope of volition. Nonetheless, when we criticise or correct others or ourselves for failing to follow the 'plans' prescribed by our epistemic normative judgments, or take steps to correct ourselves, we are in a broader sense following our plans to think in certain ways (compare Gibbard 2012: 173). Henceforth, I will use 'plan' and its cognates in this broader, extended sense.

Suppose, then, that normative judgment can be profitably conceived of as a kind of planning attitude. How should we understand the relevant domain of alternative possibilities that the contents of normative judgments distinguish between? It is here that Gibbard introduces the technical notion of a *hyperplan*. Just as a possible world provides a complete and coherent specification of a way things might be, a hyperplan provides a complete and coherent specification of what to do, think, or feel for any possible situation. As a first approximation, whereas the content of a prosaically factual belief is given by the set of worlds left open by the belief, the content of a normative belief is given by the set of hyperplans left open by the belief. The intuitive idea is that, relative to some situation, hyperplans provide prescriptions about what to do, think, or feel, and to accept some set of hyperplans is to be disposed to behave in accordance with those prescriptions. More generally, where $w$ is a centred world and $h$ a hyperplan, a subject's belief state can be modelled by the set of $\langle w, h \rangle$ pairs compatible with her beliefs.

Intuitively, by characterising logical space in terms of points of $\langle w, h \rangle$ pairs, belief contents distinguish not only between ways the world can be and ways we can be located within the world, but also between what to do, think, or feel in different scenarios. To accept a normative claim on this view is to rule in some action, thought, or feeling as being the thing to do, think, or feel. More generally, we arrive at the following view:

> **The world-hyperplan model of content**. The content of a state of belief is representable as a set of $\langle w, h \rangle$ pairs, intuitively the $\langle w, h \rangle$ pairs "left open" by what is believed. Propositions are sets of $\langle w, h \rangle$ pairs, and the propositions an agent believes are those that hold with respect to all of those $\langle w, h \rangle$ pairs the state leaves open.

From a formal perspective, the world-hyperplan model provides a conservative extension of the possible worlds model. First, hyperplans do not require introducing any new elements into the model because both possible worlds and hyperplans can be constructed out of possibilia and sets thereof (Yalcin 2012: 147; Stalnaker 2014: 130). If we think of plans as mappings from situations to prescribed outcomes, a hyperplan can be defined as a function from sets of worlds that realise possible situations to sets that realise prescribed outcomes. Second, the definitions of propositional relations and logical notions given in the previous section carry directly over to sets of <*w*, *h*> pairs. In this way, the world-hyperplan model treats logical complexity in factual and normative thought in exactly the same way as the possible worlds model. Third, as I argued in the previous section, the possible worlds model already needs enriching to account for indexical content. So there is no in principle reason why we cannot add further complexity to the model to characterise normative thought.

In the next section, I raise a challenge for the world-hyperplan model in relation to what it is to accept world-hyperplan contents. In the remainder of this section, I consider four objections to the model. These objections will help to clarify the view further and to highlight the explanatory burdens faced by the view.

First, however, it should be noted that the general approach of developing an expressivist theory of propositions within the modal conception is not wedded the world-hyperplan model as such. I opt for this model because of its familiarity within the metaethical literature and its clear associations with expressivism. Moreover, as we will see later in the chapter, one of the best developed expressivist theories that embraces the modal conception appeals to hyperplans (though I will go on raise some objections to it). But different notions might model normative content as conceived by expressivism just as well. For instance, Charlow (2014) proposes that expressivists might appeal to the kind of imperatival content that characterise the 'To-Do Lists' employed by Portner (2007) to model the pragmatics of imperatives.[4] And Chrisman (2016) proposes that the semantics for 'ought' might appeal to Castañeda's (1975) notion of a *practition* to explain agential 'ought'-sentences. Another approach would be to appeal to something like *standards* understood as orderings of possible worlds in the style of Kratzer's semantics for deontic modals. For

[4] It is unclear whether Charlow's proposal is consistent with the Portnerian framework he employs. This is because for Portner, "a modal sentence is proposed for addition to the Common Ground, while an imperative is proposed for addition to the addressee's To-Do List." (2007: 363) By contrast, Charlow suggests that deontic normative sentences (a) are modal sentences and (b) express imperatival contents.

example, Silk (2013) and Ridge (2014) endorse something close to this.[5] As all of these proposals could be modelled within an extension of the possible worlds framework, they are all available to expressivists who embrace the modal conception of content. However, I will henceforth focus solely on the planning model. Importantly, the objection raised in the next section will apply equally to any of these alternatives, so this particular choice point will not matter too much for the purposes of this chapter.

### *4.2.2 The world-hyperplan model: four objections*

I will now consider four possible objections to the world-hyperplan model of content. This subsection has two main aims. The first is to clarify the commitments of the world-hyperplan model and to forestall possible worries that are based on a misunderstanding of the view's commitments. The second is to highlight the explanatory burdens left undischarged by the world-hyperplan model. This discussion will therefore provide criteria of adequacy for an expressivist theory that accepts the world-hyperplan model.

#### *Truth*

The first objection to the model concerns the role of truth. It might be thought that an essential component of the possible worlds model is the idea of a proposition holding or being true at a world. However, one might worry that whereas we have a clear idea of how something could be true at a world, we lack any such conception of a proposition holding or being true at a world-hyperplan pair. For example, if we interpret 'truth' as some kind of correspondence relation between a proposition and a world, then if normative content is nonfactual there would be nothing for the *h* parameter to correspond to. Consequently, the notion of a proposition being true at a world-hyperplan pair fails to get a grip.

The immediate response to this objection is that it misidentifies the role that holding or being true at a world-hyperplan pair plays within the model. As Stalnaker notes in relation to the standard possible worlds model, "[t]he role of the values *true* and *false* is simply to distinguish the possible worlds that are members of the selected subset from those that are not." (1987: 2) The point here is that from the perspective of the model itself, a proposition being true at

[5] More work would need to be done here to connect standards with the intuitive characterisation of normative attitudes given above. This is because standards are less obviously connected to action than planning and To-Do Lists. Ridge achieves this by appealing to standards of practical reasoning – see Chapter 2.

a world simply consists in that world being a member of that proposition. Similarly, on the world-hyperplan model, a proposition being true at a world-hyperplan pair consists simply in that pair being a member of that proposition. Thus, like other logical and propositional notions, truth at a world or a world-hyperplan is defined set-theoretically: p is true at <*w*, *h*> just in case <*w*, *h*> is a member of p. As such, the possible worlds model does not presuppose a correspondence theory of truth, or any other theory of truth. The minimal role assigned to truth and falsity can apply just as much to sets of world-hyperplan pairs as to sets of possible worlds. (Note also that being true at a world or <*w*, *h*> pair is a technical notion and does not necessarily correspond to our ordinary conception of 'truth', whatever that is.)

*Unjustifiably ad hoc?*

A second objection grants that although there is no in principle reason that a nonfactual parameter *cannot* be added to our model of belief content, such an extension is unjustifiably ad hoc. It cannot be that we are justified in adding a nonfactual parameter simply because it helps expressivists. What we need, rather, is independent motivation for positing this parameter to model normative belief. I think this is a fair worry to raise, but that a number of things can be said in response. The first is simply that modelling normative belief content in terms of (centred) worlds alone is an instance of the representational picture of normative thought. Insofar as expressivists have good reasons for rejecting this picture, they have good reasons for rejecting this model of normative belief content. Relatedly, insofar as expressivists have good reasons for thinking that the kind of discriminations we make in normative judgments are in some sense practical, then given discrimination relativity, they have good reasons for positing something like a hyperplan parameter. So general arguments for expressivism should provide some justification for the added complexity. In the next chapter, I will explore in more detail exactly how standard arguments for expressivism might be employed to justify the world-hyperplan model of content.

Quite independently of expressivism, however, another reason to add complexity comes from an argument due to Alex Silk (2013). Silk points out that it is plausible that some normative propositions are necessarily true. For example, it is plausible that it is necessarily true [that torturing the innocent for fun is bad]. If this proposition is necessarily true, then it is true in all possible worlds. Therefore, the possible worlds model of content would characterise this proposition as the set of all possible worlds *W*. From this, it apparently follows that understanding the proposition [that torturing the innocent for fun

is bad] consists in being able to distinguish *W* from the empty set. Moreover, it apparently follows that this proposition is identical to any other necessarily true proposition. But both claims are implausible. We have already encountered the problem of distinguishing between necessarily equivalent propositions above, and this is just an instance of the more general problem. Note, however, that the metalinguistic response seems far less plausible here. A more ready response is to introduce an additional parameter to model normative content that is not world-characterising. Understanding a normative claim can then be taken to involve distinguishing (say) which hyperplans are ruled in and ruled out by some normative belief. A necessarily true normative proposition might then be true relative to all possible worlds but not to all hyperplans. This shows that quite independently of considerations in favour of expressivism, there is reason to introduce further complexity into the model.

*The disjunction problem*

The third objection is known as the disjunction problem. I here follow the presentation of the problem given by Schroeder (2015: 12ff), but it is also raised by Charlow (2015: 10ff) and Starr (2016: 373f). Earlier, I contrasted accepting factual contents as ruling out possible ways the world might be with accepting normative contents as ruling out what to do, think, or feel. A natural thought might then be that an agent's total belief state supervenes on her total factual belief state and her total normative belief state. This could then be represented as the pair <*W*, *H*> where *W* is the set of worlds and *H* is the set of hyperplans left open by her total belief state. For any arbitrary proposition [p], an agent believes [p] just in case <*W*, *H*> is a subset of [p].

The problem with this view, however, it that it fails to characterise mixed disjunctive beliefs where the subject does not believe either disjunct. Thus, where [p] is a factual proposition and [q] is a normative proposition, it follows from the definition of disjunction given above that a subject believes [p or q] just in case <*W*, *H*> is a subset of the union of [p] and [q]. Suppose, however, that a subject believes [p or q] but does not believe [p] and does not believe [q]. The problem with current view is that there is no region of <*W*, *H*> that represents this belief. Because worlds and plans are independent parameters, the union of [p] and [q] will always be a subset of either [p] or [q]. As such, the model will always characterise a subject's belief [p or q] as believing at least one disjunct.

The correct response to the disjunction problem is to point out that the world-hyperplan model, as defined above, does *not* model an agent's total

belief state as the total set of worlds and hyperplans <*W*, *H*> compatible with her beliefs. Rather, according to the world-hyperplan model, an agent's total belief state is the total set of world-hyperplan *pairs* compatible with her beliefs. Because this provides us with a more fine-grained notion of belief content, it allows us to adequately characterise mixed disjunctions. This is because by representing points in logical space as <*w*, *h*> pairs, accepting a mixed disjunction involves ruling out certain combinations of <*w*, *h*> pairs without ruling out any *w* or any *h*. And if accepting a proposition does not rule out any *w* or any *h*, then accepting that proposition does not entail accepting any purely normative or descriptive belief.

Although this provides a technical fix to the problem, it raises a question of its own. For any formal distinction in our theory of belief content, we need a psychological distinction that the formal distinction models. However, Schroeder is sceptical that the world-hyperplan model can be grounded in a satisfactory philosophy of mind. For although individuating a subject's belief states in terms of some <*W*, *H*> pair faces the disjunction problem, we nonetheless have a clear grip on the functional roles of the factual and normative beliefs that are meant to ground this assignment. By contrast, the world-hyperplan model "needs to give a richer and more powerful characterization of the functional role of belief, which can distinguish between *arbitrary* sets of world-norm pairs, and not just by their projections onto the world or norm axes." (Schroeder 2015a: 21) And Schroeder is sceptical that any such characterization will be forthcoming. I agree with Schroeder that this is precisely the question that the proponent of the world-hyperplan model needs to answer. But I am less sceptical about the prospects for providing a satisfactory answer. In the next chapter, I propose a theory of belief that aims to provide a characterisation of belief that can adequately ground the world-hyperplan model.

*Inconsistency*

The fourth and final objection concerns whether the world-hyperplan model is entitled to the notion of inconsistency defined above. To present this problem, I will first explain Gibbard's own interpretation of the world-hyperplan model and how it faces a problem explaining inconsistency. I will then go on to explain how the interpretation being offered in this chapter differs from Gibbard's and why it provides a more attractive alternative.

How should we interpret the world-hyperplan model? Gibbard provides a psychologistic interpretation of the model according to which sets of <*w*, *h*> pairs are nothing other than a useful heuristic to represent properties of mental

states. Specifically, he argues that sets of <*w*, *h*> pairs represent properties of *hyperplanners*, ideal agents who are fully opinionated with respect to what to do and how things are. On this view, invoking an abstract space of worlds and hyperplans is fruitful only insofar as it is isomorphic with properties of hyperplanners. Ultimately for Gibbard, world-hyperplan contents can be dispensed with in an explanatory account of belief. However, this aspect of his view raises some problems.

One such problem is whether Gibbard's view can provide a satisfactory explanation of inconsistency in belief. If belief contents are merely a useful heuristic that can be dispensed with in explanatory account of belief, then we cannot appeal to the fact that two beliefs have inconsistent contents to explain why those beliefs are inconsistent. Rather, the definition of inconsistency must be explained in terms of the more basic properties of mental states. To explain this difference, Schroeder (2008: 48) distinguishes between A-type and B-type inconsistency, where A-type inconsistency involves a single "inconsistency-transmitting" attitude towards inconsistent content, and B-type inconsistency involves distinct and "logically unrelated" attitude types toward the same content. (Baker and Woods [2015] point out that B-type inconsistency is better defined as inconsistency in attitude that is not A-type, because Schroeder's definitions are not exhaustive.) Schroeder then argues that whereas A-type inconsistency is well understood and theoretically respectable, B-type inconsistency is not theoretically respectable "because there are no good examples of it." (2008: 48) Given that Gibbard's explanation of inconsistency appeals to the basic inconsistency-properties of our beliefs, he is therefore helping himself to a B-type notion of inconsistency that stands in need of explanation.

One might worry that Schroeder's claim that there are no good examples of B-type inconsistency is too strong. Baker and Woods (2015) argue that there are a number of theoretically respectable examples of B-type inconsistency that are relatively well understood. For example, consider comparative preferences. Comparative preferences are not incompatible in virtue of having inconsistent contents; after all, the content of any comparative preference is a pair of inconsistent propositions. Credal states provide another example. Again, inconsistent credal states are not inconsistent simply in virtue of their contents, but partly in virtue of the specific credence one places in those contents. A further example is the inconsistency that holds between intentions and beliefs, for example if I intend [p] and intend [q] but believe [that p only if not-q]. Holding these attitudes seems to be a clear case of inconsistency, but it cannot be A-type inconsistency for the reason that it involves distinct attitude types.

Second, and relatedly, A-type inconsistency, even if it is familiar, stands in need of *some* explanation. On this score, Gibbard (2012: appendix 2) argues that his view is no worse off than rivals that appeal to A-type inconsistency. Whereas Gibbard helps himself to unexplained notions of disagreement and conceptual entailment between mental states, rival approaches help themselves to something like *conceptual consistency of content*, which explains inconsistency for A-type attitudes such as factual belief. Gibbard argues that this notion is unexplained because conceptual consistency cannot be reduced to alethic or metaphysical possibility. For example, although it is not possible both that I am drinking water and I am not drinking $H_2O$, it is not *conceptually* inconsistent to believe that both are true. So Gibbard argues that the explanatory burdens of his view and his opponent's are more or less parallel, as each approach starts off helping itself to something unexplained.

So I think it is too quick to dismiss Gibbard's explanation of inconsistency simply on the grounds that B-type inconsistency is not as theoretically respectable as A-type inconsistency. Even so, I think that there are other good reasons for rejecting a B-type explanation of inconsistent normative beliefs. First, B-type explanations eschew explaining inconsistency in terms of believing inconsistent contents. However, as I argued in Chapter 1, there are good reasons in general to think that propositions should play an explanatory role in one's theory of attitudes, and so we should expect propositions to play a role in explaining inconsistency. Second, while there are a number of different ways in which an agent can be rationally incoherent, when an agent holds inconsistent normative beliefs, we take her to be (in some sense) *logically* inconsistent. And A-type explanations in particular seem well placed to explain this kind of rational incoherence (we'll examine this issue further in the next section). Third, as Yalcin (2012; 2018a) argues, there is no reason why we cannot accept the world-hyperplan model while rejecting Gibbard's psychologistic interpretation. Just as the standard possible worlds model assigns abstracta as the contents of beliefs without interpreting these abstracta as being properties of mental states, we can interpret the world-hyperplan model as assigning abstracta as the contents of beliefs without interpreting these abstracta as being properties of mental states. Rather, we are interpreting these abstracta respectively as being possible ways the world might be and possible ways of acting (broadly construed).

This last point shows that despite my use of Gibbard's formal apparatus, the view being offered is somewhat different to his overall picture, or at least how he presents it. But it also raises a question. If the world-hyperplan model assigns abstract contents to normative beliefs, in virtue of what is this assignment correct? Following Yalcin (2018a), I call this *the foundational*

*question*. The expressivist who accepts this view of normative belief content needs an answer to this question compatible with expressivism. However, the foundational question is a familiar question that any theory of content needs to answer, and there are a number of competing answers within the philosophy of mind already available. In this way, I think that conceiving normative beliefs as A-type attitudes is theoretically more conservative than Gibbard's approach (see Yalcin 2018a for discussion). In this chapter and the next, I examine two existing frameworks for answering the foundational question and show how expressivists might utilise them. By contrast, Gibbard attempts to build a foundational philosophy of mind from the ground up in order to meet the needs of his expressivism. It would be beyond the scope of this discussion to provide a complete assessment of Gibbard's (2012) foundational picture, and nothing I have said here rules out such an approach. However, in showing that the expressivist can utilise the world-hyperplan model while adopting a more conventional approach to explaining mental content, I hope to show that such a radical departure is under-motivated.

## 4.3 The foundational question

To repeat, the foundational question is this: In virtue of what is an agent's state of belief well modelled by a given set of world-hyperplan pairs? This question applies to any given model of belief content, and this is no less true in the present case. So the world-hyperplan model is incomplete as a full account of normative belief. We can answer the foundational question by providing what I will call a *foundational theory of belief*. This is a theory that explains the more fundamental facts about the nature of belief and propositional attitudes more generally that ground the facts about what attitudes have what contents. Notice that the foundational question as it is posed here is just a more specific instance of the more general question that motivated the unity requirement in Chapter 2. Namely, in virtue of what do our attitudes have the content that they have?

In what follows, we will therefore need to consider whether the answers on offer adequately respect the unity requirement. After rejecting Yalcin's answer to the foundational question in the next section, I will go on to provide my own answer in the next chapter that both respects the unity requirement and answers the objections that were left unresolved in the previous section. First, however, I want to argue that although any theory of content needs to answer the foundational question, answering this question is more pressing for expressivists. This is because expressivists face a particular challenge

explaining (a) inconsistency in belief and (b) why normative beliefs are nonfactual.

### *4.3.1 The negation problem (again)*

The first problem relates to how the expressivist explains inconsistent normative attitudes on the current picture. Among other things, a theory of normative belief should explain when and why two normative beliefs are inconsistent. To recap our discussion of this problem in Chapter 1, consider the following belief states:

(1) Believing that one should help the meek.
(2) Believing that one should not help the meek.
(3) Believing that it is not the case that one should help the meek.

Certain relations hold between the mental states described in (1)-(3). For example, (1) is inconsistent with (2) and (3). (2) commits one to (3) but not vice versa. Traditional versions of expressivism have trouble explaining these facts. This, we saw in Chapter 1, is known as the negation problem (Unwin 2001; Schroeder 2008). Suppose that 'should' expresses the attitude of *requiring*. We can then say that (1) and (2) consist in the following:

(1*) Requiring that one helps the meek.
(2*) Requiring that one does not help the meek.

What about (3)? First, note that (3) must be distinct from:

(4) Not believing that one should help the meek.

Which would consist in:

(4*) Not requiring that one helps the meek.

This is because (3) consists in a settled view on the permissibility of not helping the meek, whereas (4) consists in lacking the mental state described by (1). Thus, (4) could be true while (3) false in cases where one has not considered the question, has no view of the matter, or suspends judgment. A natural response to this problem would be to say that (3) consists in:

(3*) Permitting that one does not help the meek.

Intuitively, this suggestion seems plausible. However, note that we have postulated a distinct attitude type to account for (3). This looks problematic because we have simply postulated an attitude type with the relevant properties required to capture the relations that hold between (3) and other attitudes, whereas our theory was meant to *explain* these relations. Now, given the discussion in the previous section of A- and B-type inconsistency, perhaps some more concrete account of this attitude might be forthcoming. However, in lieu of such an account, we seem to lack an explanation of inconsistency for normative attitudes. Hence, the negation problem.

Initially, it might seem as if the world-hyperplan model is in a good position to explain inconsistency and solve the negation problem. This is because normative attitudes can be explained as inconsistent in virtue of having inconsistent contents. Given that the definition of inconsistency in content given above applies equally to sets of world-hyperplan pairs as it does to sets of possible worlds, it might seem that the explanation of inconsistency is just as good as the standard explanation of inconsistent factual beliefs. However, while the world-hyperplan model mirrors the *form* of A-type inconsistency as it applies to factual belief, we haven't yet been given any *explanation* of why accepting inconsistent world-hyperplan contents engenders attitudinal inconsistency. This is because we have yet to be given a story about why it is wrong to accept a normative proposition and its negation (compare Starr 2016: 368f; Willer 2017: 197ff).

Compare this with the representational picture of belief. For our prosaically factual beliefs, the standard view is that accepting a proposition and its negation is inconsistent in virtue of the representational properties of belief. Factual beliefs represent the world as being some way or other. Because there is no way the world can be such that some factual proposition and its negation are both true, it is inconsistent to accept both claims for the reason that one's beliefs will necessarily fail to fulfil their representational function. Exactly how one fleshes out these claims about factual belief will vary depending on one's preferred view of the nature of belief and mental representation, a topic we will return to. But the basic picture seems clear enough: A-type inconsistency for ordinary belief is not left unexplained (pace Gibbard). Rather, the representational account seems to provide a straightforward and attractive explanation of inconsistency. But it is one unavailable to the expressivist to explain normative belief.

So although adopting the world-hyperplan model allows expressivists to retain the formal definition of propositional inconsistency, they cannot retain the standard explanation of inconsistency in belief. They need some other

explanation. So without answering the foundational question, the world-hyperplan model leaves inconsistency in normative attitudes unexplained. In this respect, the model itself does little to improve on the more traditional expressivist notion of primitive inconsistency in attitude. Moreover, given the unity requirement, we should expect the explanation of inconsistency to be fully general across factual and normative beliefs. This means that fundamentally, inconsistency in factual beliefs will also need to be cashed out in nonrepresentational terms. ('Fundamentally' is important here because the expressivist does not need to deny that inconsistent factual beliefs are representationally inconsistent; it will just be that such inconsistency is a species of a more general genus.)

### *4.3.2 The neutrality of the world-hyperplan model*

Another reason why the foundational question poses a particular challenge for expressivism is that although the world-hyperplan model of content is designed for a nonfactualist theory of normative thought and discourse, it is in fact compatible with factualism about normative thought and discourse as well. The formalism itself does not entail the truth or falsity of factualism or nonfactualism about normative belief. For example, a normative realist could maintain that the $w$ parameter of a world-hyperplan proposition picks out the non-normative core of possible worlds and the $h$ parameter picks out the prescriptions mandated by what reasons objectively exist (Sinnott-Armstrong 1993: 300f; Kalderon 2007: 73).

Similarly, a number of relativist theories also appeal to a distinct normative parameter to model normative content without thereby accepting expressivism about normative thought and discourse (e.g. MacFarlane 2014). Here, the normative parameter is taken to be the normative standards that a speaker or community implicitly accept, but the normative beliefs are themselves taken to be representational in the same way that a prosaically factual belief is. It is just that the relevant states of affairs are represented relative to a certain perspective. In short, the truth of the world-hyperplan model does not entail expressivism. Whether the world-hyperplan model supports nonfactualism depends on how we answer the foundational question (Stalnaker 2012: 201f makes a similar point). So the expressivist's answer to the foundational question must explain why accepting a world-hyperplan proposition does not involve representing the world as prescribing the actions not ruled out by those hyperplans.

## 4.4 Yalcin's plan expressivism

In this section, I examine an answer to the foundational question provided by Yalcin (2018a). Yalcin follows Gibbard in construing belief contents as sets of world-hyperplan pairs. Like Gibbard, he also takes it that such contents model plan-laden belief states, understood as states of mind that involve a view of what to do as well as a view of how things are. To answer the foundational question and ground the world-hyperplan model within a foundational philosophy of mind, Yalcin appeals to Lewis' analytic or common-sense functionalism (see in particular Lewis 1986, 1999 and Stalnaker 2004 for discussion).

According to Lewis, propositional attitudes are constitutively rational mental states that cause an agent's behaviour. He argues that folk psychology provides an implicit theory of the functional roles of mental states, which implicitly defines mental states according to certain principles of rationality. He then argues that there are physical states (e.g. neural states) that approximately realise the constitutively rational causal roles specified by the theory. Folk psychological states can thus be reduced to physical states via a Ramsey sentence that corresponds to our folk theory. Given this picture, the foundational question about our mental states in general can be answered as follows: "The intentional mental states of an agent have the content that they have in virtue of the fact that the behaviour of the agent can be explained as rational behaviour on the hypothesis that the agent's mental states have that content." (Stalnaker 2004: 205)

To see how this foundational theory of mind grounds a particular model of content, we first need to understand the rational principles that constitutively govern our attitudes. Focusing first on the centred worlds model, Yalcin (2018a: 25) proposes that this correctly models agents' belief states in virtue of the following rational principle that constitutively governs belief, desire, and action:

(P1) If agent *A* is in a belief state with the centered content *B* and in a desire state with centred content *D*, then *A* is disposed to act in ways that would tend to bring it about that he is located within *D*, were it the case that he occupied a centered world within *B*.

Although this is perhaps the most prominent principle of folk psychology, it is not exhaustive of the principles governing content determination. Yalcin also notes that for Lewis, an agent's belief state is not simply that which is most apt to produce behaviour compatible with (P1). Lewis also suggests an *eligibility*

constraint, which rules out deviant interpretations that are intuitively incorrect but give the right predictions, of the kind given by Putnam's (1981) model-theoretic argument against determinate reference. Very roughly, the constraint is that reality has a structure, and our beliefs are sensitive to that structure (in contrast to a "gruesome" or gerrymandered structure). While Yalcin is right to include this principle in addition to (P1), both are in fact part of the more general principle that folk psychology "is the constitutive theory not just of instrumental rationality but of rationality generally." (Lewis 1986: 39) So all such principles applied in determining an agent's beliefs are principles of rationality, and there are many principles of rationality other than (P1) and the eligibility constraint (see Lewis 1999: 320).

In the next chapter, I will examine further the idea that folk psychology is the constitutive theory of rationality generally. For now, however, we can ask how the world-hyperplan model might fit into Lewis' foundational philosophy of mind. To this end, Yalcin (2018a: 26) postulates the following interconnections between normative attitudes and other attitudes in addition to (P1):

(P2) Agents are disposed to act in ways which would conform with their plans, in centred worlds with respect to which their (purely factual) belief content is true.
(P3) Where agents' plans leave several options open, agents tend to elect those options which would serve best to satisfy their preferences.
(P4) Where agents find themselves in unplanned for situations, we appeal only to belief and desire.

Although Yalcin puts forward these suggestions as initial approximations, it seems plausible that these sort of principles would determine something like the world-hyperplan model as the correct model of agents with planning attitudes. Moreover, it also seems plausible that the kind of planning attitude implicitly defined by (P2)-(P4) is suitably nonfactual and close to how expressivists understand normative thought. So Yalcin's answer to the foundational question looks initially promising.

However, I think that Yalcin's proposal should be rejected for two reasons. The first reason is that Yalcin's account fails to adequately explain inconsistency in normative belief. It seems plausible that it is rationally incoherent in some way to accept incompatible plans. We adopt plans in order to achieve our ends, and in adopting incompatible plans, we are likely to thwart our ends. However, as I argued in the previous section, when we accept inconsistent normative propositions, we are rationally incoherent in a

particular way. Specifically, we are *logically* inconsistent. Moreover, we seem to be logically inconsistent in exactly the same way as when we accept inconsistent factual claims. However, it is unclear how Yalcin's account is meant to explain how accepting inconsistent normative propositions engenders logical inconsistency. Moreover, it is unclear how it is meant to explain this in a way that respects the unity requirement. This is because the fundamental explanation of what makes normative beliefs inconsistent seems to be different to the fundamental explanation of what makes factual beliefs inconsistent. More generally, despite having the form of an A-type explanation of inconsistency, it is unclear whether Yalcin's theory is really any better than Gibbard's in relation to explaining inconsistency, as Yalcin's theory also helps itself to an unexplained notion of inconsistency in attitude.

The second reason that Yalcin's account should be rejected is that it faces the disjunction problem. Although Yalcin's official line is that propositions are world-hyperplan pairs, his foundational theory of belief actually supports a different view. Specifically, it supports the view discussed by Schroeder that an agent's total belief state is characterised by the pair <*W*, *H*> of worlds and hyperplans compatible with her belief state. This is because the constitutive principles (P1)-(P4) do not individuate a functional role that distinguishes between arbitrary sets of <*w*, *h*> pairs. Rather, it individuates distinct functional roles for factual belief and normative belief. This means that mixed beliefs will supervene on combinations of purely factual and purely normative beliefs. But as we have seen, this view fails to characterise mixed disjunctions. So Yalcin fails to provide a foundational theory of attitudes that adequately answers the foundational question.

## 4.5 Conclusion

I think that the failure of Yalcin's proposal ultimately comes down to a failure to respect the unity requirement. That the failure to respect the unity requirement leads to problems explaining inconsistency is fairly clear. But arguably the same failure leads to the disjunction problem. After all, the disjunction problem arises because we lack a suitably general attitude that can distinguish between arbitrary sets of world-hyperplan pairs. If this is correct, then what expressivists need to answer the foundational question is a unified foundational theory of propositional attitudes that fundamentally characterises factual and normative beliefs as one and the same kind of attitude. (Again, 'fundamentally' is important here because the expressivist will still want to say that there is a distinction in kind between normative and

factual beliefs; it is the fundamental characterisation of these attitudes that must be the same.)

I propose that the best way for expressivists to provide an answer to the foundational question that respects the unity requirement is to provide a fully general account of belief that applies equally to factual and normative beliefs. This account should explain a number of things. First, it should explain why belief as a general attitude is *not* essentially representational — if belief as a general attitude were essentially representational, then there would be no room for an expressivist account of normative belief. Second, it should explain why *some* beliefs are representational, even if belief as such is not representational. This is because expressivists want to maintain a distinction between normative and descriptive beliefs. Third, it should adequately ground the world-hyperplan model of content — according to which points in logical space are <*w*, *h*> pairs and not individual *w* and *h* coordinates — in order to avoid the disjunction problem. Although contemporary expressivists generally accept that normative statements express normative beliefs, this is typically maintained in a minimal or deflationary sense. By contrast, I am proposing that expressivists should develop an explanatory but nonrepresentational theory of belief and propositional attitudes more generally in which to situate their theory of normative thought and discourse. In the next chapter, I provide such a theory.[6]

Before proceeding, it might be worth forestalling a possible objection. I claim that expressivists need a general theory of belief in which to situate normative judgment. However, isn't expressivism by definition the denial that normative judgments are non-deflationary beliefs? Isn't belief eo ipso representational? My answer is twofold. First, there already exist accounts of belief that are nonrepresentational in the relevant sense, so simply assuming otherwise seems to beg the question (e.g. pragmatist, deflationary, and perhaps certain varieties of inferentialist and functionalist theories). Second, as mentioned in Chapter 1, the sense of 'representational' in question is ultimately a theoretical notion. So the expressivist can grant that it might be a conceptual truth that belief is 'representational' in some platitudinous sense. However, if this is a platitude then it must be respected as such and cannot presuppose a metaphysically robust interpretation (see Divers and Miller 1995). Moreover, there are a

[6] One might wonder whether Horgan and Timmons' (2006) 'cognitivist expressivist' theory provides such an account of normative belief. I'm not sure it does. Although Horgan and Timmons make a good case defending the *claim* that normative judgments are beliefs, they do not have a general theory of belief that explains *why* normative judgments are beliefs. Another example might be Schroeder's (2013) account of belief as states of *being for*. I examine this suggestion in the Chapter 6.

number of distinctions that expressivists can appeal to here. For example, one might distinguish between two senses of 'world' — the world as everything that is the case, and the world as an ontologically robust environment. While all beliefs represent the world in the former sense, the expressivist might claim, only some beliefs represent the world in the latter sense (compare Price's [2013] distinction between *i*-representation and *e*-representation).

# 5
# Interpretative Expressivism

This chapter motivates and develops an expressivist theory of normative belief. By combining an interpretationist account of propositional attitudes with the world-hyperplan model of belief content examined in the previous chapter, I argue that expressivists can and should maintain that our ordinary normative commitments are beliefs not merely in a deflationary sense but in a theoretically robust sense. The plan is as follows. First, I situate the proposal that expressivists should develop a general theory of belief within a broader dialectic (section 5.1). Second, I introduce the interpretationist framework and make the case for an expressivist theory of normative belief within this framework (section 5.2). I call the resulting view 'interpretative expressivism'. Third, I argue that interpretative expressivism has a number of benefits over rival versions of expressivism (section 5.3). Before concluding, I consider and respond to an objection concerning the normativity of attitude attributions and whether this is compatible with interpretative expressivism (section 5.4).

## 5.1 Introduction

In the previous chapter, I examined the modal conception of content, the idea that content should be explained in terms of possibilities. I argued that this provided a suitably unified framework in which to develop a nonfactual notion of normative propositional content. In particular, I suggested that expressivists would do well to adopt the world-hyperplan model of content. However, I argued that any such approach needs to be grounded in a foundational theory of mind compatible with expressivism. And I argued that this should take the form of a fully general theory of belief according to which normative beliefs are nonrepresentational. This chapter aims to provide such a theory, therefore completing the account of normative content articulated in the previous chapter. However, before proceeding, it might first be helpful to take stock and situate the account proposed in this chapter within a wider dialectical context.

In a nutshell, expressivism is the view that normative statements express nonrepresentational, nondescriptive, or nonfactual states of mind. Early versions of expressivism took this to mean that normative statements express

emotions or desires (Ayer 1936; Stevenson 1937). More contemporary versions maintain that normative statements express sui generis normative attitudes with a distinctive nonrepresentational functional profile (Gibbard 1990, 2003; Blackburn 1998). Other versions claim that they express complex attitudes such as a belief-desire pair (Ridge 2014). Although contemporary expressivists claim that normative attitudes can be beliefs in a minimal sense, what unites these views is the claim that, fundamentally, normative attitudes have a different psychological profile to prosaically factual beliefs.

Notoriously, expressivists have a hard time explaining what attitudes are expressed by logically complex statements with normative elements (the literature is vast — see Woods [2017] for an up-to-date overview). While it might seem intuitive that "stealing is wrong" expresses something like *disapproval* of stealing, it's less clear what attitude might be expressed by "stealing is wrong or it doesn't cause pain". After all, one can seemingly accept this statement without disapproving anything. Contrast this with a complex descriptive statement like "stealing is not enjoyable or it doesn't cause pain". Here, the statement expresses a prosaically factual belief, the complexity of which is explained in terms of its content. That is, it expresses the belief in the proposition [that stealing is not enjoyable or it doesn't cause pain].

The driving idea of this thesis has been that expressivists would do well to mirror the descriptivist explanation in the normative case by providing a theory of normative propositions that can play the role of belief contents. The expressivist could then explain the attitude expressed by a complex statement with normative elements in terms of the content of the belief it expresses. In the previous chapter, I examined a view according to which normative belief contents are modelled in terms of the sets of nonfactual possibilities ruled out by one's normative beliefs. Such a view, pioneered by Allan Gibbard (1990, 2003), has recently been enjoying a resurgence (Yalcin 2012, 2018a; Silk 2013, 2015; Charlow 2014; Pérez Carballo and Santorio 2016). Further, we saw that it has also been employed to develop expressivist views in a number of other domains, such as epistemic modality (Yalcin 2007, 2011; see also Field 2009, 2018), mathematical thought (Pérez Carballo 2016), and know-how (Santorio 2016). Given that expressivists about normative thought are typically sympathetic to expressivist treatments of other domains, this approach to modelling nonfactual content should therefore be attractive for expressivists.

However, recent proponents of this approach have generally said little about the constitutive question of *what it is* for a thought to have nonfactual content. This has led to some scepticism about how much is achieved by characterising normative thought and discourse in terms of sets of nonfactual contents (e.g. Schroeder 2015; Starr 2016; Willer 2017). More generally,

expressivists who appeal to nonfactual contents need to explain the role these contents play within an independently plausible and satisfactory philosophy of mind. Those who have addressed the constitutive question have typically tried to find a suitable noncognitive attitude that has the structure ascribed to it by their formal models. For example, Silk (2015) proposes that normative contents model properties of *weak preference*, of the kind studied by decision theory. Yalcin (2018a) proposes they model properties of *planning states*, a kind of sui generis state defined in terms of its unique functional role (see Bratman 1987; Gibbard 2003).

It seems to me, however, that expressivists should take a more radical approach. With the introduction of normative content, it is open to expressivists to answer the constitutive question in terms of what it is to *believe* some normative content. While most contemporary expressivists claim that normative statements express normative beliefs in a minimal or deflationary sense, they deny that this provides any informative answer to the constitutive question. By contrast, I am proposing that expressivists should claim that normative statements express normative beliefs in a theoretically robust sense, where this is explicated by an independently plausible theory of propositional attitudes. I will argue that given the role that certain theories of propositional attitudes ascribe to belief contents, a suitably nonfactualist theory of belief can be achieved by incorporating nonfactual contents into the theory.

The remainder of this chapter is divided into two main sections. In the first, I explore how expressivism can combine with an *interpretationist* theory of propositional attitudes to deliver a theory of belief that is suitably nonfactualist.[1] 'Interpretationism' covers a fairly diverse family of theories (most prominently Davidson 1980; Lewis 1984; Dennett 1987a; see Child 1996 for an overview). Here, I will adopt something close to Dennett's *intentional stance strategy*. There are two parts to this strategy. The first part is to highlight the role that propositional attitudes play in the explanation and prediction of action or behaviour, broadly construed. The second is to argue that our understanding of propositional attitudes is exhausted by this role. Very roughly, for an agent A to believe [p] just is for A to be disposed to act as if A

[1] Blackburn (1998) also adopts a broadly interpretationist strategy in developing his expressivism; however, he accepts a representationalist construal of belief within this theory, answering the constitutive question in terms of the noncognitive attitude of *valuing*. Gibbard (1990) provides some intimation towards an interpretationist view of belief, though provides no elaboration and in later work seems to take a somewhat different view. Ridge (2018) argues that expressivists can adopt an interpretationist construal of *normative certitude* in terms of counterfactual betting behavior; in certain respects, the view developed in this chapter can be seen as extending this approach to normative propositional attitudes in general.

believes [p] according to the best theory of interpretation, where this theory specifies the conditions under which it is correct to ascribe attitudes to an individual in order to predict and make intelligible her behaviour. In this context, the expressivist's task is to explain why the best theory of interpretation for agents with normative attitudes should appeal to nonfactual contents. I argue that standard arguments for expressivism can be utilised for this purpose.

With the interpretationist cum expressivist theory of normative belief explained, I then outline the benefits of accepting such a view over versions of expressivism that explain descriptive and normative judgments as being fundamentally different kinds of attitude. I argue for four claims. First, interpretative expressivism has the virtue of theoretical parsimony, in that it postulates the existence of one rather than two (or more) kinds of attitudes. Second, it coheres well with an independently plausible theory of propositional attitudes. Third, it is well placed to explain mixed disjunctions and mixed attitudes more generally. Fourth, it is well placed to explain normative inconsistency. So not only does this theory of normative belief provide a distinct and novel way of implementing the expressivist programme, if my arguments are correct it also improves upon existing versions of expressivism.

The resulting view not only constitutes an improvement in the prospects for expressivism. It also challenges a widely assumptions about the nature of belief. First, it challenges the assumption that beliefs by their nature have a distinctive mind-to-world direction of fit. This follows from the thesis that normative beliefs are nonrepresentational. Second, it challenges the assumption that beliefs by their nature are motivationally inert. This follows from the thesis that normative beliefs are practical. However, I will argue that these assumptions are not built into the interpretationist view of belief and that it is thus well placed to make sense of a nondescriptivist conception of belief. With the motivation for the view located within a wider dialectical context, I will now proceed to explain the view.

## 5.2 Interpretative expressivism

In this section, I explain how expressivism about normative thought can combine with an interpretationist theory of propositional attitudes to provide a suitably nonfactualist theory of belief. First, I set out the interpretationist framework and show how the modal conception of content finds a natural home within this framework. Second, I show how the world-hyperplan model

can be employed within the interpretationist framework to provide an expressivist-friendly theory of belief.

### *5.2.1 The interpretationist framework*

Interpretation is the process of ascribing beliefs, desires, and other propositional attitudes to subjects or agents on the basis of what they do and say. It is something we do all the time. Suppose Meredith is running across Covent Garden. Why is she doing this? Because she *wants* to see *La Traviata* and *believes* it's about to start. She hasn't picked up her ticket yet, so we can predict that she'll first go to the box office as she *knows* she won't get in without a ticket. Unfortunately, Meredith has forgotten her glasses and so she *hopes* that her seat isn't too far back. Upon *discovering* that she is up near the gods, we can expect Meredith to be *disappointed*. However, music lover that she is, we also expect her to overcome her disappointment and *tell* all her colleagues the day after that it was a wonderful performance. The predictions and explanations offered by such a narrative are not infallible and require a background of other assumptions. But the idea that we ascribe attitudes to others (and ourselves) in the service of the explanation and prediction of action should be familiar enough.

Dennett (1971; 1987a) proposes that we make sense of this kind of interpretation by distinguishing between three different predictive strategies. First, we have the *physical stance*, which involves predicting behaviour on the basis of the physical state of an object together with the laws of nature. Second, we have the *design stance*, which involves predicting behaviour on the basis of an object's function (e.g. artefacts and biological objects). The design stance allows us to make predictions without any knowledge of the physical underpinnings of the object in question. For example, I do not need to know the physical laws governing my alarm clock in order to predict when it will ring. Third, we have the *intentional stance*, which involves ascribing intentional attitudes to an object under the assumption that it is a rational agent, and then predicting it to behave in ways that are rational given its attitudes. This is the stance we took towards Meredith and that we utilise in folk psychology more generally. Although less reliable than the other stances, the intentional stance allows us to effortlessly understand and predict a vast amount of human behaviour, a task of otherwise immense complexity.

The intentional stance is first and foremost an epistemology of propositional attitudes. Dennett's further metaphysical claim is that for an agent to possess propositional attitudes just is for that agent to be "reliably and voluminously" predictable using the intentional stance (1987b: 15). Thus, all it is for an agent

A to believe [p] is for the best (i.e. most predictive) interpretation of A to assign to A the belief [p]. Importantly, however, whether an agent believes [p] is an objective matter, as the predictive success of the interpretation is grounded in real patterns of behaviour (Dennett 1987b: 25ff; 1991). That said, interpretationists like Dennett accept that there is often a certain amount of indeterminacy about what an agent believes, though this is seen as reflective of the phenomena rather than as a problem for the view. Overall, Dennett's interpretationism can be seen as a kind of dispositional view of belief — for an agent to believe some proposition is for that agent to be disposed to behave in certain ways under certain conditions (Dennett 1987c: 50). Spelling out exactly what ways and under what conditions is the task of intentional systems theory, or the theory of interpretation, which aims to make explicit the rules of attribution implicit in folk psychological practice.

An initial gloss on these rules can be given as follows (Dennett 1987b: 17):

> Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many — but not all — instances yield a decision about what the agent ought to do; that is what you predict the agent *will* do.

Thus according to Dennett, the interpretative principles that constitutively govern propositional attitudes are *rational* principles about what agents ought to think, want, and do.[2] Beliefs and desires are therefore constitutively rational in the sense that they are dispositions whose manifestation conditions are specified in terms of norms of rationality. Accordingly, we can understand what it is for something to be a belief, say, rather than a desire, in terms of the respective norms by which it is constitutively governed.

What exactly are these norms? While Dennett takes the assumption of rationality to play a "crucial role" (1987d: 96) in his theory, he claims that the relevant concept of rationality is "systematically pre-theoretical" and broad in scope (1987d: 98; Davidson 1980: 241 and Lewis 1986: 38f, 1996: 320ff make similar claims). Because of this, Dennett resists any attempt to provide a precise characterisation of the relevant notion of rationality. Nonetheless, he provides a number of general suggestions as to the kind of norms he takes to be

[2] Does it follow that attitude attributions are normative? I return to this question below in section 5.4.

constitutive of rationality. Most centrally, beliefs and desires are governed by the principle of instrumental rationality: rational agents will act in ways that satisfy their desires in light of their beliefs. Beliefs are also distinctively governed by norms of theoretical rationality, such as norms relating belief to truth and evidence. Thus, rational agents will have mostly true beliefs about those parts of the world they have had exposure to, relative to their perceptual capacities and what is relevant to their interests (Dennett 1987b: 19). Sophisticated agents will also acquire new beliefs via reasoning, and eliminate inconsistencies when brought to light, relative to interests and resource limits (ibid: 20; c.f. Harman 1995). What desires a rational agent ought to have will be determined in part by the nature of the agent in question. In virtue of the biologically evolved nature of human beings, Dennett proposes that we attribute to people a stock of basic desires such as "survival, absence of pain, food, comfort, procreation, entertainment" and so on (Dennett 1987b: 20; see also Anscombe 1957: §36-37; Lewis 1999: 320). Derived desires can be attributed on the basis of what the agent believes, such as desiring what an agent takes to be a means to an already desired end.

And so on. Although the precise details are no doubt important and interesting in their own right, it will not be necessary to settle these details in relation to the more general purpose of this chapter of making the case for an expressivist theory of normative belief. Moreover, I freely admit that there is a lot of room for debate about these details and the overall viability of interpretationism in the philosophy of mind more generally (for discussion of some of these issues see Dennett 2009), though I do think that interpretationism is an independently attractive view. But the general idea that I want to work with in addressing the specific topic of this chapter is that we assign to an agent the propositional attitudes that best rationalise that agent's behaviour in light of her nature, capacities, and sensory history. A believes [p] just in case A's dispositional profile is best rationalised by A believing [p], where this is fixed by the norms implicit in folk psychology. It is matter of contention precisely what these norms are, and indeed if the relevant notion of rationality is ultimately codifiable (this notion comes from McDowell [1979] in relation to practical rationality, but Child [1996: ch.2] argues that it is central to interpretationism more generally). While below I will make substantive claims about the distinctive norms that govern normative attitudes, their precise formulation and codifiability will not be at issue. Dennett's interpretationism and the expressivist theory of normative belief developed below are compatible with a range of answers to these questions. So although at the end of the day these details will need to be ironed out, it is not necessary at this stage of our inquiry to settle these questions in advance.

Although we will return to this in more detail below, we can already begin to see why Dennett's interpretationism is particularly amenable to an expressivist theory of normative belief. Interpretationism denies that beliefs are *intrinsically* or *fundamentally* representational, whatever exactly that might amount to. Rather, insofar as an agent represents the world in believing [p], she does so *implicitly* in virtue of her dispositional profile. This opens up the possibility that the dispositions involved in believing a normative proposition do not involve any implicit representation of one's environment. Moreover, if we have in hand a notion of nonfactual belief content that can adequately characterise normative attitudes, then we can put this to work in individuating the relevant disposition that constitutes believing a normative proposition. This is because interpretationism is a *non-reductive* theory of intentionality. Recall that according to interpretationism, A believes [p] just in case A is best interpreted as believing [p]. Here, belief and belief contents figure in the right-hand side of the analysis, meaning that propositions will play an explanatory role in characterising belief.

In the next subsection, I will argue that the world-hyperplan model can be put to work within a Dennettian theory of interpretation to provide an expressivist theory of normative belief. However, it's first worth highlighting that the modal conception of content provides the most natural picture of content for the kind of interpretationist view being offered here.[3] To see why, consider again the simplest version of the modal conception:

> **The possible worlds model of content**. The content of a state of belief is representable as a set of possible worlds, intuitively the worlds "left open" by what is believed. Propositions are sets of possible worlds, and the propositions an agent believes are those true with respect to all of those worlds the state leaves open.

Something very close to the possible worlds model of content is endorsed by Dennett (1987e; other dispositionalist views similar in spirit also endorse this model — see Lewis 1986; Stalnaker 1987). There are a number of reasons why this model fits well with Dennett's interpretationism. Most of these reasons have already been touched upon in the previous chapter as general attractions of the possible worlds model, but it will be worthwhile seeing how these general attractions relate to interpretationism in particular.

First, consider how interpretation is a holistic process in which the assignment of any particular belief is always dependent on an agent's other

[3] Of course, this will not be true of other kinds of interpretationist views, such as Davidsonian views that give an essential place to language in interpretation.

beliefs. This feature of interpretation is nicely captured by the possible worlds model in how the model takes an agent's *total* belief state to be primary, rather than individual beliefs. Second, it is an important part of Dennett's view that we can take the intentional stance towards non-human agents. Because possible worlds propositions are unstructured, agents are not required to possess language or entertain structured thoughts in order to have beliefs. Third, interpretationism typically admits that there is often a certain amount of indeterminacy concerning whether an agent in fact believes some proposition. In the previous chapter, we saw that the possible worlds model captures this in several ways. One way was by positing fragmented belief states to agents. Another way was to characterise the domain of possible worlds as *discrimination relative* (Dennett 1987e: 207 explicitly endorses this view). Because agents with different discriminatory abilities will divide up possibilities in different ways, this can result in indeterminacy when an agent's beliefs cut up reality less finely than the proposition under consideration. Plausibly, this will often be the case when attributing beliefs to non-linguistic animals. Fourth, and relatedly, Dennett (1987c: 48-9) stresses that we are intentional systems in virtue of our evolved nature. Discrimination relativity allows us to capture the developmental continuum from differentially responding to an environment to possessing the discriminative capacities characteristic of complex language use.

Appreciating the way in which belief content is discrimination relative is central to understanding the sense in which beliefs are representational on the interpretationist view. When I believe [that Tibbles is on the mat], this proposition makes explicit what is represented by my belief state. However, in believing this proposition, I represent Tibbles being on the mat *implicitly* in virtue of my sensory and behavioural relation to my environment. Specifically, I respond rationally to the sensory information that indicates that there is a certain cat on the mat (given my history and other attitudes) and I behave in ways that would be rational given my belief about Tibbles (given my other attitudes). It is in virtue of these facts that it is rational to attribute to me a certain discriminatory ability, namely the ability to divide up a domain of possibilities into those in which Tibbles is on the mat and those in which Tibbles is not.

As we saw in the previous chapter, discrimination relativity is significant because it allows the expressivist to argue that the particular discriminatory ability involved in believing a normative proposition consists in something other than an ability to discriminate between alternative factual possibilities. By drawing a comparison between normative judgments and plans, I

suggested that the discriminatory abilities involved in believing normative propositions might ground the following model of content:

> **The world-hyperplan model of content**. The content of a state of belief is representable as a set of <*w*, *h*> pairs, intuitively the <*w*, *h*> pairs "left open" by what is believed. Propositions are sets of <*w*, *h*> pairs, and the propositions an agent believes are those that hold with respect to all of those <*w*, *h*> pairs the state leaves open.

If the possible worlds model provides the right model of content for Dennett's interpretationism, then we might hope that the world-hyperplan model provides the right model of content for an interpretationist cum expressivist theory of normative belief. Our task will be to argue that the interpretation of agents with normative attitudes involves attributing discriminative abilities that are characterised by the world-hyperplan model. Moreover, we must do so in a way that is compatible with expressivism. If successful, then we will have adequately answered the foundational question with respect to the world-hyperplan model raised in the previous chapter.

### *5.2.2 Nonfactual belief in interpretationism*

To recap, the view being offered is that beliefs are behavioural dispositions individuated by constitutive principles of rationality. What makes some attitude a belief and not another attitude is the particular principles by which it is constitutively governed and which supply the basis for interpretation. What makes it the case than an agent believes [p] rather than [q] is that her actions are best predicted by interpreting her as believing [p] rather than [q] under the assumption of rationality. As such, beliefs are not intrinsically representational. Rather, that one represents the world in believing [p] is a consequence of the predictions this interpretation affords. In the previous chapter, I examined a view according to which belief contents encode not only factual information about an agent's environment but also nonfactual information of a directive nature. In this section, I argue that by incorporating this view of content within interpretationism, a case can be made for an expressivist theory of normative belief.

The basic idea is this. Agents with normative attitudes require a richer interpretative framework than that provided by the possible worlds model of content. This fact is explained in terms of distinctive rational principles that respectively govern purely normative and purely factual beliefs. However, these differences notwithstanding, both kinds of attitude are governed by the

constitutive norms of belief *tout court*. Thus, on the interpretationist view, both attitudes are fundamentally beliefs. My argument for this view proceeds in two steps. First, I mobilise traditional arguments for expressivism to support the claim that normative and factual beliefs are respectively governed by distinctive rational norms. Among other things, these arguments explain why believing a normative proposition does not involve an implicit representation of how things are normatively. Second, I argue that normative judgments are fundamentally beliefs because they are subject to the constitutive norms of belief. As well as providing an expressivist theory of normative belief, this view has the interesting corollary that belief as such is nonrepresentational. While this might seem surprising, I argue that given interpretationism this is in fact less surprising than it might appear.

*Step 1*

The first claim to defend is that the interpretation of normative agents (i.e. agents with normative attitudes) requires (at least) the richer framework of the world-hyperplan model of content as opposed to the possible worlds model. The justification for this claim must ultimately come from the constitutive principles of rationality governing our attitudes. I suggest two principles which taken together might provide this justification.

The first principle comes from the debate about motivational internalism. Arguably, something like the following is true of first-person ought-beliefs (see Ridge 2015):

> **Normative Internalism**. Necessarily, for any fully rational agent A, if A believes [that A ought to Φ in C], then A will intend to Φ in C.

Normative Internalism receives support from the observation that in a number of contexts, an agent's believing [that she ought to Φ in C] is sufficient to explain her Φ-ing in C. For example, if I come to believe [that all things considered I ought to give more money to famine relief], and a UNICEF collector comes to my door to collect money, then one would expect me to give money (Smith 1994: 6). If I did not give money, then we would search for some countervailing factor, such as an overriding desire to hold onto my money, or my not having any money to give. But if I do give money, this seems perfectly intelligible even if I have no particular desire to give to the collector. However, if I had no particular desire to give to the collector, then my giving money requires some other explanation. If Normative Internalism is true, then my action can be explained by my judgment (and the intention it brings about). So

we might think of Normative Internalism as a norm of rationality that constitutively governs our attitudes.

Some care here needs to be taken to explain what is meant by the claim that A's believing [that she ought to Φ in C] is *sufficient* to explain her Φ-ing in C. For it cannot mean that an interpretation that assigned *only* this belief could explain A's action. This is because interpretation is essentially holistic — whether an agent possesses some attitude always depends on the totality of her beliefs *and* desires. Moreover, it is clear that in the UNICEF example we implicitly attribute many other beliefs and desires in order to make sense of my action (e.g. the belief [that the collector is from UNICEF], the desire [that if I give the collector money then this will contribute towards famine relief], and so on). So the notion of one's normative belief "sufficiently explaining" one's action needs some other explication. The suggestion being offered is that an explanation of an action typically requires not merely the presence of some related desires but an independent desire *to do that thing* (under some suitable description). Normative Internalism then claims that if an agent believes [that she ought to Φ in C], then, other things being equal, she will Φ in C regardless of whether she has an independent desire to Φ.

Normative Internalism alone does not support expressivism because it leaves open the possibility that normative beliefs are both intrinsically motivating *and* representational. So we need a second principle to complete step one of the argument. Noncognitivists often appeal to internalism in conjunction with a Humean belief-desire psychology according to which beliefs are motivationally inert. If normative statements express beliefs as I am arguing here, then the Humean picture must be rejected as it stands. However, it is still possible to respect some of the intuitions motivating the picture within our interpretationist framework.

Arguably, one idea driving the Humean view is that "cold representations" of the world just aren't the kind of thing that alone could rationalise action. If an agent's total belief state is given in terms of the set of world-hyperplan pairs it leaves open, we can pick out the *purely factual* or *plan-invariant* subset of beliefs as those beliefs where no *h* is ruled out. The suggestion is then that these beliefs are constitutively governed by the following principle:

> **Representational Inertness**. Necessarily, for any fully rational agent A, action Φ and set of purely factual propositions P, A's believing P is not sufficient to explain A's Φ-ing.

(Where "sufficient to explain" is understood as above.) Arguably, Representational Inertness captures what motivates the Humean view of belief

once we drop the representationalist assumption that all beliefs represent reality. Importantly, the principle is compatible with the view that *some* beliefs are sufficient to explain actions. It is just that such beliefs cannot be purely factual.

Thus, taken together, Representational Inertness and Normative Internalism support the claim that an agent's normative beliefs must be something other than beliefs in purely factual propositions. Moreover, if we individuate beliefs using <*w*, *h*> contents, the link between an agent's normative beliefs and her actions is reflected in what is prescribed by the set of hyperplans she accepts, which makes explicit what she accepts in virtue of her dispositional profile. While this provides us with a suitably nonfactualist view of normative belief, it's important to see that this is not entailed by the world-hyperplan model itself. As we saw in the previous chapter, the model is consistent with a realist view according to which *w* denotes reality stripped of its normative features and *h* denotes the norms prescribed by objective reasons (Sinnott-Armstrong 1993: 300f; Kalderon 2007: 73), as well certain relativist views (e.g. MacFarlane 2014). Instead, the nonfactuality of normative belief follows from the fact that the disposition individuated by world-hyperplan contents on this view contains no implicit representation of the agent's environment, i.e. the world she inhabits. In other words, the psychological profile of accepting a normative proposition entails no ability to discriminate between normative ways the world might be and entails no sensitivity to normative reality. Moreover, if we did take this sort of directive disposition to contain an implicit representation of the agent's environment, it would violate Representational Inertness.[4]

What reason is there to accept Representational Inertness? One kind of consideration invoked by expressivists derives from Moore's open question argument. The idea is that for any factual proposition that we might believe, there is always an "open" question of what to do about it, or more generally what normative significance to give it. For any belief about how the world is, "[e]ven if that belief were settled, there would still be issues of what importance to give it, what to do, and all the rest. For we have no conception of a 'truth condition' or fact of which mere apprehension by itself determines practical

[4] This last claim is complicated by the fact that one might interpret 'purely factual propositions' in the realist-friendly way suggested in the text so that it is compatible with Representational Inertness. Three things can be said in response. First, although such a view might be possible, I think that it is undermotivated given the dispositional profile of normative beliefs on this picture. Second, the expressivist arguments for Representational Inertness discussed below constitute arguments against this view. Third, insofar as expressivists think there are general reasons to be sceptical of descriptivist views, these arguments provide indirect support for the expressivist interpretation of 'purely factual'.

issues. For any fact, there is a question of what to do about it." (Blackburn 1998: 70; see also Darwall, Gibbard and Railton 1992: 116ff) However, one might wonder if this is just a restatement of Representational Inertness rather than an argument for it. So the expressivist needs to find some non-question begging way to establish the thesis.

That said, if the principle is a constitutive norm of belief, one might wonder what sort of argument would establish this. Compare: what sort of argument would establish that the principle of instrumental rationality was a constitutive norm of belief? One might think that the dialectical burden here is negative in the sense that the defender of Representational Inertness would need to show that putative explanations of an agent's actions in terms of purely factual beliefs alone are in fact irrational or unintelligible. So the kind of argument the expressivist gives might depend on the nature of facts purportedly believed in normative belief. It would be beyond the scope of the scope of this chapter to survey the arguments, but see Horgan and Timmons (1992) for argument that natural facts cannot play this role and Dreier (2015a; 2015b) for argument that non-natural facts cannot play this role.

Expressivists might also appeal to facts about disagreement to motivate the appeal to nonfactual contents. The problem arises in cases of fundamental disagreement where two individuals have a different conception of the subject matter in question and systematically respond to different features of reality (Björnsson 2017: 277). For instance, consider Hare's (1952: 148) example of the Missionary and the Cannibals. In Hare's example, the Missionary learns that the Cannibals have a term similar to 'good' in that it is a general adjective of commendation. However, whereas the Missionary applies the term to people who are meek and gentle, the Cannibals apply it to people who collect a large number of scalps. In such a case, it seems plausible that the two parties have a different conception of the subject matter of what is good. However, it also seems plausible that the two parties are in disagreement. This seems to stand in contrast to purely factual disagreement. In the purely factual case, when two parties have a radically different conception of the subject matter or are systematically responding to different features of reality, we interpret their respective beliefs to be about different things, and so not in disagreement. Arguably, therefore, we cannot capture fundamental normative disagreement in terms of two parties disagreeing over some purely factual proposition. Thus, if the disagreement is a disagreement in belief, then this supports the claim that normative beliefs are individuated by nonfactual contents (Field 2018: 15 makes a similar argument).

Again, this is not the place to settle this debate. My aim is rather to show that arguments traditionally used to support the claim that normative attitudes

are noncognitive can be used to support the claim that they are beliefs with nonfactual content, in the sense spelled out above. Given that expressivists hold something like Normative Internalism and Representational Inertness anyway, proceeding this way should look attractive to expressivists. This concludes the first step of the argument, which was to defend the claim that the interpretation of normative attitudes requires world-hyperplan contents in virtue of the distinctive norms of rationality that govern normative and factual beliefs.

*Step 2*

Given that I have argued that factual and normative beliefs are constitutively governed by different norms, it might seem natural to conclude that factual and normative beliefs are fundamentally distinct kinds of attitude. I want to resist this thought. In the remainder of this section, I argue that both kinds of attitude are fundamentally beliefs because they are both governed by the constitutive norms of belief tout court, and so are fundamentally beliefs according to interpretationism.

First, both normative and factual beliefs are governed by the same norms of theoretical rationality, which apply exclusively to beliefs. For example, the theory of interpretation involves the assumption that agents more or less follow certain rules of logic and reasoning (Dennett 1971: 95). These rules apply to an agent's beliefs. Thus, we predict that an agent will acquire new beliefs and eschew old ones through inference and reasoning, as well as eliminating inconsistencies in light of new evidence. Given that sets of <*w*, *h*> pairs follow the same logic as sets of worlds, all the same predictions can be made about how an agent will reason to and from beliefs with normative content. The non-reductive nature of interpretationism is important here, because there is no requirement to explain intentional notions like inconsistency in content in more basic, non-intentional terms (for example, in terms of representational failure — we'll return to this in the next section).

Further, the holistic nature of belief explains how an agent can acquire normative beliefs in virtue of responding to features of reality without the need for explicit reasoning. For example, if belief acquisition is something that applies to total belief states, then an agent who (a) accepts hyperplans that rule out inflicting gratuitous pain and (b) perceives someone inflicting gratuitous pain can be predicted to (c) acquire the belief that that person's actions are impermissible without the need for explicit reasoning. So the rules of interpretation pertaining to theoretical rationality apply just as much to beliefs with world-hyperplan contents as beliefs with possible worlds contents. Given

that normative agents are in fact disposed to act in these ways, this supports the claim that these attitudes are beliefs.

Moving to the constitutive norms of practical rationality that govern beliefs, consider the principle of instrumental rationality: that rational agents act in ways that fulfil their desires given their beliefs. If ever there was a constitutive norm of belief, presumably this is it. It might appear less obvious that this principle applies to normative beliefs as well as factual beliefs. First, if we accept Normative Internalism, then the paradigm case of normative beliefs motivating actions will be directly rather than via their interactions with desires or other attitudes. However, it might also appear less obvious because the principle of instrumental rationality is often cashed out in metaphysically robust terms. For example, desire satisfaction is often cashed out in terms the factual content of the desire being realised, where realisation is a metaphysically substantive notion. However, given the non-reductive ambitions of interpretationism, there is no requirement to reduce intentional notions to metaphysical ones. So we can still talk about an agent's desires being "realised" even if these desires involve nonfactual content.

For example, suppose Alex believes that tax avoidance is wrong. Other things being equal, it follows from Normative Internalism that we should predict that Alex will pay his taxes in full. However, suppose that things are not equal. Suppose that Alex has a strong desire to earn as much money as possible, and this desire leads him to avoid paying his taxes in full whenever the opportunity arises. After some soul searching, however, Alex comes to form an overwhelming desire to avoid wrongdoing. The next day it's time for Alex to complete his tax returns. What will he do? It would be reasonable to expect that Alex will pay his taxes in full, as we know he believes that doing so will be a way of "bringing it about" that he avoids wrongdoing, where the content of his desire is specified in terms of set of worlds in which he acts in ways required by the hyperplans left open by first order normative inquiry. So Alex's belief that tax avoidance is wrong is subject to the principle of instrumental rationality in much the same way as any of his purely factual beliefs are. (Ridge [2018: 10, 18] makes a similar point that so long as expressivists can appeal to normative propositions, they can capture these kinds of explanations.)

In sum, I have argued that our normative and factual beliefs are both subject to the same constitutive norms of belief tout court. It follows given interpretationism that both attitudes are fundamentally beliefs. I have also argued that the additional constitutive norms of normative and factual belief in particular explain how an agent's normative beliefs cannot be interpreted in terms of factual contents alone. The resulting picture is a robust theory of belief

according to which some but not all beliefs are representational. In contrast to expressivist theories that only allow for a deflationary or quasi-realist conception of normative beliefs, the view being offered here allows for a theory of normative belief in the full-blooded sense of the term that is compatible with expressivism. Further, if some beliefs are nonrepresentational, it follows that belief as such is nonrepresentational. This might sound alarming to many. However, if we accept the interpretationist picture, I think this is less surprising than it might seem. If beliefs are fundamentally dispositions, then the sense in which our beliefs are representational is already derivative. And once we attend to the sorts of dispositions characteristic of normative belief, and we see that it is not at all obvious that such a disposition implicitly represents the world, we are further loosened from the grip of the representationalist picture.

As I have indicated at various points, there are many explanatory burdens that one would need to take on to fully defend the sort of view outlined here. But these are burdens that already exist for any interpretationist theory and any expressivist theory. So if my arguments are correct, then the view outlined above is at least as good as rival expressivist theories that claim that normative and factual beliefs are fundamentally different kinds of attitudes. In the next section, I argue that the view on offer is preferable to rival expressivist views.

## 5.3 Pros

Most contemporary expressivists accept that normative statements express normative beliefs. This is typically embraced in the quasi-realist spirit of accommodating the realist-sounding features of normative discourse within an expressivist framework (e.g. Blackburn 1993). While such expressivists accept normative belief-talk "at the end of the day", the challenge for them is to earn their right to normative belief-talk given a noncognitivist account of normative thought according to which normative judgments are fundamentally desire-like attitudes rather than beliefs. This differs from the proposal set out above. The traditional approach is to start with the thesis that factual and normative beliefs are fundamentally different kinds of attitudes, and then go on to explain how both can be properly thought of under the title of 'belief'. My approach is to start with the thesis that factual and normative beliefs are fundamentally the same kind of attitude (i.e. belief), and then go on to explain how this is compatible with an expressivist account of normative thought. Thus, the order of explanation is reversed.

Call these respective approaches *bifurcated attitude expressivism* and *unified attitude expressivism*. Given that both approaches aim to accommodate

normative belief, it is natural to ask which approach is comparatively more attractive. In this section, I argue that unified attitude expressivism has four important benefits: (i) it is more theoretically parsimonious; (ii) it coheres with an independently plausible philosophy of mind; (iii) it is better placed to explain mixed attitudes; and (iv) it is better placed to explain inconsistency. These last two benefits answer two of the problems raised in the previous chapter for expressivist theories that accept the world-hyperplan model of content. Comparing the two approaches will also clarify some of the commitments of the unified approach. Not only this, however, but it will exhibit certain resources available to expressivists that embrace a robust theory of normative belief that are otherwise not available. As such, these considerations show how the account being offered improves upon rival expressivist theories.

### *5.3.1 Simplicity*

Unified attitude expressivism is simpler than bifurcated attitude expressivism in that the former only postulates the existence of a single attitude type where the latter postulates the existence of two or more attitude types. Consider the two attitudes denoted by 'the belief that grass is green' and 'the belief that murder is wrong'. According to unified attitude expressivism, these are two instances of a single attitude type that are distinguished by their contents. According to bifurcated attitude expressivism, these are distinct kinds of attitude. Whereas the former is a belief, the latter is some distinct kind of nonrepresentational attitude. While the bifurcated approach as such is compatible with a number of ways of cashing out this distinction, it is typically explained in terms whether the attitude has a representational or nonrepresentational *characteristic functional role*, where factual beliefs have the former and normative beliefs have the latter.

If normative and factual beliefs are fundamentally distinct kinds of attitude, it is likely that attitude types will multiply further. Indeed, it has been argued that certain versions of expressivism imply an infinite hierarchy of such attitudes (Schroeder 2008: 49ff; though see Shiller 2016 for an argument that this implication is unproblematic). This will be (at least) to account for mixed attitudes, such as that picked out by "the belief that grass is green or murder is wrong". Given the normative component, this attitude cannot be fundamentally representational in nature. Given the representational component, it cannot be fundamentally directive in nature. So it seems that we need to introduce some third kind of attitude to account for mixed thoughts. Perhaps this might be something like Blackburn's (1988) notion of *being tied to*

*a tree*. However, given that this attitude is not fundamentally representational or directive, yet is somehow made up of components that are both, we need some characterisation that explains this attitude.

While simplicity may not count for too much on its own, it is worth emphasising that bifurcated attitude expressivism posits differences where it appears that there are none. This is because we use all the same locutions to talk about normative beliefs as we do other kinds of beliefs. From a pre-theoretical perspective, normative beliefs seem to be just one among many other kinds of beliefs. Moreover, from a theoretical perspective, normative beliefs have all the same properties that are central to other kinds of beliefs, such as their inferential and inconsistency properties, as well as a number of phenomenological properties (Horgan and Timmons 2006). Given that we seem to have a unified explanandum, we should expect a unified explanans. But this is exactly what bifurcated attitude expressivism denies. Thus, simplicity here is not just a theoretical virtue in and of itself, but seemingly demanded by the very facts we are attempting to explain (though see Ridge 2009 for an attempt to meet this objection head on).

This point is strengthened when we observe that expressivists not only owe us an account of normative and mixed beliefs, but of other types of normative attitude, such as desires, hopes, doubts, presuppositions, etc. This is known as *the many attitudes problem* (Rosen 1998: 393ff; Schroeder 2010: 83f). If the expressivist explains normative belief as distinct from factual belief, then it would seem that she would have to provide piecemeal explanations of all other normative propositional attitudes as well. By contrast, unified attitude expressivism provides a unified account not only of belief but of propositional attitudes more generally, and so no such problem arises.

### *5.3.2 Theoretical consonance*

As developed above, unified attitude expressivism coheres well with an independently plausible philosophy of mind. This is for the simple reason that we began by adopting an interpretationist framework and then developed an expressivist theory within that framework. By contrast, metaethical debates about the nature of normative judgment tend to remain neutral about the exact nature of belief and the mind more generally. While this is fine as far as it goes, we should expect our theories in each domain to converge given the overlap in subject matter. Moreover, I think that the discussion above shows that separation of the subject matter can obscure some of the theoretical possibilities available in metaethics.

Further, consider an analogy with the relation between metaethics and the philosophy of language. Traditionally, expressivists have rejected truth-conditional semantics for normative language. This means having to construct an alternative semantic theory from the ground up. A number of philosophers complain that given the orthodoxy and general utility of truth-conditional semantics, it is a large cost for expressivists to abandon this general framework for considerations local to metaethics. Indeed, a number of philosophers now argue for reconciliation between expressivism and truth-conditional semantics (e.g. Ridge 2014; Chrisman 2016; Köhler 2018).

Arguably, bifurcated attitude expressivism faces an analogous problem in that it must construct an alternative philosophy of mind from the ground up (at least for expressivists who claim that normative judgments are a sui generis kind of attitude). Moreover, it must do so for reasons local to metaethics. While this objection is in no way decisive, it seems good methodological practice to integrate our best theories from different domains where we can. And this is exactly what unified attitude expressivism aims to do.

### *5.3.3 Mixed attitudes and the disjunction problem*

The previous considerations put forward in favour of unified attitude expressivism concern its more general theoretical virtues. However, bifurcated attitude expressivism faces a number of challenges that unified attitude expressivism is better placed to solve or perhaps avoid. To be clear from the outset, I am not claiming that unified attitude expressivism faces *none* of the problems faced by bifurcated attitude expressivism, or that it faces *no* challenges explaining mixed attitudes and inconsistency. Rather, the claim is that the unified approach is *better placed* to solve or avoid these problems. The first problem I will examine is the problem of mixed disjunctions and mixed attitudes more generally.

The mixed attitudes problem refers to the problem of explaining the nature of attitudes that involve both normative and non-normative content, such as the belief [that stealing is wrong or it never causes pain]. According to bifurcated attitude expressivism, to believe [that stealing never causes pain] is to be in a state with a representational functional role which aims to track the world, and to believe [that stealing is wrong] is to be in a state with a conative functional role which aims to guide action. If these are fundamentally distinct state types, what type of state is the mixed belief?

As we saw in the previous chapter, a natural suggestion is that mixed beliefs are combinations of purely normative and purely factual beliefs. However, we also saw that this views fails to explain mixed disjunctions. Because it is

possible to believe a disjunction without believing either disjunct, the belief [that stealing is wrong or it never causes pain] cannot be captured by any combination of purely normative and purely factual beliefs (Schroeder 2015: 12ff; see also Charlow 2015: 10ff; Starr 2016: 373f). There is then some pressure to explain the disjunctive state as some kind of inferential commitment (Blackburn 1988; Chrisman 2016: 178ff). However, we are now owed an account of this third type of state and how it relates to atomic factual and normative beliefs.

By contrast, there is no special account needed for mixed beliefs if we adopt unified attitude expressivism. On the version I have outlined above, mixed beliefs are just beliefs with world-hyperplan contents. The nature of such beliefs is explained in terms of the conditions under which the theory of interpretation says it is correct to attribute beliefs with this kind of content. And this will be explained fundamentally in terms of an agent's overall dispositional profile, where the attribution of a logically complex belief will involve (inter alia) dispositions to reason in certain ways, broadly construed. However exactly this explanation goes, it is not fundamentally different in kind to explanations involving purely factual or purely normative beliefs.

So an agent A will believe a mixed disjunction [p or q] without believing [p] or believing [q] when the best interpretation of A's actions assigns only the first belief and not the other two. Here is one such possible situation: A is *not* disposed to act as if she believes [p] and she is *not* disposed to act if she believes [q], but A *is* disposed to form the belief [p] upon learning [not-q] and to form the belief [q] upon learning [not-p]. If we could only interpret A's beliefs using the unstructured set <*W*, *H*> compatible with her beliefs, then assigning to A the belief [p or q] without assigning either [p] or [q] would not be an intelligible predictive stance that we could take towards A. This stance is made intelligible, however, if we adopt an interpretative framework that employs sets of <*w*, *h*> pairs. Moreover, because to believe [p or q] without believing [p] or believing [q] *just is* to be predictable in this sort of way according to interpretationism, there is no further question about what this belief must be like in order to have these properties.[5]

It's worth stressing the role of the holistic aspect of belief in this response to the disjunction problem. Recall that one of Schroeder's (2015: 21) worries about solving the disjunction problem with the world-hyperplan model of content was that we have no grip on the functional role of a state that discriminates between arbitrary sets of world-hyperplan pairs. However, this objection

[5] There may be a further question about the sub-personal mechanisms in virtue of which human beings are intentional agents. But this is an empirical question to be answered by the cognitive sciences. See Hornsby (2000) for discussion.

presupposes that beliefs are individuated atomistically. Specifically, that they are individuated by their canonical functional role. By contrast, the interpretationist view individuates individual beliefs holistically relative to an agent's total belief (and desire) state, which is primary, where this latter assignment is governed by the totality of constitutive norms of rationality contained within the theory of interpretation. Thus, the "richer and more powerful characterisation" of belief provided by the interpretationist view comes not from the richer functional role of individual beliefs but from the holistic interpretative nexus by which we predict normative agents. Moreover, not only does the world-hyperplan model allow us to make the predictions we need of normative agents, it makes their actions *intelligible* in terms of the distinctive rational principles that govern these predictions. So there is something right to the suggestion that logically complex beliefs are characterised by their inferential commitments. But that is because *all* beliefs are characterised by their inferential commitments according to interpretationism, though of course not exclusively.

### *5.3.4 The inconsistency problem*

As we saw in the previous chapter, expressivists notoriously have a difficult time explaining how it is that normative beliefs can stand in the right sort of inconsistency relations with other beliefs (see, for example, Wright 1988; Unwin 2001; Schroeder 2008). Introducing nonfactual contents within the modal conception of content allowed the expressivist to define two beliefs as being inconsistent when the intersection of their contents is empty. This is fine as far as it goes, but the problem for this view is that it doesn't tell us *why* those beliefs are inconsistent (Starr 2016: 368f; Willer 2016: 197ff). Simply assigning formally inconsistent contents is not sufficient to explain why any two states are inconsistent in the right sense. After all, my desire that I eat cake and my desire that I not eat cake take inconsistent propositions as their contents; however, these attitudes are not inconsistent in the relevant sense. In short, we need some explanation of why beliefs with contradictory contents are inconsistent.

Assuming bifurcated attitude expressivism, a natural answer would be to look to the characteristic functional roles of each attitude type and explain how their constitutive functions are frustrated when an agent believes contents with empty overlap. In the descriptive case, we might say that beliefs with empty intersection necessarily fail to represent the world as being some way. This is because there is no way the world can be such that the contents of both beliefs are true. Thus, descriptive inconsistency engenders a failure of the

representational function of factual judgments to track the way reality is. In the normative case, if we assume that normative judgments are kinds of planning states, we might say that beliefs with empty intersection necessarily fail to prescribe a coherent contingency plan. This is because there is no way of acting such that both plans can be realised. Thus, normative inconsistency engenders a failure of the practical function of normative judgments to settle the thing to do.

There are a number of problems with this approach, however. First, one might worry whether the sort of practical inconsistency appealed to by the expressivist is of the right kind to ground *logical* inconsistency. If I plan to eat cake and to not eat cake, I presumably have an incoherent plan, but this does not obviously make me logically inconsistent. However, where [p] is a normative proposition, it is surely logically inconsistent to believe both [p] and [not-p]. Second, even assuming a suitable sense of practical inconsistency is specified, we again have distinct explanations for what appears to be a unified phenomenon. This approach therefore fails to respect the unity requirement. Inconsistency in belief appears to be of the same kind regardless of whether the beliefs are normative or factual. So given bifurcated attitude expressivism, we have a bifurcated explanation for a seemingly unified explanandum. Moreover, combinations of normative, factual, *and* mixed beliefs can be logically inconsistent. It is difficult to see how this could be true if normative and factual inconsistency were different in kind.

By contrast, because unified attitude expressivism counts all beliefs as instances of a single kind of attitude, inconsistency is explained as the same across normative, factual, and mixed contexts. Given our interpretationist theory of belief, what exactly is this explanation? The first thing to note is that given the theory's non-reductive commitments, we should not expect inconsistency in belief to reduce to some non-intentional notion, such as some kind of functional failure. Thus, trivially, to believe just is to believe as true, and believing inconsistent propositions necessarily results in believing a falsehood. If we suppose that the theory of interpretation contains or entails some epistemic truth norm for believing only what is true, then inconsistent beliefs will violate this principle.

Further, consider that interpretation is based on how an agent acts, where the attitudes ascribed to the agent provide explanatory reasons that make her actions intelligible. If beliefs are constitutively governed by rational principles, this suggests another way of thinking about what is wrong with believing inconsistent claims. Specifically, we ask what would be wrong with a fully rational agent who believes inconsistent claims. If we think of believing propositions in terms of ruling out sets of world-hyperplan pairs, then to

believe inconsistent propositions is to rule out all doxastic possibilities. Which is to say the agent has no beliefs which could serve as reasons for action. Interpretation simply breaks down. Importantly, this explanation of what is wrong with believing inconsistent propositions applies equally to normative, factual, and mixed beliefs. As less than fully rational agents with limited and fragmented cognitive powers, we find that sometimes it is intelligible to assign inconsistent beliefs to an agent. But this requires us to invoke the idea of fragmented belief systems or some other notion that can explain this. And if an agent fails to exhibit some degree of stability and consistency, then they will fail to be interpretable.

To sum up, as well as being more complex and less integrated into existing theories of mind, bifurcated attitude expressivism has a difficult time explaining mixed attitudes and inconsistency in belief. Because unified attitude expressivism provides unified explanations of these phenomena, I have argued that it is better equipped to explain these features of belief. No doubt much more needs to be said on behalf of unified attitude expressivism in order to fully explain these features of belief. But these are substantive matters that any theory of belief needs to explain. Seeing as whatever explanations interpretationism gives for these phenomena must be grounded in facts that do not fundamentally appeal to representational notions, I see no reason, at least from the outset, that they cannot be applied to beliefs with nonfactual contents.

## 5.4 Are attitude attributions normative?

The theory of belief on offer is meant to provide an account of what it is to have normative beliefs. The theory is expressivist insofar as the view respects the two core theses set out in Chapter 1. Namely, that normative thought is fundamentally (a) nonrepresentational and (b) practical. Typically, the nonrepresentational thesis is motivated both by *psychological* claims about the nature of normative thought (as offered here) as well as *metaphysical* claims about the nature of reality. Specifically, the nonrepresentational thesis is motivated in part by the claim that normative properties, relations, facts, etc., do not exist in any robust sense. Because there are no such things according to expressivists, they cannot play an explanatory role in their theory of normative thought.

Now for the problem. Interpretationism explains beliefs in terms of the rational norms that constitutively govern our attitude attributions. However, many philosophers think that the 'ought'-claims provided by these rational norms are themselves normative. It would therefore seem to follow that the

explanations offered by interpretationism are normative explanations. But this looks to be in tension with expressivism's metaphysical thesis. If there are no normative properties, relations, or states of affairs, then how is it that we can explain the nature of propositional attitudes using normative vocabulary? In this section, I examine two ways of responding to this problem. The first approach is to accept the normativity of rationality but maintain its compatibility with interpretative expressivism. The second is to reject the normativity of rationality. This is a huge topic and what I have to say here will inevitably be far from conclusive. But I think the objection is important enough to merit some indication of how it should be answered.

### *5.4.1 Constitutive rationality as normative*

Suppose that the rational norms formulated in the theory of interpretation are genuinely normative. Why might this be incompatible with expressivism? First, note that the problem is not simply that an explanation of normative thought that employs normative notions is somehow circular. By comparison, an explanation of thoughts involving logical notions need not refrain from using logical notions (Wedgwood 2007: 21; Chrisman 2016: 204f). Second, note that invoking normative notions to explain normative thought is not incompatible with the nonrepresentational thesis as such. This is because even assuming our theory of belief invokes normative notions, it does not explain normative beliefs in terms of their *being about* the existence and distribution of normative properties, relations, or facts.

Rather, the tension arises from expressivism's ontological commitments. Roughly, expressivists maintain that what (robustly) exists is that which is described by the natural sciences. The general viewpoint is nicely summed up by Blackburn (1998: 48f):

> To be a naturalist is to see human beings as frail complexes of perishable tissue, and so part of the natural order. It is thus to refuse unexplained appeals to mind or spirit, and unexplained appeals to knowledge of a Platonic order of Forms or Norms; it is above all to refuse any appeal to a supernatural order. […] [T]he problem is finding room for ethics, or placing ethics within the disenchanted, non-ethical order which we inhabit, and of which we are a part.

Blackburn also accepts an interpretationist approach to explaining the mental (though he denies that normative commitments are beliefs) and claims that given this approach, "mental states turn out to get their identity from a network of normative considerations." (1998: 54) The worry for such a view is that the "network of normative considerations" that explains normative

attitudes looks no different from an "unexplained appeal" to an "order of norms". And this is incompatible with expressivism's naturalist commitments.[6]

One might respond to this worry by challenging the implicit assumption that the expressivist theory of thought in question is itself descriptive. That is, one might reject the assumption that the normative vocabulary employed within the expressivist theory refers to anything normative. Instead, one might provide suitably nondescriptive interpretations of these notions as they are used within the expressivist theory. This would amount to providing a nondescriptivist *meta-theory* of the expressivist theory of thought. This strategy is pursued by Gibbard (2012), who first explains meaning in terms of inferential and evidential ought-claims, and then explains ought-claims in terms of planning states. More generally, this view is compatible with naturalism because it does not refer to anything distinctively normative.

However, if our expressivist theory of thought is to explain our use of normative vocabulary in general, then it had better also explain our use of normative vocabulary within the meta-theory. Otherwise, we would not have a fully general theory of normative discourse. In this way, we can say that such a theory must be *uroborosian*, in the sense that it must be able to explain itself. Moreover, if our theory was not uroborosian, then appealing to a nondescriptivist meta-theory would do no more than move the bump under the carpet, as the original problem would simply reappear at the level of the meta-theory. So in order for interpretative expressivism to make good on this suggestion, one would need to successfully apply its analysis of ought-claims to the ought-claims made within the theory of interpretation. As a first approximation, one would need show that the claim [that an agent believes p iff she behaves as if she rationally ought to believe p] can be characterised in terms of the set of <*w*, *h*> pairs at which the agent behaves at *w* as if she believes what is rationally required by *h* in *w*. I think this suggestion has some initial plausibility and I would welcome any developments of the view that proceeds this way. But I do have a number of reservations, some of which I will elaborate in more detail in the next subsection.

First, it is not obvious that accepting interpretationism consists in having any directive dispositions of the kind relevant to normative thought. However, this is what the uroborosian view predicts. Second, the uroborosian view

[6] In a critical discussion of Blackburn (1998), Dreier (2002) explores this issue in detail, arguing that Blackburn is committed to some kind of vicious regress or circularity. However, for the reasons given in this subsection, I think that the real worry is not circularity as such but rather the compatibility of the normativity of attitudes with naturalism.

entails that whether some assignment of attitudes is correct will depend on which first-order normative view is correct. Given nonfactualism about normativity, it follows that there is no worldly fact of the matter concerning whether I believe some proposition or not. One might worry whether this is really plausible. Moreover, as we saw above it is denied by interpretationists like Dennett. Third, it also follows that any social science that employs propositional attitude psychology is not purely descriptive, which might be a cost depending on one's views of such disciplines. Fourth, the uroborosian view entails the possibility of radical disagreement over what attitudes to ascribe to people in cases where two parties hold fundamentally opposed normative principles of rationality. However, this seems less plausible in the case of attitudes than in ordinary normative disagreement. Fifth, it seems possible to judge that it would be rational for A to believe [p] without judging that A has any normative reason to believe [p], which seems to be ruled out if attitude attributions are normative. While these reasons against this approach might not be decisive, I think they are jointly sufficient to look for an alternative.

Could one maintain a descriptivist meta-theory of interpretative expressivism while accepting the normativity of attitude attributions? Blackburn seems to think so: "I think of myself only as describing: what attitudes are, what desires are, what normativity is, by placing them in some satisfactory relationship with their empirical manifestations. For that enterprise to succeed, I still do not think it matters too much if the trail of the normative serpent lies over everything." (2002a: 166) Given Blackburn's avowed naturalism, this might seem puzzling. However, it seems to me that Blackburn's response trades on an equivocation of the word 'normative'. For we can distinguish a narrower sense of 'normative' that is the concern of metaethics from a broader sense of 'normative' that applies to anything that is in some way governed by norms or standards. For instance, chess is an example of a practice that is constitutively norm-governed and so 'normative' in the broader sense. However, these norms are not plausibly normative in the narrower sense. Of course, if we are playing chess, then we probably will have reasons to follow these rules. But our reasons for following these rules will not come from the norms of chess, but from some other source, such as the enjoyment of playing chess or our desire to do so.

The comparison is worth making because Blackburn (2002a: 165) appeals to chess as another example of a constitutively norm-governed activity that is describable from a naturalistic standpoint. However, if attitudes are constitutively normative only in *this* sense, then it not obvious that interpretationism employs any normative notions in the narrower sense that

expressivists aim to explain. This is further supported by Blackburn's claim that interpretationist claims about (say) what one ought to do if one believes some proposition are "statements about what you would expect from an agent. […] [T]he 'ought' here speaks not of duties and values, but just about what you would expect." (1998: 56) More generally, it is implausible that all uses of 'ought' are normative in the narrow sense (see Chrisman 2016: 26ff and passim). So I think that Blackburn's response is better seen as rejecting the normativity (narrowly construed) of rationality.

### *5.4.2 Constitutive rationality as descriptive*

Suppose that the rational norms formulated in the theory of interpretation are not normative. On this supposition, no tension arises between rationality and naturalism, and there is no requirement to provide a nondescriptive meta-theory of the expressivist theory of normative thought. This is not to deny that interpretation is constitutively norm-governed. But it is to deny that these norms are normative in the robust sense that is the interest of metaethics. What reason is there to accept this view? I have already mentioned a number of reasons against the normativity of rationality in the previous subsection, but it will be worth exploring some of them in a little more depth. This is because I want to argue that rejecting the normativity of rationality is the better response to the initial worry because it is independently plausible that rationality is not normative.

It seems plausible that normative discourse is distinguished from other domains of discourse in terms of the distinctive practical role that it plays within our lives. Such a view is available to expressivists and descriptivists alike. I haven't said too much about the precise nature of this role, but it seems plausible that it has something to do with motivating actions and coordinating attitudes in thinking how to live. In accepting normative principles, we are disposed to be governed and guided in the ways recommended by those principles. In evaluating whether rationality is normative, one question we can therefore ask is whether accepting claims about what one rationally ought to do, think, and feel plays this role in our lives.

To be clear, the question is not whether these rational principles play a governing role in how we interpret others, for that is precisely what they do. Rather, the question is whether the rational principles by which we interpret others are ones that we accept in normative deliberation and discussion. By comparison, it might be correct to say that our thoughts and actions are governed by the principles of geometry insofar as we use those principles to guide our geometrical practice, but it does not follow that the principles of

geometry are normative in the relevant sense. For the purposes of our discussion, it is also important to frame this question in terms of the *role* of rational principles in our lives because it is coherent to maintain that (i) constitutive rationality is not normative but (ii) the correct first-order normative theory coincides with the principles of rationality (Ridge 2014: 231). What needs to be shown, then, is whether the principles of rationality *as we employ them in interpretation* are normative.

I think that there are a number of considerations that support a negative answer to this question. First, if the rational principles governing interpretation were normative, then it would follow that only agents with normative attitudes would be capable of attributing propositional attitudes. However, this seems implausible. For example, it is surely possible that some alien species similar to human beings could lack normative attitudes but nonetheless have beliefs and the capacity to attribute beliefs to others. Closer to home, some empirical studies to suggest that non-human animals possess some degree of a theory of mind (for discussion see Povinelli & Vonk 2004), but it seems implausible that all these animals have normative attitudes. Further, empirical studies show that typical human beings develop a theory of mind between the ages of 3 and 4 years old (e.g. Wimmer & Perner 1983), but it is not at all obvious that this coincides with the development of normative attitudes, though admittedly this is somewhat speculative. In short, the normativity of rationality places strong constraints on the kind of creatures that can interpret others, which arguably conflicts with empirical evidence

Second, there are a number of cases in which our judgments about what is rational and our judgments about what normative reasons we have seem to come apart. Recall, for instance, Normative Internalism, which roughly stated that one (rationally) ought to intend to Φ if one believes [that one (normatively) ought to Φ]. While this might be plausible as a norm of rationality governing normative belief, it is not plausible that believing [that one ought to Φ] gives one a normative reason (to intend) to Φ. Otherwise, this would lead to unacceptable "bootstrapping" where we can generate normative reasons simply by adopting or forming certain attitudes. (Moreover, as Kolodny [2005] argues, the norm of Normative Internalism in particular cannot be saved from bootstrapping objections by narrow-scope interpretations.) Importantly, the divergence here is not simply between our judgments about what we rationally ought to do and our judgments about what we all-things-considered ought to do. This alone would not be sufficient to support the claim that rationality is not normative because it might be that the reasons provided by rationality are outweighed by the other reasons we have. Rather, the claim is that there are cases where we may judge that it is rational to Φ without judging that there is

*any* (pro tanto) normative reason to Φ. But this possibility is ruled out if judgments of rationality are normative.

Third, even if one denies the normativity of rationality, one should not have to deny that '(ir)rational' does have normative uses and that we sometimes have normative reasons to do what we rationally ought. To explain these cases in a way that is compatible with denying the normativity of rationality, we can divide them into two groups. First, there are cases where we do have normative reasons to follow the norms of rationality but where the source of normativity resides elsewhere. Second, there are cases where we do not have any normative reason to follow the norms of norms of rationality but where we can explain the appearance of there being such a reason.

As an example of the first kind, consider an agent A who flouts the principle of instrumental rationality. In such a case, it would be natural to point out to A that she is acting irrationally by failing to adopt the necessary means to her ends while still maintaining those ends. Moreover, it would be natural to offer this as a *criticism* and correspondingly *advise* that she should alter either her ends or intentions. However, criticising and advising in this way seem to be distinctively normative: "Criticizing someone involves more than the judgement that the criticized person has violated some standard; it also involves the judgement that the standard is authoritative for her. And […] this means that the person has decisive reasons to conform to this standard." (Kiesewetter 2017: 25) This observation leads philosophers like Kiesewetter to the conclusion that rationality is normative.

However, assuming that A has a normative reason not to violate the principle of instrumental rationality, I think that this is more plausibly explained by factors other than the normativity of rationality. For example, it is plausible that the fulfilment of at least some desires or desire-like attitudes has prudential value. This obviously follows from a desire-satisfaction theory of well-being, but it is also plausible given a perfectionist or objective list theory of well-being (Dorsey 2018: 1904). In such cases, acting in accordance with the principle of instrumental rationality will instrumentally bring about or constitutively realise this prudential value. (Because the descriptive and normative uses of 'irrational' sometimes coincide in this way, Ridge [2014: 238] suggests that this provides some explanation of how 'irrational' could semantically shift from a descriptive to a normative meaning.) On the highly plausible assumption that facts about well-being generate normative reasons (see Fletcher 2019), we can explain A's reasons without appealing to the normativity of rationality. Moreover, it seems to me that well-being is a more plausible source of reasons in this case. After all, why should I care about the

principle of rationality itself, independently of how it relates to my interests or ends?

Furthermore, in cases where the satisfaction of some desire is not prudentially good, we can still explain why agents will in general be motivated to act according to the principle of instrumental rationality. First, this might be because when an agent desires something, she typically apprehends that thing as good, which is to say that it *presents* itself as normatively significant.[7] Relatedly, Kolodny (2005: 577ff) proposes that the apparent normativity of rationality can be explained in terms of the "transparency" of the rational ought from the first-person standpoint. For example, consider a case in which someone does not believe the logical consequence [q] of her belief [p]. By calling her irrational, we make a descriptive claim about her psychology violating a certain standard. However, if we make this claim *to* this person, we effectively say the following: "From *your* point of view, you ought to believe [q]". This constitutes advice insofar as it draws the agent's attention to some substantive reason which the agent already takes herself to have for her belief [p]. So from the agent's own perspective, the rational 'ought' is "transparent" because it provides no additional reason to her existing substantive reason to believe [p] and hence also its logical consequence [q].

Thus, quite independently of considerations particular to expressivism, there are a number of reasons for thinking that the rationality involved in attitude attribution is not normative. I therefore submit that denying the normativity of rationality is the best way for interpretative expressivism to avoid the objection. However, if rationality is normative, then interpretative expressivism can also accommodate this insofar as it can adequately apply its theory to itself. As I mentioned at the beginning of this discussion, this is a huge issue that would require a thesis of its own to come to any definite conclusions. But I hope to have at least shown that there are a number of considerations in favour of a descriptive view of rationality. If correct, this shows that interpretative expressivism does not use normative notions to explain normative belief. Therefore, there is no circularity objection, no requirement for the theory to reflexively explain itself, and no requirement to posit normative properties or relations in the explanation of normative thought and discourse.

[7] I understand this claim as a platitude to be accommodated by expressivism rather incompatible with it – compare Blackburn (2002b: 125).

## 5.5 Conclusion

I have argued that expressivists can and should maintain that normative statements express normative beliefs. I argued that interpretationism provides an attractive framework in which to give content to this claim. If successful, the interpretationist view provides a straightforward explanation of logically complex normative thought, mixed thoughts, and inconsistency — all of which expressivists traditionally struggle to explain. Given that I have made use of an off the shelf notion of normative content and propositional attitudes, one might wonder why expressivists have not already considered this approach. A tentative diagnosis is that those engaged in metaethical debates have assumed that any theoretically robust notion of belief must be representational and that any other notion of belief must be deflationary and non-explanatory. I hope to have shown that it is open to expressivists to develop a suitably nonrepresentational theory of belief and other propositional attitudes. However, I have not claimed that interpretationism is the only framework in which nonfactual contents can be put to work in developing an expressivist-friendly theory of belief. So the interpretationist approach must ultimately be assessed in relation to rival approaches as well as on its own terms.

# 6
# Conclusion

In this concluding chapter, I first summarise the main conclusions of the previous chapters and recap my main argument for the theory of normative belief proposed in the previous chapter (section 6.1). Next, in order to provide some illumination through contrast, I critically examine a rival expressivist theory of normative belief and belief content due to Mark Schroeder (section 6.2). Finally, I examine some of the wider implications of the main conclusions of the previous chapters (section 6.3).

## 6.1 Summary of main conclusions

Expressivists have traditionally eschewed normative propositions in their theories of thought and discourse. At the beginning of this thesis, I argued that expressivists can and should develop a theory of normative propositions to inform their theory of normative thought and discourse. By doing so, expressivists can use normative propositions to characterise and explain a number of features of normative thought and discourse that are otherwise difficult to explain. While I think that the possibility of such a view is significantly underappreciated, I have not focused too much on arguments for the general proposal, which in any case is not entirely new (see Schroeder 2013, 2015; Ridge 2014: 124ff, 2018: 10). Instead, I have tried to advance the debate by examining in detail some of the main contenders for an expressivist theory of normative propositions. Rather than building new theories from the ground up, my approach throughout has been to examine existing frameworks for theorising about content and to see what can be achieved by rejecting the representationalist assumptions presupposed by adherents of those frameworks.

The first framework I examined was that of cognitive act theories of propositions. According to cognitive act theories, propositions are types of cognitive acts. Specifically, propositions are those cognitive act types tokened in all cognitive acts with the same content, where 'the same content' is fleshed out non-circularly (e.g. in terms of representation). I took this approach as my starting point for two reasons. First, the basic idea looks particularly amenable

to expressivism because expressivists already accept normative cognitive act types in their theoretical ontology. Second, one of the few existing expressivist theories of normative propositions employs this framework, namely Michael Ridge's ecumenical expressivist theory of normative propositions.

I argued in Chapter 2, however, that Ridge's theory failed to provide any principled account of the identity conditions for normative propositions. I suggested that this resulted from Ridge's failure to respect what I called the unity requirement, which states that expressivists should provide a unified explanation of that in virtue of which our attitudes have the content that they have. This requirement was motivated by reflecting on how Scott Soames' theory of propositions, which Ridge took as a starting point, explained identity conditions in terms of the *purely representational* properties of concrete cognitive acts. Rather than accepting Soames' starting point and trying to extend the theory, I argued that an expressivist cognitive act theory should explain content in terms of a more general nonrepresentational notion that applies to both normative and representational attitudes. In a constructive spirit, I sketched a theory of propositions as acts of conceptual categorisation individuated by the conceptual roles of the constituent concepts. While I did not endorse this account, I highlighted the fact that many expressivists are independently sympathetic to conceptual role approaches to content. Because of this, I suggested that it might provide an attractive picture for many expressivists.

However, we saw that the fate of this picture was ultimately dependent on the availability of a conceptual role account of normative content compatible with expressivism. For this reason, in Chapter 3 I examined conceptual role expressivism more generally. My first task was to provide a general characterisation of the view that unites its otherwise disparate instances. To this end, I provided three core commitments to explaining normative concepts, which taken in conjunction constitute conceptual role expressivism in its most general form. My second task was to raise a general challenge that would apply to any version of this view. This was the challenge of explaining defective concepts. More specifically, the challenge was to explain why normative conceptual roles as conceived by expressivism ground meaningful content, whereas conceptual roles for candidate concepts like HIYO do not, despite also respecting the three core commitments of conceptual role expressivism.

I argued that extant versions of conceptual role expressivism lacked the resources to adequately answer the challenge. I argued that this was because each theory embraced some kind of deflationism about propositional content. Because deflationary notions cannot play an explanatory role, this leaves all the explanatory work to the conceptual role component of the theory of

content. Traditionally, however, conceptual role theories also assign an explanatory role to the contents determined by conceptual roles. This suggested that expressivists attracted to some sort of conceptual role account of content need some substantive conception of normative propositions to play an explanatory role in their overall theory of thought and discourse. However, this could not be provided by the cognitive act theory alone, as the properties of the relevant acts were themselves explained in terms of conceptual roles. I then tentatively suggested that something like Allan Gibbard's notion of planning contents might play this role, leaving a full discussion of this view for the next chapter.

In Chapter 4 I took a broader look at planning contents. First, I located planning contents within a more general view that I called the modal conception of content, which explained content in terms of possibilities. Although this is standardly associated with a possible worlds view of propositions, I argued that the basic idea is more general and encompasses other views of propositions. Moreover, I argued that there is no in principle reason why the modal view cannot explain nonfactual contents in terms of possibilities. Although there are a number of different ways of modelling nonfactual content within this framework, I found it most useful to work with the Gibbardian idea that propositions be modelled as sets of world-hyperplan pairs, where a hyperplan specifics a maximally coherent contingency plan.

I argued that although the world-hyperplan model provides an attractive framework for thinking about normative propositions for expressivists, it faces 'the foundational question' of grounding the model in a satisfactory foundational philosophy of mind. I argued that failing to do this left the world-hyperplan model at best incomplete and at worst non-expressivist. I examined one such attempt to answer the foundational question due to Seth Yalcin but argued that it failed. My diagnosis of this failure was that it postulated a problematic bifurcation between our representational and normative beliefs. What expressivists therefore need, I proposed, is a general theory of belief according to which representational and normative beliefs are fundamentally the same kind of attitude. While I arrived at this conclusion by reflecting on the failure of Yalcin's theory, I think it is exactly what we should expect of an expressivist theory of normative thought that embraces normative propositions within the framework of propositional attitude psychology.

Building on this discussion, in Chapter 5 I tried to answer the foundational question by providing an expressivist cum interpretationist theory of belief and propositional attitudes. According to this view, beliefs are dispositions individuated by principles of constitutive rationality, which provide the basis for interpretation. I argued that in order to interpret agents with normative

attitudes, we require (at least) something like the world-hyperplan model of content. Further, I argued that our normative attitudes are best interpreted by assigning nonfactual rather than factual contents to normative commitments and that this explains why normative commitments are nonrepresentational. However, because normative commitments are governed by the constitutive norms of belief, I argued that the theory explains why normative commitments are fundamentally beliefs and not some other kind of attitude. I then compared the view with versions of expressivism that deny that representational and normative beliefs are fundamentally the same kind of attitude. I argued that the former possessed a number of benefits over the latter. Because of this, I believe that interpretative expressivism has a claim to the title of the best hope for an expressivist theory of belief with normative propositional content.

Although I have argued that the interpretationist approach to belief combined with the modal conception of content provides the best package for an expressivist theory of propositions and normative belief, the aim of this thesis has just as much been to shed light on competing approaches and explore their limitations. To be clear, although I think that the worries for the competing views raised along the way should be taken seriously, I do not doubt that those sympathetic to the frameworks I have rejected will have avenues of response that I have not anticipated here. So long as my criticisms will have furthered our understanding of what questions these theories need to answer and what desiderata they need to meet, then the inquiry will have succeeded.

## 6.2 Schroeder's theory of normative propositions

One of the few philosophers to have explicitly argued that expressivists should embrace normative propositions is Mark Schroeder. The framing and motivations of this thesis owe much to Schroeder's work (see in particular Schroeder 2013, 2015). I discussed some of his arguments for the claim that expressivists should embrace normative propositions in Chapter 1 and we returned to some of these arguments in later chapters. But I haven't said anything about Schroeder's positive proposals for how expressivists should go about embracing normative propositions. Schroeder proposes both a general framework for a nondescriptivist theory of belief and propositions, as well as a theory within that framework. In this section, I examine both of these proposals.

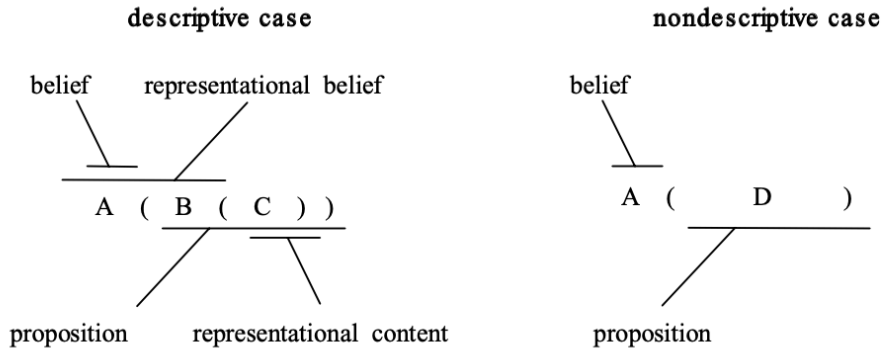
As one of the few existing nondescriptivist theories of propositions designed to be suitable for an expressivist theory of normative thought and discourse, Schroeder's theory is worth examining in its own right. However, it

is also a competitor to the interpretative expressivist view developed in this thesis. It therefore remains to be seen which theory expressivists should prefer. Here, I will argue that the interpretative expressivist view is preferable by raising some worries about Schroeder's theory. Before proceeding, it is worth highlighting that Schroeder's view is meant to provide a framework for nondescriptive propositions more generally, not normative propositions specifically. However, I will focus only on the view as it applies to metaethical expressivism.

### *6.2.1 The general framework*

After outlining a number of problems that expressivist accounts face that arise from rejecting normative propositions, Schroeder (2013: 85) proposes that we distinguish between two theoretical roles that propositions have traditionally been posited to play. The first role is that of being the objects of attitudes and the primary bearers of truth and falsity. The second role is that of carving up the world, where propositions correspond to distinctions in reality. If we assume a representationalist picture of propositional attitudes and truth, it is easy to conflate these two roles, or at least to think that each role is filled by the same class of entities. However, Schroeder argues that nondescriptivists should reject this assumption. Instead, they should assume that there are two distinct classes of entities corresponding to each role. Moreover, not only are these two classes distinct, but they are classes of "two different sorts of thing" (2013: 88). Schroeder calls the entities that play the first role *propositions* and the entities that play the second role *representational contents*.

Given the assumption that propositions are a different sort of thing to representational contents, we are led to ask the following question. If belief is a relation to a proposition, but representational beliefs also carve up the world, what must representational belief be like such that it is also a relation to a representational content? Schroeder's answer is to introduce a structural constraint on representational belief to guarantee that it consists in a relation to a proposition and a distinct relation to a representational content, depending on how you "carve up" the belief state. Diagrammatically, Schroeder (2013: 89) suggests the following picture:




So in the nondescriptive case, belief is simply a relation to a proposition. But in the descriptive case we can carve up belief in two ways. Specifically, it can be carved up as a relation A(_) to a proposition B(C) or it can be carved up as a relation A(B(_)) to a representational content C. (This isn't quite right, because as we saw in Chapter 1, beliefs are relations between *subjects* or *agents* and contents, so we would need to include another variable within the scope of the belief relation to range over subjects. I will omit this, however, for ease of exposition.)

To make the basic idea more concrete, Schroeder (2013: 89) offers a comparison to a more familiar kind of state that structurally resembles belief in the descriptive case. He asks us to consider the state of *being about to go to Paris*. What exactly does this state consist in? Well, on one way of carving up the state, it consists in an agent standing in a relation to an *act type*. Specifically, the relation of *being about to* and the act type of *going to Paris*. But on another way of carving up the state, it consists in an agent standing in a different relation to a *place*. Specifically, the relation of *being about to go to* and the city Paris. So we see that the single state of *being about to go to Paris* consists in two distinct relations to two distinct objects. Moreover, this example seems to exhibit the same abstract structure of the descriptive case in the diagram above. In the present case, 'A(_)' denotes *being about to*, 'B(_)' denotes *going to*, 'A(B(_)' denotes *being about to go to*, and 'C' denotes Paris. So we see that there is nothing inherently problematic or unusual with understanding a state as having this sort of structure.

Granting Schroeder's analysis of the state of being about to go to Paris, what should we make of the analysis as it applies to belief? Although Schroeder (2013: 92) expresses more confidence in his general framework than in his own development of it, I think the framework is pitched at too abstract a level to assess it with much confidence either way. One reason that it is difficult to assess is that before we have said anything about what kind of relation belief

is and what kind of entities propositions are, we haven't done anything to explain or justify the claim that the 'nondescriptive case' is genuinely nondescriptive. As we have seen a number of times throughout this thesis, it is not enough to simply postulate a difference in kind between normative and descriptive propositions to justify a nondescriptive construal of normative thought. We need to actually explain why normative thought is nondescriptive. In the absence of any concrete implementation of the framework, we have no way of evaluating whether the theory is genuinely nondescriptive.

Nevertheless, I think there are some things that can be said against Schroeder's framework at an abstract level. The first thing to note is that although I follow Schroeder in distinguishing the theoretical roles of propositions and representational contents, I have not assumed that each role is filled by a different kind of entity. Whereas Schroeder wants to "avoid the conclusion that propositions are representational contents" (2013: 88), I only want to avoid the conclusion that *all* propositions are representational contents. This is because I see no reason to rule out views according to which *some* propositions are representational contents, as long as some others are not. Or, at least, that some are not *purely* representational contents, if we want to include a theory like Ridge's. This is perfectly compatible with the view that the object-of-attitudes role is distinct from the marking-distinctions-in-reality role. It is just that while the occupiers of the first role are given by the class of propositions, the occupiers of the second role are given by a proper subset of that class. So it's not obvious that respecting the distinction between these two roles requires accepting Schroeder's framework, though it is one way of respecting the distinction.

However, not only is Schroeder's framework not required to respect the distinction, but arguably the alternative is preferable. One reason is that the alternative is simpler and more parsimonious. Whereas Schroeder's framework postulates two distinct relations involved in representational belief and two distinct objects of representational belief, the view that representational contents are a kind of proposition postulates only one object and correspondingly one relation. Moreover, the view is simpler without the loss of explanatory power. We can see this by asking whether the simpler view is any less well placed to explain the four points that Schroeder takes his framework to be able to explain (2013: 89f).

First, assuming the simpler view, we can straightforwardly explain how representational beliefs consist in relations both to propositions and representational contents; this is because representational contents just are propositions on this view. Second, we can straightforwardly explain why it is

easy to overlook the fact that not all beliefs consist in relations to representational contents; this is because the paradigm cases of belief consist in relations to representational contents. Third, we can straightforwardly explain nonrepresentational beliefs; this is because not all propositions are representational contents. Finally, we can straightforwardly explain how the view generalises to other attitudes; this is because we can simply change the relevant relation to the same propositions, whether they are representational or not. So by adopting a simpler framework, we do not lose any explanatory power.

In addition, recall that, according to Schroeder's account, representational propositions necessarily consist in some binary relation between something and a representational content. This is designed to make the necessary distinctions between representational and nonrepresentational beliefs. However, this places a fairly strong constraint on the kind of entity a representational proposition can be. For example, it seems to rule out the possibility of an unstructured view of representational propositions. Perhaps there are good reasons for ruling out such a view, but it shouldn't be ruled out simply by the framework for distinguishing representational and non-representational propositions. Moreover, when the constraint is taken in conjunction with the other constraints already in play for any theory of propositions, this strengthens the constraint considerably. For example, recall that propositions must have the right sort of logical properties, broadly construed, to be able to stand in entailment and inconsistency relations with other propositions. The class of candidates for entities that meet both kinds of constraint is not likely to be very large. At the very least, this class will be considerably smaller than the class of candidates available to nondescriptivists who reject this constraint.

To be clear, I don't take these considerations to be compelling against any particular instance of the framework. But I do think that these considerations give us reason to doubt that an expressivist theory of propositions must employ this framework. Ultimately, however, the framework will have to be assessed on the merits of its implementation. I will therefore now examine Schroeder's implementation of the framework.

### *6.2.2 Biforcated attitudes semantics*

Schroeder develops a nondescriptivist theory of propositions based on his *biforcated attitude semantics*, developed in his book *Being For* (2008). In its original context, biforcated attitude semantics is presented as a theory in natural language semantics to rival truth-conditional and propositional

approaches. However, Schroeder (2013) goes on to use this theory as the basis for his expressivist theory of belief and normative propositions. My focus here will be solely on the latter use.

An attractive feature of Schroeder's view is that he simultaneously provides a nondescriptivist theory of belief and propositions, rather than providing a nondescriptivist theory of propositions and then trying to find a suitable psychological story to ground this theory. It therefore faces no 'foundational question' (see Chapter 4) and no awkward questions about how realists can apparently say all the same things about normative concepts other than that they are also representational (see Chapter 3). In explaining Schroeder's view, it will be helpful to start with his account of belief and then proceed to his account of propositions.

The basic attitude Schroeder postulates to play the role of belief is the attitude of *being for*. We can say that a subject has the attitude of being for when she is *for* something. As a first approximation, what one is for is a certain kind of property. So being for is a relation that holds between subjects and properties. More specifically, when one is for something, then other things being equal, one *does* that thing (2008: 84). So the class of properties that being for relates subjects to is the class of possible *act types*, broadly construed. We can then define *being for* as the state "whose functional role is to lead one to acquire that property [one is for], other things being equal" (2013: 92). (Although Schroeder claims that this definition is not essential to his theory, for the sake of concreteness I will assume this definition throughout.) If believing some proposition consists in being for some property, then at a first pass we can say that propositions are the properties that one is for when one believes that proposition. If the relevant class of properties are act types, then identifying the proposition [p] will involve specifying what an agent is typically motivated to do when she believes [p].

This general account of belief is perfectly suited for an expressivist theory of normative belief. Recall that expressivists think that normative thought is essentially directive or action-guiding. If believing consists in being for, then it turns out that all beliefs are essentially directive or action-guiding. The directive nature of normative belief is then just a particular instance of the directive nature of belief in general. Further, we can suppose that the sense in which beliefs are 'directive' is broad is enough capture an agent's being for having certain attitudes as well as performing certain actions. So if we suppose, as the expressivist might well do, that normative expressions like 'wrong' express noncognitive attitudes like *disapproval*, we can incorporate this idea into Schroeder's theory by analysing the belief [that murder is wrong] as consisting in *being for disapproving of murder*. Thus if believing just is being for,

then the proposition [that murder is wrong] just is the property of disapproving of murder. In terms of the diagram, we can now say that the relation A is that of being for, and the proposition D is the act type that one is for when one believes D.

With the nondescriptive case explained, we now need to explain the descriptive case. If believing is being for, and being for involves being motivated to do something, broadly construed, what kind of things are we motivated to do when we believe representational propositions? That is, if propositions are akin to "sets of instructions about what to do" (Schroeder 2015: 17), what kind of instructions are encoded by representational propositions? Given the constraints of the general framework, they must be instructions that relate to representational contents. Schroeder suggests that we understand these instructions in terms of the state of *proceeding as if*, which relates agents to representational contents. In terms of the diagram, this relation is denoted by 'B'. Believing a representational proposition therefore consists in *being for proceeding as if p*, where '*p*' denotes a representational content. Thus, for instance, believing [that the cat is on the mat] consists in *being for proceeding as if the cat is on the mat*, where '*the cat is on the mat*' denotes a representational content. It will not matter too much what kind of entity we take representational contents to be, but Schroeder (2013: 94n) suggests that we can think of representational contents in terms of orthodox conceptions of propositions, for example sets of possible worlds or Russellian propositions.

What is it, then, to proceed as if *p* and to be for proceeding as if *p*? Schroeder provides the following answer: "to proceed as if *p* is to take *p* as settled in deciding what to do. Assuming that being for has the motivational property that someone who is for α will tend to do α, other things being equal, it follows that someone who believes that *p* will tend to proceed as if *p*, other things being equal. That is, she will tend to treat *p* as settled in deciding what to do." (Schroeder 2008: 93f; see also Schroeder 2013: 94) The basic idea is that this notion captures Ramsey's metaphor that (representational) belief is the map by which we steer our way through the world. And at a descriptive level, Schroeder's suggestion does seem plausible. When I believe [that my desk is wooden], other things being equal I do proceed as if my desk is wooden, in that I take this fact as settled in deciding to place my coffee cup on my desk, to keep it away from fire, and so on. So if believing just is being for, then the proposition [that my desk is wooden] is the property of proceeding as if *my desk is wooden*. This illustrates the basic idea of Schroeder's theory.

(However, this is in fact an oversimplification. For in order to adequately explain negated beliefs, we need the objects of beliefs not simply to be

properties, but to be *pairs* of properties. To see why we, consider the representational belief [p]. Given our initial analysis, this belief consists in:

(1) being for pai *p*.

As Schroeder (2008: 64) notes, the most natural interpretation of truth-functional logical connectives on this picture is to consider them as operating on the properties that are the contents of beliefs. Although this seems to predict the right results in the nondescriptive case, it raises a problem in the descriptive case. On the natural interpretation of 'not', our simple analysis predicts that the belief [not-p] consists in:

(2) being for *not* pai *p*.

This is because the if negation operates on the object of belief, then it will negate the property *pai p*, which gives us *not pai p*. However, this seems to give us the wrong prediction, because intuitively we should expect the belief [not-p] to consist in:

(3) being for pai *not*-p.

Because (2) and (3) are not equivalent, the theory seems to over-generate belief states (Schroeder 2008: 96). To solve this problem, Schroeder proposes that we understand belief as relating agents to *pairs* of properties, where one property entails the other. Thus, the belief [p] consists not in (1) but in:

(4) <being for(pai *p*), being for(¬pai ¬*p*)>

Truth-functional logical connectives are then interpreted as operating on the pair of properties. The belief [not-p] is then analysed as consisting in:

(5) <being for(¬¬pai ¬*p*), being for(¬pai *p*)>
= <being for(pai¬*p*), being for(¬pai *p*)>

(Schroeder orders the pairs such that the strongest property always comes first.) While this might look somewhat more complex, it is not obviously significantly stronger than the initial analysis. This is because it is plausible that it is inconsistent to proceed as if *p* and to proceed as if not-*p*. So on the simple analysis, being for proceeding as if *p* already *commits* one to being for not proceeding as if not-*p* (Schroeder 2008: 98). In any case, although this is an

important feature of the theory, it will not be too important for the ensuing discussion, so I will stick to the simpler analysis for ease of exposition.)

Schroeder's theory of belief and propositions has a number of virtues. First, it provides a principled account of the distinction between representational and nonrepresentational propositions within a more general theory of belief and propositions. It therefore answers the foundational question and respects the unity requirement. Second, it is highly plausible that it is inconsistent to be for inconsistent properties. The theory therefore well placed to explain why it is inconsistent to believe inconsistent propositions. Third, given the functional characterisation of being for, we can explain inconsistency in belief in terms of its necessarily frustrating its constitutive function. Moreover, given the account of representational belief, we can still explain how inconsistency in representational beliefs is also a kind of representational failure. In the previous chapters, we examined a number of different approaches that struggled to meet one or more of these desiderata. So Schroeder's theory is in many ways attractive. However, in the next subsection, I will raise some worries about the theory.

### *6.2.3 Some objections to Schroeder's theory*

My main objection to Schroeder's theory concerns the notion of *proceeding as if*. This notion plays a vital role for Schroeder in distinguishing those beliefs which are representational from those which are not. However, it's unclear to me that proceeding as if can in fact play this role. In its most basic form, the problem is that the act or activity of proceeding as if seems to be just as present in the nondescriptive case as in the descriptive case. If I believe [that my desk is wooden], then a plausible description of my state is that I am for proceeding as if my desk is wooden. However, it is no less plausible to describe my belief [that murder is wrong] in terms of my being for proceeding as if murder is wrong. After all, other things being equal, if I have this belief, then I will take the wrongness of murder as settled in deciding what to do. So intuitively the notion of proceeding as if fails to distinguish between representational and nonrepresentational beliefs.

One response to this objection would be to point out that 'proceeding as if' is defined as a relation between (a) states of being for and (b) representational contents. If we reject the existence of normative representational contents, then there is nothing that can fill the second argument place of proceeding as if in the normative case. So because 'murder is wrong' expresses only a proposition and no representational content, and because by definition proceeding as if is a relation between states of being for and representational contents, and not

between states of being for and propositions, there is no possible state of being for proceeding as if murder is wrong.

However, although it is possible to define 'proceeding as if' in this way, the stipulation that it applies only to representational contents is not grounded in the functional characterisation of the attitude. If we take 'proceeding as if' as "a shorthand for the general relation of taking something as settled in one's deliberative activity" (Schroeder 2013: 94), then it is clear that the notion applies to propositions, even nonrepresentational propositions, because it is clear that I can take the proposition [that murder is wrong] as settled in my deliberative activity. Perhaps one could amend the functional characterisation to range over representational contents explicitly, so that 'proceeding as if' is shorthand for the relation of taking *representational contents* as settled in one's deliberative activity. However, it is unclear that we have any grasp of this functional role over and above that of the more general functional role. Moreover, to echo a criticism of Schroeder's in a different context, explicitly ranging over representational contents in this way looks more like a criterion that the functional role of proceeding as if must meet rather than an explanation of what this functional role is actually like such that it only ranges over representational contents. So we have not yet been given any psychological distinction to go along with our distinction between <pai *p*> and <pai [that p]>.

Indeed, insofar as 'proceeding as if' has its home in deliberative activity, then even in the descriptive case it seems more plausible within Schroeder's framework that proceeding as if is a relation to propositions and not representational contents. For deliberative activity will constitutively involve propositional acts and attitudes, and so the most natural way of characterising this activity will appeal to the propositional contents of those acts and attitudes, including in the descriptive case. Moreover, it would be strange if the same activity of proceeding as if took a different object in the descriptive case, even though the same kind of object (propositions) is also available.

However, this might suggest another response to the objection, and one that Schroeder seems to endorse in a slightly different context. While Schroeder explains representational belief in terms of proceeding as if, he is also aware that it is perfectly acceptable to describe normative beliefs in terms of proceeding as if (2008: 155). If this is right, then how should we explain the difference between the descriptive and nondescriptive case? Schroeder suggests that we explain the difference in terms of what proceeding as if *consists in* in each case. Thus, an expressivist might claim that while it is correct to describe the belief [that murder is wrong] as being for proceeding as if murder is wrong, proceeding as if murder is wrong *just is* disapproving of

murder. By comparison, while it is also correct to describe the belief [that my desk is wooden] as being for proceeding as if my desk is wooden, proceeding as if my desk is wooden *just is* proceeding as if a certain representational content is true.[1]

However, this response faces the same objection as before. The problem is that we do not have any account of what it is to proceed as if some representational content is true. For again, insofar as we understand what it is to proceed as if, it applies to propositions and not representational contents. Further, if we assumed that the very same relation applied to both propositions and representational contents, then arguably this would over-generate beliefs. This is because for every representational state picked out by the expression 'being for proceeding as if p', 'p' would be ambiguous between denoting a proposition and a representational content. Given Schroeder's assumption that these are two distinct kinds of entity, it follows that there are two states of belief where intuitively it seems that there is only one. This is because beliefs are individuated by their contents, and so if descriptive beliefs consist in being for proceeding as if p, we can differentiate between two states of being for depending on how we disambiguate 'p'.

I haven't said anything here that rules out the possibility of providing a characterisation of 'proceeding as if' that applies only to representational contents. However, for the reasons set out above, this relation is *not* "the general relation of taking something as settled in one's deliberative activity." Instead, it will be some distinct relation, call it *pai**. Perhaps one could then say that in the descriptive case, the property <pai [that p]> just is <*pai* p*>. However, this cannot be right for the simple reason that by hypothesis, propositions and representational contents are distinct kinds of entity. Therefore it is not possible that [p] = *p*. Because relations are partly individuated by their relata, it is therefore not possible that <pai [p]> just is <*pai* p*>.

Another objection concerns the plausibility of generalising Schroeder's theory to other propositional attitudes. Schroeder (2013: 90) rightly claims that any implementation of his general framework will have the right sort of structure to explain other propositional attitudes. In terms of the above diagram, it is simply a matter of replacing the belief relation A with other relevant relations, such as the desire relation. Importantly, because A is the same in both the descriptive and non-descriptive case, this allows us to have a

[1] Schroeder's discussion of this issue occurs prior to his endorsement of a separation of the roles of propositions and representational contents, so this isn't exactly what he says. But I take it that this is what the original suggestion would amount to within the new framework. His explanation of what makes proceeding as if p descriptive is if "it is possible to understand what it would be for it to be the case that P" (2008: 156).

uniform treatment of attitudes across both cases. However, while this is a virtue of the general framework, it is not obvious how to implement it within the biforcated attitude theory.

Consider, for instance, an ordinary desire, such as the desire to go for a walk. If desires have propositions as their objects, we might suppose that the relevant content of this desire is the proposition [that I am walking]. Given that this is a descriptive proposition, according to Schroeder's theory this proposition will be the property of proceeding as if I am walking. So what relation do we bear to this property in desiring to go for a walk? Note that standard analyses of desires will not apply here. For example, if propositions are centred worlds, it is natural to understand the functional role of desire in terms of an agent being motivated to acquire the property that is the object of the desire. But if we accept Schroeder's theory of propositions, this relation more or less just is the belief relation. So it's not entirely clear to me what relation could play the desire role on Schroeder's picture given his account of propositions.

Further, it's worth stressing the dialectical burden that this places on Schroeder's theory. By providing a unified framework for descriptive and nondescriptive belief, Schroeder's theory avoids having to provide unacceptably ad hoc analyses of nondescriptive desires, hopes, doubts, and so on. However, if believing is being for and propositions are sets of instructions about what to do, then it's hard to see how to generalise the theory to other propositional attitudes. It seems that we are going to require a new analysis of each attitude kind to explain how that attitude consists in standing in a relation to a set of instructions about what to do, even in the descriptive case. By comparison, the interpretationist view set out in the previous chapter retains a fairly conventional view about the desire relation, so long as we can make sense of the notion of *bringing it about that some proposition is true* non-reductively, so that it can apply to nonrepresentational propositions as well as representational propositions.

To sum up, although Schroeder's theory has a number of virtues, I think that the above objections give us sufficient reason to doubt its success. Although I haven't said anything to rule out this theory entirely, I hope to have shown that as it currently stands, it does not succeed in its primary task, which is to explain the difference between representational and nonrepresentational belief. I therefore conclude that the theory is not preferable to the interpretationist view developed in the previous chapter.

## 6.3 Theoretical implications and directions for future study

In the final section of this chapter, I want to discuss some of the wider theoretical implications of the conclusions reached in this thesis, together with further avenues of research. Perhaps most obvious is the implications for debates about expressivism within metaethics. Expressivism has often been characterised as the denial that there are such things as normative beliefs together with the denial that there are such things as normative propositions (except in a deflationary sense). Opponents of expressivism often take issue with these denials and argue that expressivism has serious shortcomings because of them. Clearly, however, if the interpretative expressivist view sketched in the previous chapter is correct, then expressivism has no such shortcomings. Or at least, if it does, it can no longer be explained simply in terms of its rejection of normative belief and propositions. As I have highlighted, it would be fair to point out that the modal conception of content as well as interpretationism have problems of their own, and these will no doubt be inherited by the interpretative expressivist view of normative belief. However, this would be true of any theory of belief and any theory of propositions that expressivists might utilise, so the charge is dialectically ineffective — though of course these problems will have to be addressed at the end of the day.

Another implication is that it upsets many of the standard taxonomies so dear to metaethicists (see, for example, Miller 2013: 8; Kalderon 2007: 140; Parfit 2017: 56). For it is no longer open to define the opposition between expressivism and descriptivism as a difference between whether normative statements express normative propositions or beliefs. Moreover, the discussion of this thesis shows that the taxonomies employed within metaethics are not as theoretically neutral as they are often assumed to be. This is because they are not free of presuppositions regarding the nature of belief, propositions, and so on. Although such taxonomies can be a pedagogically helpful heuristic, I think this shows that it is a mistake to insist that all metaethical theories fit neatly within a commonly shared taxonomy. While at the end of the day any theory will have to be compared with any other, there is no reason to insist that everyone comes to the table with the same starting assumptions. Moreover, there is no reason why we should not view the background assumptions as part of the overall package to be assessed. To do otherwise obscures the theoretical possibilities available within metaethics.

Looking beyond metaethics, the discussion has wider implications for the philosophy of mind and language as well as for metaphysics. For instance, if

something like interpretative expressivism is along the right lines, this would undermine a commonly accepted picture of the nature of meaning and intentionality. Specifically, it would undermine the picture of meaning and intentionality as essentially representational. Further, if certain domains of discourse are nonrepresentational, this raises the further question about whether providing a metaphysics for that domain as traditionally conceived is a legitimate enterprise. This is because given a nonrepresentational construal of some domain of discourse, there is no feature of reality that that discourse is about.[2] Looking forward, this naturally raises questions about whether the expressivist view of normativity can be generalised to other domains that lend themselves to an expressivist analysis. I have already noted that the modal view of content has also been employed to provide nonfactualist characterisations of mathematical thought, know how, and epistemic modality. So I think there is some license for optimism that the view of belief will generalise to other domains, though ultimately this remains to be seen.

While I have argued that the interpretative expressivist view is the most promising approach for an expressivist theory of belief, I have not fully defended this view. If one were to fully defend the view, one would need to do two things. First, one would need to provide a defence of interpretationism. Second, one would need to provide a defence of the particular principles of rationality that explained why normative beliefs have nonfactual propositions as their contents. Although these tasks would be crucial to establishing the interpretative expressivist view, each task would deserve a thesis of its own, and so I have not made any claim to have conclusively established these positions. However, for what it is worth, I do believe that interpretationism provides one of the most promising approaches for explaining propositional attitudes and folk psychology. And the kind of arguments I appealed to in motivating the relevant principles of constitutive rationality are just the kind of arguments expressivists employ anyway, so I think expressivists should feel comfortable in their ability to defend these principles.

I also mentioned at the end of Chapter 5 that while I take interpretationism to be independently plausible and particularly amenable to expressivism, I have said nothing to rule out other foundational theories of belief as providing suitable frameworks for expressivists. Perhaps another plausible candidate would be that of analytic or 'common-sense' functionalism (Lewis 1999; Braddon-Mitchell and Jackson 2007). Analytic functionalism is in many ways similar to the interpretationist view I have adopted. Both claim that the mental is constitutively rational and both accept a holistic approach to belief

[2] Though see Dreier (2015c) for an argument that quasi-realist expressivism does not successfully deflate all metaphysical questions about normativity.

attribution. The analytic functionalist makes the further claim that we can (approximately) reduce mental properties to physical properties (i.e. brain states) by providing a Ramsey sentence of the folk psychological characterisation of an agent's attitudes. Another possibility would be to develop a dispositionalist account that does not rely on principles of rationality to individuate beliefs. For example, Schwitzgebel (2002) suggests that a belief's dispositional profile might include phenomenal dispositions. Perhaps normative beliefs can then be characterised by some kind of affective or sentimental disposition. A further possibility would be to adopt a psychofunctionalist view that embraces the representational theory of mind. The expressivist would then need to provide a suitably nondescriptivist psychosemantics for normative language (Horwich [2010] can be read way; Gibbard [2012] arguably provides a different kind of sententialist theory of belief). No doubt there are other possibilities, all of which I welcome.

However, I think that interpretative expressivist view provides a promising, novel, and interesting framework for thinking about normative belief and belief in general. It provides new resources for bringing to fruition some of the suggestive but sketchy ideas of the new proponents of the Gibbardian approach to content. Moreover, as well as furthering the expressivist research programme, it also provides the grounds for a reimagining of interpretationism in a number of ways. One way in which it does so is by providing a new way of thinking about pragmatist views that take all belief-talk to be nonrepresentational. Moreover, by emphasising the role that propositions play within interpretationism, the inquiry suggests a tighter integration of interpretationism with the modal view of content. More generally, it calls for a greater appreciation of the explanatory role of propositions within interpretationism, including the variety of views that might follow depending on one's choice of propositions within the theory. So as well as being an attractive view in and of itself, it suggests a number of further avenues of research both within metaethics and the philosophy of mind.

# References


Adams, Robert M. (1974). 'Theories of Actuality'. *Noûs*, 8(3): 211-231.

Anscombe, G. E. M. (1957). *Intention.* Oxford: Blackwell.

Ayer, A. J. (1936). *Language, Truth, and Logic*. London: V. Gollancz.

Baker, Derek Clayton and Woods, Jack (2015). 'How Expressivists Can and Should Explain Inconsistency'. *Ethics*, 125(2): 391-424.

Barnes, W. H. F. (1934). 'A Suggestion about Value'. *Analysis,* 1(3): 45-46.

Båve, Arvid (2013). 'Compositional Semantics for Expressivists'. *Philosophical Quarterly*, 63(253): 633-659.

Belnap, Nuel (1962). 'Tonk, Plonk and Plink'. *Analysis*, 22(6): 130-134.

Björnsson, Gunnar (2017). 'The Significance of Ethical Disagreement for Theories of Ethical Thought and Talk'. In McPherson, T. and Plunkett, D. (eds.) *Routledge Handbook of Metaethics*. London: Taylor and Francis.

Blackburn, Simon (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon.

Blackburn, Simon (1988) 'Attitudes and Contents'. Reprinted in Blackburn (1993).

Blackburn, Simon (1993). *Essays in Quasi-Realism*. New York: Oxford University Press.

Blackburn, Simon (1998). *Ruling Passions: A Theory of Practical Reason*. Oxford: Clarendon Press.

Blackburn, Simon (2002a). 'Replies'. *Philosophy and Phenomenological Research*, 65(1): 164–176.

Blackburn, Simon (2002b). 'Precis of *Ruling Passions*'. *Philosophy and Phenomenological Research*, 65(1): 122-135.

Blackburn, Simon (2006). 'The Semantics of Non-factualism, Non-cognitivism, and Quasi-realism'. In Devitt, M. and Hanley, R. (eds.) *The Blackwell Guide to the Philosophy of Language* (Malden, Mass.; Oxford: Blackwell).

Block, Ned (1986). 'Advertisement for a Semantics for Psychology'. *Midwest Studies in Philosophy*, 10(1): 615-678.

Block, Ned (1998). Conceptual Role Semantics. In Craig, E. (ed.) *Routledge Encyclopedia of Philosophy*. London; New York: Routledge.

Braddon-Mitchell, David and Jackson, Frank (2007). *Philosophy of Mind and Cognition* (2nd edition). Malden, Mass.: Blackwell.

Brandom, Robert (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.

Bratman, Michael (1987). *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.

Carnap, Rudolf (1935) *Philosophy and Logical Syntax*, London: Routledge.

Carnap, Rudolf (1947). *Meaning and Necessity*. Chicago: University of Chicago Press.

Carruthers, Peter (1989). *Tractarian Semantics: Finding Sense in Wittgenstein's Tractatus*. Oxford: Basil Blackwell.

Castañeda, Hector-Neri (1975). *Thinking and Doing: The Philosophical Foundations of Institutions*. Dordrecht, Holland.

Charlow, Nate (2014). 'The Problem With the Frege-Geach Problem'. *Philosophical Studies*, 167(3): 635-665.

Charlow, Nate (2015). 'Prospects for an Expressivist Theory of Meaning'. *Philosophers Imprint*, 15: 1-43

Chrisman, Matthew (2007). 'A Dilemma for Moral Fictionalism'. *Philosophical Books*, 49(1): 4-13.

Chrisman, Matthew (2012). 'Epistemic Expressivism'. *Philosophy Compass,* 7(2): 118-126.

Chrisman, Matthew (2016). *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics*. New York: Oxford University Press.

Chrisman, Matthew (2017). 'Conceptual Role Accounts of Meaning in Metaethics'. In McPherson, T. and Plunkett, D. (eds.) *The Routledge Handbook of Metaethics*. London: Taylor and Francis.

Child, William (1996). *Causality, Interpretation and the Mind*. Oxford: Clarendon.

Churchland, Paul M. (1981). 'Eliminative Materialism and the Propositional Attitudes'. *The Journal of Philosophy*, 78(2): 67-90.

Crane, Tim (1990). 'An Alleged Analogy Between Numbers and Propositions'. *Analysis*, 50 (4): 224-230.

Darwall, Stephen, Gibbard, Allan and Railton, Peter (1992). 'Toward Fin de Siècle Ethics: Some Trends'. *Philosophical Review*, 101(1): 115-189.

Davidson, Donald (1975). 'Thought and Talk'. In Guttenplan, S. D. (ed.) *Mind and Language*. Oxford: Clarendon Press.

Davidson, Donald (1980). *Essays on Actions and Events*. Oxford: Clarendon.

Davis, Wayne A. (2003). *Meaning, Expression, and Thought*. Cambridge: Cambridge University Press.

Davis, Wayne A. (2019). 'Cognitive Propositions and Semantic Values. *Inquiry* (published online).

Dennett, Daniel C. (1969). *Content and Consciousness*. London: Routledge & Kegan Paul.

Dennett, Daniel C. (1971). 'Intentional Systems'. *Journal of Philosophy*, 68: 87-106.

Dennett, Daniel C. (1987a). *The Intentional Stance*. Cambridge, Mass.: MIT Press.

Dennett, Daniel C. (1987b). 'True Believers'. In Dennett (1987a).

Dennett, Daniel C. (1989c). 'Three Kinds of Intentional Psychology'. In Dennett (1987a).

Dennett, Daniel C. (1989d). 'Making Sense of Ourselves.' In Dennett (1987a).

Dennett, Daniel C. (1989e). 'Beyond Belief'. In Dennett (1987a).

Dennett, Daniel C. (1991). 'Real Patterns'. *Journal of Philosophy*, 88(1): 27-51.

Devitt, Michael (2010). *Putting Metaphysics First: Essays on Metaphysics and Epistemology*. Oxford: Oxford University Press.

Divers, John and Miller, Alexander (1995). 'Platitudes and Attitudes: A Minimalist Conception of Belief'. *Analysis*, 55(1): 37 - 44.

Dorsey, Dale (2018). 'Prudence and Past Selves'. *Philosophical Studies*, 175(8): 1901-1925.

Dreier, James (1996). 'Expressivist Embeddings and Minimalist Truth'. *Philosophical Studies*, 8(1): 29-51.

Dreier, James (2002). 'The Expressivist Circle: Invoking Norms in the Explanation of Normative Judgment'. *Philosophy and Phenomenological Research*, 65(1): 136–143.

Dreier, James (2015a). 'Can Reasons Fundamentalism Answer the Normative Question?' In Björnsson, G., Strandberg, C., Olinder, R., Eriksson, J., and Björklund, F. (eds.) *Motivational internalism*. New York: Oxford University Press.

Dreier, James (2015b). 'Another World'. In Johnson, R. and Smith, M. (eds.) *Passions and Projections: Themes from the Philosophy of Simon Blackburn*. New York: Oxford University Press.

Dreier, James (2015c). 'Explaining the Quasi-Real'. In Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics, Volume 10*. New York: Oxford University Press.

Dummett, Michael (1973). *Frege: Philosophy of language*. London: Duckworth.

Dummett, M. (1991). 'Frege's Myth of the Third Realm'. In *Frege and Other Philosophers*. Oxford: Clarendon.

Egan, Francis (2014). How to Think About Mental Content. *Philosophical Studies,* 170(1): 115-135.

Eklund, Matti (2009). 'The Frege–Geach Problem and Kalderon's Moral Fictionalism'. *Philosophical Quarterly*, 59(237): 705-712.
Field, Hartry (1977). 'Logic, Meaning, and Conceptual Role'. *Journal of Philosophy*, 74(7): 378-409.
Field, Hartry (1978). 'Mental Representation'. *Erkenntnis*, 13: 9-61.
Field, Hartry (2009). 'Epistemology Without Metaphysics'. *Philosophical Studies*, 143(2): 249-290.
Field, Hartry (2018). 'Epistemology from an Evaluativist Perspective'. *Philosophers' Imprint* 18: 1-23.
Fletcher, Guy (2019). ''Taking Prudence Seriously'. In Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics, Volume 14*. New York: Oxford University Press.
Fodor, Jerry A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: The MIT Press.
Geach, P. T. (1965). 'Assertion'. *Philosophical Review*, 74(4): 449-465.
Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, Mass.: Harvard University Press.
Gibbard, Allan (2003). *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.
Gibbard, Allan (2012). *Meaning and Normativity*. Oxford: Oxford University Press.
Godfrey-Smith, Peter (2005). 'Folk Psychology as a Model'. *Philosopher's Imprint* 5: 1-16.
Hanks, Peter (2015). *Propositional Content*. Oxford: Oxford University Press.
Harman, Gilbert (1973). *Thought*. Princeton, N.J.; Princeton University Press.
Harman, Gilbert (1995). 'Rationality'. In Smith, E. E. and Osherson, D. N. (eds.) *Invitation to Cognitive Science*. Cambridge Mass.: MIT Press.
Hare, R. M. (1952). *The Language of Morals.* Oxford: Clarendon.
Horgan, Terence and Timmons, Mark (1992). 'Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived'. *Philosophical Papers,* 21(3): 153-175.
Horgan, Terence and Timmons, Mark (2006). 'Cognitivist Expressivism'. In Horgan, T., and Timmons, M. (eds.) *Metaethics after Moore.* Oxford: Clarendon.
Hornsby, Jennifer (2000). 'Personal and Sub-Personal; A Defence of Dennett's Early Distinction'. *Philosophical Explorations*, 3(1): 6-24.
Hornsby, Jennifer (2013). 'Basic Activity'. *Aristotelian Society Supplementary Volume*, 87(1): 1-18.
Horwich, Paul (1993). 'Gibbard's Theory of Norms'. *Philosophy and Public Affairs*, 22(1): 67-78.
Horwich, Paul (1998). *Meaning*. Oxford: Clarendon.
Horwich, Paul (2010). 'The Motive Power of Evaluative Concepts'. In *Truth-Meaning-Reality*. Oxford: Clarendon.
Joyce, Richard (2012). 'Review of Kalderon's *Moral Fictionalism*'. *Philosophy and Phenomenological Research*, 85(1): 161-173.
Kalderon, Mark E. (2005). *Moral Fictionalism*. Oxford: Clarendon.
Kenny, Anthony (1963). *Action, Emotion and Will*. London: Routledge & K. Paul.
Kiesewetter, Benjamin. (2017). *The Normativity of Rationality*. New York: Oxford University Press.
Köhler, Sebastian (2017). 'Expressivism, Belief, and All That'. *Journal of Philosophy*, 114(4): 189-207.
Köhler, Sebastian (2018). 'Expressivism, Meaning, and All That'. *Canadian Journal of Philosophy*, 48(3-4): 337-356.
Kolodny, Niko (2005). 'Why Be Rational?'. *Mind*, 114(455): 509-563.
Kratzer, Angelika (1977). 'What 'Must' and 'Can' Must and Can Mean'. *Linguistics and Philosophy*, 1(3): 337--355.
Lewis, David K. (1974). 'Radical Interpretation'. *Synthese*, 27:331-344
Lewis, David K. (1979a). 'Attitudes De Dicto and De Se'. *Philosophical Review*, 88(4): 513-543.

Lewis, David K. (1979b). 'Scorekeeping in a Language Game'. *Journal of Philosophical Logic*, 8(1): 339-359.

Lewis, David K. (1980) 'Index, Context, and Content'. In S. Kanger and S. Ohman (eds.) *Philosophy and Grammar*. Dordrecht, Holland: D. Reidel.

Lewis, David K. (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.

Lewis, David K. (1999). 'Reduction of Mind'. In *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Liao, Shen-yi (2012). 'What Are Centered Worlds?' *Philosophical Quarterly*, 62(247): 294-316.

Loar, Brian (1981). *Mind and Meaning*. Cambridge: Cambridge University Press.

MacFarlane, John (2014). *Assessment Sensitivity: Relative Truth and its Applications*. New York: Oxford University Press.

Matthews, Robert (2007). *The Measure of Mind: Propositional Attitudes and Their Attribution*. Oxford: Clarendon.

McDowell, John (1979). 'Virtue and Reason'. *The Monist*, 62(3): 331-350.

McGrath, Matthew and Frank, Devin. (2018) 'Propositions'. In Edward N. Zalta (ed.), The Stanford Encyclopedia of Philosophy (Spring 2018 edition). https://plato.stanford.edu/entries/propositions/.

Miller, Alexander (2003). *An Introduction to Contemporary Metaethics* (2nd edition). Cambridge; Malden, Mass.: Polity Press.

Moore, Joseph G. (1999). 'Propositions Without Identity'. *Noûs*, 33(1): 1-29.

Morgan, Mary S., and Morrison, Margaret (1999). *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Mourelatos, Alexander P. D. (1978). 'Events, Processes, and States'. *Linguistics and Philosophy*, 2(3): 415-434.

Ninan, Dilip (2010). 'Semantics and the Objects of Assertion'. *Linguistics and Philosophy*, 33(5): 355-380.

Ogden, C. K., and Richards, I. A. (1923). *The Meaning of Meaning : A Study of the Influence of Language Upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul.

Parfit, Derek (2017). *On What Matters: Volume Three*. New York: Oxford University Press.

Peacocke, Christopher (1992). *A Study of Concepts*. Cambridge, Mass.: The MIT Press.

Peacocke, Christopher (1996). 'Concepts Without Words'. In Heck, R. G. (ed.) *Language, Thought, and Logic: Essays in Honour of Michael Dummett*. New York: Oxford University Press.

Pérez Carballo, Alejandro (2016). 'Structuring Logical Space'. *Philosophy and Phenomenological Research*, 92(2): 460-491.

Perry, John (1977). 'Frege on Demonstratives'. *Philosophical Review*, 86(4): 474-497.

Porot, Nicolas, & Mandelbaum, Eric (forthcoming). 'Empirical Models of Belief: A Progress Report'. Wires Cognitive Science.

Portner, Paul (2007). 'Imperatives and Modals'. *Natural Language Semantics*, 15(4): 351-383.

Povinelli, Daniel J. & Vonk, Jennifer (2004). 'We Don't Need a Microscope to Explore the Chimpanzee's Mind'. *Mind and Language* 19(1): 1-28.

Price, Huw (1994). 'Semantic Deflationism and the Frege Point.' In Tsohatzidis, S. L. (ed.), *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*. Routledge.

Price, Huw (2011). Expressivism for Two Voices'. In Knowles, J. and Rydenfelt, H. (eds.) *Pragmatism, Science and Naturalism*, 87-113. Bern, Switzerland: Peter Lang D.

Price, Huw (2013). *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.

Prior, A. N. (1960). 'The Runabout Inference-Ticket'. *Analysis*, 21(2): 38.

Putnam, Hilary (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

Rabern, Brian (2012). 'Against the Identification of Assertoric Content with Compositional Value'. *Synthese*, 189(1): 75-96.
Ridge, Michael (2009). 'Moral Assertion for Expressivists'. *Philosophical Issues*, 19(1): 182-204.
Ridge, Michael (2014). *Impassioned Belief*. Oxford: Oxford University Press.
Ridge, Michael (2015). 'Internalism'. In Björnsson, G., Strandberg, C., Olinder, R., Eriksson, J., and Björklund, F. (eds.) *Motivational internalism*. New York: Oxford University Press.
Ridge, Michael (2018). Normative Certitude for Expressivists. *Synthese* (published online).
Rosen, Gideon. (1998). 'Blackburn's *Essays in Quasi-Realism*'. *Noûs*, 32(3): 386-405.
Santorio, Paolo (2016). 'Nonfactual Know-How and the Boundaries of Semantics'. *Philosophical Review*, 125(1): 35-82.
Schroeder, Mark (2008). *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Clarendon.
Schroeder, Mark (2010). *Noncognitivism in Ethics*. London: Routledge.
Schroeder, Mark (2013). 'Two Roles for Propositions: Cause for Divorce?' reprinted in Schroeder (2015).
Schroeder, Mark (2015). *Expressing Our Attitudes: Explanation and Expression in Ethics* Volume 2. Oxford: Oxford University Press.
Schroeter, Laura and Schroeter, Francois (2003). 'A Slim Semantics for Thin Moral Terms?' *Australasian Journal of Philosophy*, 81(2): 191–207.
Schiffer, Stephen (2003). *The Things We Mean*. New York: Oxford University Press.
Schwitzgebel, Eric (2002). 'A Phenomenal, Dispositional Account of Belief'. *Noûs*, 36(2): 249-275
Shiller, Derek (2016). 'A Primitive Solution to the Negation Problem'. *Ethical Theory and Moral Practice* 19(3): 725-740.
Shiller, Derek (2018). 'The Unity of Moral Attitudes: Recipe Semantics and Credal Exaptation'. *Canadian Journal of Philosophy*, 48(3-4): 425-446.
Silk, Alex (2013). 'Truth Conditions and the Meanings of Ethical Terms'. In Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics, Volume 8*. Oxford: Oxford University Press.
Silk, Alex (2015). 'How to Be an Ethical Expressivist'. *Philosophy and Phenomenological Research*, 91(1): 47-81.
Sinclair, Neil (2018) 'Conceptual Role Semantics and the Reference of Moral Concepts'. *European Journal of Philosophy*, 26(1): 95-121.
Sinnott-Armstrong, Walter (1993). 'Some Problems for Gibbard's Norm-expressivism'. *Philosophical Studies*, 69(2-3): 297-313.
Smith, Michael (1994). *The Moral Problem*. Oxford: Blackwell.
Soames, Scott (2010). *What Is Meaning?* Princeton, N.J.: Princeton University Press.
Soames, Scott (2014). 'Cognitive Propositions'. In King, J., Soames, S., and Speaks, J. (eds.) *New thinking about propositions*. Oxford: Oxford University Press.
Soames, Scott (2015). *Rethinking Language, Mind, and Meaning*. Princeton, N.J.; Oxford: Princeton University Press.
Stalnaker, Robert (1987). *Inquiry.* Cambridge, Mass.: MIT Press.
Stalnaker, Robert (2004). 'Lewis on Intentionality'. *Australasian Journal of Philosophy*, 82(1): 199-212.
Stalnaker, Robert (2012). *Mere Possibilities: Metaphysical Foundations of Modal Semantics*. Princeton, N.J.: Princeton University Press.
Stalnaker, Robert (2014). *Context*. New York: Oxford University Press.
Starr, William B. (2016). 'Dynamic Expressivism about Deontic Modality'. In Starr, W. B. and Chrisman, M. (eds.) *Deontic Modality*. New York: Oxford University Press.
Stevenson, C. L. (1937). 'The Emotive Meaning of Ethical Terms'. *Mind*, 46: 14.
Stevenson, C. L. (1944). *Ethics and Language*. London: Oxford University Press.

Steward, Helen (1997). *The Ontology of Mind: Events, Processes, and States*. Oxford: Clarendon.

Toppinen, Teemu (2013). 'Believing in Expressivism'. In Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics, Volume 8*. Oxford: Oxford University Press.

Unwin, Nicholas (1999). 'Quasi-Realism, Negation and the Frege-Geach Problem'. *Philosophical Quarterly*, 49(196): 337-352.

Unwin, Nicholas (2001). 'Norms and Negation: A Problem for Gibbard's Logic'. *Philosophical Quarterly*, 51(202): 60-75.

Vendler, Zeno (1967). *Linguistics in Philosophy*. Ithaca, N.Y.: Cornell University Press.

Wedgwood, Ralph (2001). 'Conceptual role semantics for moral terms'. *Philosophical Review* 110(1): 1-30.

Wedgwood, Ralph (2007). *The Nature of Normativity*. Oxford: Clarendon.

Willer, Malte (2017). 'Advice for Noncognitivists: Advice for Noncognitivists'. *Pacific Philosophical Quarterly*, 98(1): 174-207.

Wimmer, H., and Perner, J. (1983) 'Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception'. *Cognition*, 13: 103-128.

Whiting, Daniel (2007). 'Conceptual Role Semantics'. *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/.

Woods, Jack (2017). 'The Frege-Geach Problem'. In McPherson, T and Plunkett, D. (eds.) *The Routledge Handbook of Metaethics*. London: Taylor and Francis.

Wright, Crispin (1988). 'Realism, Antirealism, Irrealism, Quasi-Realism'. *Midwest Studies in Philosophy*, 12(1): 25-49

Yalcin, Seth (2007). 'Epistemic Modals'. *Mind*, 116(464): 983-1026.

Yalcin, Seth (2011). Nonfactualism about Epistemic Modality. In Egan, A. & Weatherson, B. (eds.) *Epistemic Modality*. New York: Oxford University Press.

Yalcin, Seth (2012). Bayesian Expressivism. *Proceedings of the Aristotelian Society*, 112(2): 123-160.

Yalcin, Seth (2014). 'Semantics and Metasemantics in the Context of Generative Grammar'. In Burgess, A. and Sherman, B. (eds.) *Metasemantics: New Essays on the Foundations of Meaning*. Oxford: Oxford University Press.

Yalcin, Seth (2018a). 'Expressivism by Force'. In Fogal, D., Harris, D. and Moss, M. (eds.) *New Work on Speech Acts*. New York: Oxford University Press.

Yalcin, Seth (2018b). 'Belief as Question-Sensitive'. *Philosophy and Phenomenological Research*, 97(1): 23-47.

Yalcin, Seth (2018c). 'Semantics as a Model-Based Science'. In Derek, B. and Rabern, B. (eds.) *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*. Oxford: Oxford University Press.