

# Facts and the Function of Truth

Extended Edition

Huw Price

May 2024

This is a draft of the new extended edition of *Facts and the Function of Truth*, forthcoming from Oxford University Press. If you wish to cite it before the final version appears, please refer to it as '*Facts and the Function of Truth, Extended Edition* (draft)', including the URL at PhilPapers, and date of access.

## Contents

Preface to the Second Edition	ii
Preface to the First Edition	v
0. Introduction to the Second Edition	1
1. Introduction	31
<b>Part I: A Sceptic's Guide to the Matter of Fact</b>	<b>41</b>
2. The Place of Truth	42
3. The Insignificance of Mood	56
4. A Retreat to the Head	66
5. Fact and Psychological Function	81
<b>Part II: Truth and Its Limits: An Explanatory View</b>	<b>99</b>
6. The Function of Truth	100
7. The Power of Negative Thinking	117
8. The Limits of Intolerance	133
9. End to Enquiry	151
<b>Part III: Thirty Five Years Downstream</b>	<b>161</b>
10. Embracing Minimalism	162
11. Cambridge Expressivism	186
12. Generality and Causality	214
13. Neighbours and Rivals I: No Fault Disagreement	239
14. Neighbours and Rivals II: Explaining Truth	259
Bibliography	281

## Preface to the Second Edition

This new edition of *Facts and the Function of Truth* (FFT) has had a long gestation, even by the standards of the elephant it has become. It was conceived in 2000, in the first months of the new century. In March that year I had lunch in London with Peter Momtchiloff, then and for more than two decades afterwards Commissioning Editor for Philosophy at OUP, Oxford. As always, Peter was keen to hear about publishing plans. Mine were about to change.

We were in London for a conference on Pragmatism. One of the other speakers was Bob Brandom (only the second time we had met in person, I think). My talk, an early version of (Price 2003), developed ideas about truth from FFT as a criticism of Rorty. Brandom commented afterwards that he hadn't read FFT, 'but should.' That may have been the seed. The 1988 edition had been out of print for several years at that point, and Blackwells had relinquished the rights. Where would Brandom find a copy?

A few weeks later, picking up from our conversation at lunch, I emailed Peter to ask whether OUP might be interested in publishing a second edition. Saying that it was 'now common for me to meet people who don't know [FFT], but are interested in the ideas', I added that there was 'no better example than Brandom, at the conference in March. ... So I think there'd be a market.'

This boutique marketing strategy did the trick. Peter responded positively, and encouraged me to think about adding new material, at the beginning or end: '[I]he more you are able to add at the beginning or end of the book, the more attractive a new edition would be.' Trying to set myself a realistic goal, I proposed simply a new Introduction, relating the original text to recent work by me and by others – another 10,000 words or so. (I said that I was particularly keen to stress the links between FFT and what I was then coming to call subject naturalism.) OUP accepted the proposal in this form, and we signed a contract a few months later.

Various factors then got in the way of progress, for many years. Other projects and deadlines always seemed more pressing, more manageable, or both. I discovered that there were two worsening consequences of the delay itself. The mountain kept getting higher, in the sense that the amount of new material I would need to discuss was always growing. And as the original edition grew more distant, and my felt connections with its author more tenuous, the prospect of returning to take ownership seemed less and less attractive.

So, as the century entered its late teens, it would have been reasonable to predict that the new edition might never fall from the press, in Hume's famous phrase, still-born or otherwise. But two things changed, one local and one global. The smaller thing, substantial from my personal perspective, was that Lionel Shapiro and Johannes Haag kindly proposed a workshop in Potsdam, Germany, in June 2018, to mark the thirtieth anniversary of the first edition.

The Potsdam workshop gave me a deadline to think about the original project of the book, and its contemporary relevance. It convinced me that the package of views on offer in FFT remains interesting and unorthodox – and, importantly, 'not one that a reader can now find on the shelf in other, perhaps more digestible, versions', as I put it below (§0.2). So it renewed my sense of the usefulness of a new edition, and gave me a set of notes which, as I realised at the time, could well be the basis of my new Introduction.

Still, the project made no further progress, for almost two years. I had many commitments in Cambridge, and little time for writing. Then, in March 2020, the whole world changed. When Covid came to the UK, an accident of timing found me locked down Down Under, many hours

ahead of my Cambridge colleagues. We were all working from home, but for me, visiting family in Sydney, that meant evening commitments. Suddenly there was daylight in my diary, in daylight hours.

Thinking that it might be *weeks* before my wife and I could return to Cambridge, I looked around for writing projects, to take advantage of this windfall. With my Potsdam remarks on hand, the new Introduction looked eminently achievable. Twenty years to the month from conception, I set to work, at the same time extracting a clean version of the text of the first edition from a pirate scan I had once found online.

Six months later, we were still in Sydney, with no near-term prospect of the borders opening up. This was months before vaccines, and in Australia, behind its quarantine barrier, there was very little Covid. It was beginning to look as though we might never return to Cambridge. I was retiring from my chair at that point in any case, though I had a continuing half-time role as Academic Director of the Leverhulme Centre for the Future of Intelligence.

However, my elephant was approaching its current size and shape. I had a 17,000 word draft of the new Introduction, plus another 40,000 words towards, at that stage, three new chapters to be added at the end. (There are now five.) Embarrassed by these dimensions, I was relieved to rediscover Peter Momtchiloff's initial assurance that the more I was able to add, the more attractive the new edition would be. I am delighted to have this opportunity to express my gratitude to Peter for taking the elephant in his stride, as well as for his support and patience at every point.

Completing the project took another three years. In part, this is because it produced so many opportunities for digressions. Some of these were detailed comparisons of my own approach with the work of near-neighbours: John MacFarlane, Crispin Wright, Allan Gibbard, and most recently Bernard Williams. Three of these digressions produced stand-alone papers, now in print (Price 2022b, 2022d, 2023a), and much of this material was then incorporated into the final new chapters, as I finished them off. These three papers were written in response to approaches from the editors of journal special issues, Matej Drobnák, Tom Kaspers, and Thomas Spiegel; I am grateful to them for the invitations and for comments on early drafts. In the case of the comparison with Bernard Williams, it was prompted by my discovery of excellent recent work on Williams and genealogy by Matthieu Queloz, to whom I'm indebted for comments and discussion.

Another welcome source of delay was an opportunity at the University of Bonn. After my full retirement from Cambridge in October 2021, I was delighted to be offered a two-year part-time Distinguished Emeritus Professorship at the Centre for Science and Thought in Bonn. This involved two three-month periods of residence in Bonn, in the northern summers of 2022 and 2023. I am grateful to Markus Gabriel, Jan Voosholz, Christiane Schäfer, and Barbara Wiehlpütz for their help and hospitality there.

Among other things, these visits to Bonn provided an opportunity for me to present some of this new material at the REWE2023 conference there, organised in association with Christine Tiefensee (Frankfurt School) and Charlotte Gauvry (Bonn), as well as at meetings in Bologna and Alghero. Earlier, thanks to Céline Henne and Oscar Westerblad, I had had the opportunity to discuss my draft chapters in an online reading group on Pragmatism, organised from HPS, Cambridge. Finally, I am especially indebted to Lionel Shapiro and Joshua Gert, for written comments on various drafts.

Many participants in such events, in Potsdam, Munich, Helsinki, Bonn, Bologna, Alghero, and variously online, gave me helpful comments on this and related material. My debts certainly outrun my memory, at this point. I hope it won't be invidious to thank, in addition to folk already mentioned, Helen Beebee, Filippo Ferrari, Yvonne Hütter, Luca Incurvati, Jenann Ismael, Jonathan Knowles, Cheryl Misak, Carlo Rovelli, Mat Simpson, Amie Thomasson, Crispin Wright, and José Zalabardo.

## Preface to the First Edition

Some years ago I worked on the philosophy of probability. For a long time I thought about probabilities – what kinds of things they could be, and so on. Then it struck me that this might be the wrong approach. We didn't have to assume that there must be something that probabilistic judgements are characteristically *about*, and so concern ourselves with the nature of that distinctive subject matter. Instead we could interpret probabilistic judgement in terms of a characteristic way of talking and thinking about other things. At the time it seemed quite a revelation. Of course I soon realized that others had expressed the same view about probability, as about several other topics. But in the case of probability, at least, the view was not well developed. I went on to write a thesis (Price 1981) that argued its advantages over other views of probability, and defended it against certain objections raised against similar 'non-factualist' views in other areas.

By contemporary standards the unusual feature of this view of probability is that it takes a linguistic approach to metaphysical issues. Non-factualist theories were a common and characteristic product of the mid-century linguistic movement in philosophy; but these days that movement is widely felt to be discredited. Contemporary metaphysics regards itself as free of the peculiar concern with language that characterized that period. So as I eventually came to appreciate, a defence of such a treatment of probability (or indeed other topics) calls for some defence of the linguistic stance. More importantly, it calls for an elucidation of the cluster of notions whose proper application non-factualists are concerned to dispute – notions such as *assertion* and *statement of fact*. In this respect, as in others, the linguistic approach needs to be grounded in the philosophy of language.

I found that such a grounding had not in practice been obtained. This was not simply because it had usually not been sought. Contemporary philosophy of language has appeared to have reason to dismiss non-factualism – a reason that turns on the role of the notion of truth in Fregean linguistic theory. In effect, non-factualism and contemporary philosophy of language have wanted to draw the limits of fact-stating discourse at different places. This conflict has not had the attention that it deserves, I think. The linguistic approach to difficult areas of philosophy has more to be said for it than its reputation suggests. And the philosophy of language has much more to learn from the encounter than the history would indicate.

This book is in part an attempt at mediation. I try to show that with a proper understanding of what an account of statement of fact should amount to the conflict can be resolved. Neither side escapes unscathed. The resolution turns, in particular, on the rejection of a shared assumption about the appropriate goal of an elucidation of the notion of statement. But both sides benefit from the new perspective thus required. The major benefit is a particularly fruitful perspective on the nature and origins of truth.

I appreciate that the approach may seem misguided. Many philosophers will feel that the conflict described is long since decided in favour of the philosophy of language. I suspect that others, on both sides of the fence, will not have noticed that it ever existed. I can only urge such people to read the book – or at least to read chapter 1, which contains a more detailed guide to the nature and origins of the conflict, and the structure and benefits of the resolution I propose.

In order to make it a little easier to read the book, I have starred a number of sections of a more technical or incidental nature. These may be omitted without serious discontinuity. Chapter 1 mentions a further possibility, which involves beginning at chapter 6.

The book was written during my tenure of an Australian National Research Fellowship, in the School of Philosophy at the University of New South Wales. The Editor of *Mind* has kindly allowed me to reproduce parts of 'Truth and the nature of assertion' (Price 1987), which appear mainly in chapter 2. I am also indebted to Philip Pettit, whose assistance and encouragement long since transcended the merely editorial; and to Peter Menzies and Phillip Staines, for their insight and patience in many conversations on this material.

## Chapter 0

### Introduction to the Second Edition

This new edition of *Facts and the Function of Truth* (FFT) contains six new chapters, in addition to the nine chapters of the first edition (which are unchanged here, except for minor grammatical and typographical corrections). The new material comprises this new Introduction (Chapter 0), and the new Part III (Chapters 10–14).

The new chapters have two main aims. I want to try to make FFT as accessible as possible to new readers, and I want to discuss some connections between the themes of FFT and later developments, both in my own work and in the field at large. These aims overlap, of course. One way to make FFT accessible to contemporary readers will be to relate its themes to ideas and terminology that have since become familiar elsewhere.

For both tasks, it will be helpful to explain what the author of FFT had in mind. This isn't as straightforward as it might seem. Memory doesn't get me into his head, at this point, at least in crucial respects. What follows is therefore my best shot at rational reconstruction, informed by textual evidence in FFT and elsewhere, as well as my recollections of my circumstances and influences, during the relevant period.

The remainder of this chapter is in three main parts. §0.1 explains where FFT came from – its origins in my early work. §0.2 provides a big picture view of FFT's structure and main themes. §0.3 offers an extended reading guide, describing the main arguments and conclusions of each chapter, and relating some of them to later work. (Chapters 10–14 take up some connections of this kind in a lot more detail.)

#### 0.1 The origins of FFT

FFT was written in Sydney in the period 1985–87, some five years after I completed my Cambridge PhD thesis (Price 1981). It is aligned with the conclusions of my thesis in some respects but diverges markedly in others. This combination is a clue to how I conceived the project. So I'll describe my thesis, and then the divergence. Let's begin with some historical context.

Since the early days of analytic philosophy, it has often been suggested that some of the claims we make in language are not what they seem, and that this has consequences for traditional philosophical debates. In Cambridge and Vienna a century ago, for example, this view was linked to the attempt to distinguish a 'primary' factual or scientific language from various 'secondary' uses of language – including various 'pseudo-statements', as it was sometimes put.

In recent decades this kind of claim is especially familiar in metaethics – indeed, many who write about it in metaethics seem unaware that it has other applications. Metaethical *non-cognitivists* say that moral claims are not genuine factual claims, and that this avoids metaphysical puzzles about the nature of moral properties. The assumption that there are such things as moral properties, for philosophy to concern itself with, rests on a mistake about the nature of moral language. (This way of putting things reflects the fact that when I was writing, and by some still today, non-cognitivism was thought to combine a metaphysical thesis of its own – the claim that there are no moral properties – with its thesis about the function of moral language. More below on the project of pulling these things apart.)

My PhD thesis defended a view of this kind about probabilistic claims – in particular, claims of the form ‘It is probable that P’, or ‘It is unlikely that Q’, in which the probabilistic element takes the form of a sentential operator. I argued that claims of this kind should be understood not as assertions about a distinctive realm of probabilistic facts, but rather as expressions of degrees of confidence, or ‘partial beliefs’. On this view, the claim ‘It is probable that P’ stands to a (highish) degree of confidence that P just as the assertion ‘P’ stands to a (fullish) belief that P. In other words, ‘It is probable that P’ is to be understood not as an expression of a *full belief* (that it is probable that P), but as an expression of a *partial belief* (that P).

Contemporary readers may know this view in a form proposed by Yalcin (2012), who calls it *credal expressivism*. As Yalcin says, the view

develops the thought that in asserting something like

(1) Allan is probably in his office,

one may express an aspect of one’s credal state, without describing that state. One expresses one’s confidence, that is, without literally saying that one is confident. The relevant credal state expressed is ... not a state tantamount to full belief in a proposition. (Yalcin 2012, 125)

My thesis argued that a great advantage of this approach is that it avoids a problem that plagues all rivals, that of explaining why a full belief that it is probable that P is, or should be, normally accompanied by a high degree of confidence that P (and by the typical manifestations of such a degree of confidence, such as a preference for a bet that P over a bet that not-P, other things being equal). This problem has a long history. One familiar manifestation of it is the challenge of justifying what David Lewis (1986) calls the ‘Principal Principle’ – a principle linking beliefs about chances to credences, or partial beliefs. It is the probabilistic analogue of what in FFT I called ‘the approval problem’, the challenge of explaining why a belief that it is good that P is, or should be, typically accompanied by a desire that P – by a preference for options such that P, other things being equal, at least.

In both the probabilistic and evaluative cases, one tradition attempts to meet this challenge by defining probability or value in terms of *rational* degree of belief, or *rational* preference. (See, e.g., Mellor (1971) in the probability case, and Smith (1994) in the moral case.) In my thesis I was already sceptical about this approach, which seemed to me simply to move the bump in the carpet. Why should our actions be guided by beliefs about rationality? That’s the same problem all over again.

At any rate, I took this to be a reason for thinking that ‘It is probable that P’ is not an ordinary assertion with a content concerning the probability that P. Instead it is a different kind of speech act altogether – what I termed a ‘partial assertion’ – with content *P*.<sup>1</sup> So this was a non-cognitivist

---

<sup>1</sup> I may well have found the idea in Toulmin (1950), as characterised in turn by John Passmore. In his 1957 book *A Hundred Years of Philosophy* – recommended to me by Hugh Mellor in 1976, as suitable reading for an overland journey from Australia to Europe! – Passmore gives Toulmin’s view as an example of Oxford ordinary language philosophy.

Austin’s line of reasoning is taken over, to cite one of many examples, by S. E. Toulmin – of Cambridge origins but subsequently an inhabitant of Oxford – in his ‘Probability’ [1950] ... [Toulmin] thinks that to say ‘S is probably P’ is to make a guarded and restricted statement: it is to commit ourselves to a certain degree ... There is no particular ‘thing’, Toulmin concludes, that probability statements are about ... [A]

account of ‘It is probable that P’, defended, as was common in the metaethical case, by the argument that it makes better and easier sense of the agreed *practical relevance* of claims of this kind – of their link to choice behaviour.

In addition to this *pragmatic* case for the account, as we might term it, I thought I had a novel argument that claims such as ‘It is probable that P’ lack truth conditions. It turned on the fact that probabilistic claims of this kind can give rise to what I termed ‘non-critical disagreements’ – cases in which an apparent disagreement turns out to involve no fault on either side, because two speakers had access to different bodies of evidence. I argued that a genuine disagreement about a factual matter should not allow this possibility.

I also considered at length the objection that this kind of ‘No Fault Disagreement’ (the term I used in FFT) simply reflects an implicit relativisation of probability claims to a body of evidence. I argued that no construal of what this evidence might be gets things right. Make it too subjective, and the case becomes an exchange in which, as Ramsey puts it, ‘we are each saying something about ourselves which pass by one another like “I went to Grantchester”, “I didn’t”’ (Ramsey, 1929, 137) – there is no sense of disagreement in the first place, in other words. Make it too objective, and we lose touch with the kind of evidence on which we actually rely in making such judgements. (We often know that there’s more evidence out there, but have to act without it.)

As I say, I thought of these points about disagreement as an additional argument that ‘It is probable that P’ lacks genuine truth conditions. I backed this argument up with a kind of just-so story. We all agree (I took it) that questions of the form ‘Is it the case that P?’ lack truth conditions. But we could imagine a language that I called (not very imaginatively) *Q-English*, in which such a question could be asked in a superficially declarative way, by saying ‘It is queried whether P’. ‘That’s true’ and ‘That’s false’ could then be used by another speaker to indicate, in effect, that she also wished to ask, or declined to ask, the original question. Because speakers can quite properly have different reasons for asking questions, this use of ‘That’s false’ would not be associated with the view that the other speaker was mistaken, in many circumstances. And I took this to reflect the fact that the Q-English statement ‘It is queried whether P’ would not be genuinely truth conditional, despite its declarative form.

In summary, this is where my thesis put me. I was advocating a fairly orthodox kind of ‘non-factualism’ (my term from FFT) for the probabilistic case, building a case by two routes, one familiar, one less so. The familiar route argued that unlike its factulist rivals, this view could explain the practical relevance of probabilistic judgements in decision-making under uncertainty. The less familiar route appealed to the possibility of No Fault Disagreements about probabilistic judgements, and argued that this shows that they are not genuinely truth-conditional. (I also defended the view against the well-known Frege-Geach argument, on which more in §10.4 below.)

I had a postdoc at ANU, Canberra, in 1981–1983, the years immediately following my PhD. This gave me time to write up some of the material from my thesis, and to explore new directions. Some of the central arguments from the thesis appeared in (Price 1983a), a paper called ‘Does “Probably” modify sense?’ My answer was, ‘No, it modifies *force*’. The title reflected my growing interest in the central distinction between sense and force in a Fregean approach to meaning. One of my influences on this topic was Michael Dummett. I had cited Dummett’s work at a

---

probability statement is not distinguished from other statements by having a special subject-matter, but by involving a special degree of commitment. (1957, 458)

couple of points in my thesis, and read more with the encouragement of Christina Slade, then working on Dummett's verificationist theories of meaning for her ANU thesis.

As well as Dummett's classic work on Frege (Dummett 1973), I certainly found congenial his insistence that we should think about what we *do* with the notions of truth and falsity. In FFT I quote the following passage from one of Dummett's earliest papers:

At one time it was usual to say that we do not call ethical statements 'true' or 'false', and from this many consequences for ethics were held to flow. But the question is not whether these words are in practice applied to ethical statements, but whether, if they were so applied, the point of doing so would be the same as the point of applying them to statements of other kinds, and, if not, in what ways it would be different. (Dummett 1959, 144)

Dummett's influence shows up in two other papers I published in 1983, as well as in a fourth paper written at the same time, parts of which later emerged in Price (1994) and Price (2004a). Originally entitled 'The Use of Force in a Theory of Meaning', this fourth piece earned an invitation to revise and resubmit from *Philosophical Review*, before other things, among them fatherhood, consigned it to a bottom drawer.<sup>2</sup> It drew on material on the Frege-Geach argument from my thesis, but also argued that the sense–force distinction is not as clear cut as usually assumed. I propose that negation, for example, might begin life as a force modifier, and earn the right to behave like a sense modifier. This material was certainly moving in the direction of FFT.

Of the two other published papers, the first (Price 1983b) is a defence of classical logic – in particular, the principle of Double Negation Elimination – against Dummett's famous criticisms. The defence relies on the suggestion that meaning be determined by both assertion and denial conditions, rather than, as Dummett himself proposed, one or the other. As I note, the suggestion requires that denial be treated as distinct kind of speech act from assertion, rather than defined as assertion of a negation:

The account of sense and of negation here proposed requires that denial be taken as a separate activity from assertion. Denying that S cannot be defined as asserting that not-S; on the contrary, the account of negation appeals to the notion of denial. (Price 1983b, 171)

But Dummett himself invokes an objection of Frege's – the original form of the Frege-Geach argument – against this idea. As I point out, however, it is easy to avoid this objection if we allow that the force–sense parsing of may be non-unique, so that 'Not-P' may be described, equivalently, either as an assertion that not-P or as a denial that P. And this seemed to me to be no big price to pay.

[W]hy should we impose on a theory of meaning the requirement that the meaning of every utterance should have a unique resolution into a component due to sense, and a component due to force? After all, in saying 'Eric is flying and Eric is dreaming', do I make one assertion (of a conjunction) or two? We could ignore the latter interpretation if there wasn't a general use for a notion of conjunction external to force indicators (for example, in describing the form of an utterance such as 'I'm resting, and don't disturb me again!'). Similarly we could do without an independent notion of denial, if its only role was to provide a (redundant) account of the form of negated sentences. But if the sense

---

<sup>2</sup> Rescued from an ancient hard drive, it is now accessible as (Price 1983d).

of an assertoric sentence is partially determined by its denial, or rejection, conditions, then the interpretation of negated sentences is far from the only use of a notion of denial; and we must live with the resulting redundancy. (1983b, 172)

Whatever the merits of this little argument, it is clear that I was moving in the direction of the view that differences of *sense*, or *content* – e.g., that between the sense of ‘Not-P’ and the sense of ‘P’ – could ‘sit on top of’ differences in *force* (i.e., something more pragmatic or functional in nature). As I said a moment ago, this was certainly a move in the direction of FFT. I explore the move further for negation in (Price 1990), while in (Price 1994) it develops into my general response to the Frege-Geach argument – more on this in Part III (§10.4).

My last Dummett-influenced paper from that period is a little-known piece called ‘Could a Question be True?: Assent and the Basis of Meaning’ (Price 1983c). This, too, reflects my interest in disagreement and in the sense–force distinction, and in the connections between the two. By this point, my earlier concern with what I saw as ‘non-standard’ ascriptions of truth and falsity – ascriptions to utterances such as probability judgements and the indicative queries of Q-English, that are not genuinely truth-conditional – had led me to Dummett’s question: Why do we call each other’s utterances ‘true’ and ‘false’ in the first place? What is the *point* of doing so, as Dummett puts it?

The paper proposes an answer to this question. It is the beginnings of an account of the role of truth and falsity that plays a central role in FFT (and in later work, such as Price 2003). The paper puts the proposal so clearly that I think it is worth quoting from it here. I begin by asking ‘why it is not appropriate to assent to or dissent from utterances such as questions, commands and requests’ (1983c, 355). I note that ‘[o]ne function of the terms “true” and “false” is to provide a uniform means of endorsing or rejecting a statement made by a previous speaker’, and that this is in effect Strawson’s (1949) account. ‘However’, I say,

if this were all that ‘true’ and ‘false’ did, it would be unclear why they shouldn’t be applied to utterances other than assertions – to questions, for example. There are many circumstances in which it would be useful to have a simple device for putting one’s weight behind a question asked by someone else; for indicating, in effect, that one is asking the same question oneself. Similarly, there are situations in which it would be useful to be able to indicate in a simple way that a question did not have one’s endorsement. ... If a major function of the terms ‘true’ and ‘false’ is to provide a simple means of so endorsing or rejecting a previous utterance, why should these terms be applied only to assertions, and not to utterances of other force-types?<sup>3</sup> (1983c, 355)

I note that another possible answer is ‘that questions (commands, requests, etc.) don’t have truth-conditions, and are therefore not properly described as true or false’ (1983c, 355). But I argue that this looks ‘suspiciously circular’, especially if ‘holding true’ and ‘holding false’ are supposed to provide the evidential basis for a theory of meaning, as in the Tarski-Davidsonian framework I had in the background. I suggest that a more promising approach

lies in the claim that, as Dummett puts it, ‘the roots of the notions of truth and falsity lie in the distinction between a speaker’s being, objectively, *right* or *wrong* in what he says when he makes an assertion’ (1978, p. *xviii*). That is, to call an utterance ‘true’ is to say that it (or its speaker) is *correct*, or *right*; to call it ‘false’, to say that it (or the speaker) is *incorrect*, or *wrong*. (1983c, 355)

---

<sup>3</sup> As I say in FFT (§2.2) and in (Price 1987), Bernard Williams (1966) had made this point about Strawson’s account.

‘This is a step in the right direction’, I say, ‘but not a solution’:

For it has not been explained how to call an utterance ‘correct’ is to do more than simply endorse it (in the sense seen to be equally appropriate for questions); and to call it ‘incorrect’, to do more than to reject or decline to endorse it. It is no use appealing to further ways of saying that an utterance is incorrect, or wrong (that it is ‘at fault’, say, or ‘mistaken’). Rather we need an account of the function of this collection of descriptions, and some understanding of their consequences in typical conversations. (1983c, 355–356)

In effect, I was suggesting that Dummett does not push his own question about truth far enough. We need to ask about the point of having some such norm in the first place.

I then offer my own proposal, which, in four short paragraphs, is the core of the dialectical account of the functions of the norms of truth and falsity that I was to develop in FFT.

It seems to me that the primary significance of these forms of criticism lies in the fact that they constitute a challenge to a speaker to *justify* an utterance, and an indication of readiness on the part of the critic to engage in a *dispute*. In such a dispute, rival, incompatible views are exposed to common scrutiny. Ideally the more well-justified prevails, and one speaker recants, accepting the view of the other. Plausibly, there is enough of a general advantage in such dispute behaviour to explain the existence of a powerful linguistic device to facilitate it (i.e., the use of ‘true’ and ‘false’). This advantage will be explained in terms of the behavioural consequences of particular views, and the consequent benefits of basing one’s views on as wide a body of experience as possible.

For the deliberately vague term ‘views’ here, it is natural to read ‘beliefs’. There would seem too little point in such a “dispute” with respect to utterances expressing, say, desires, which could reasonably vary from speaker to speaker, even in the face of the same evidence as to matters of fact. In other words, the suggestion is this: some utterances (call them ‘assertions’) characteristically express states of mind (‘beliefs’) with respect to which there is reason to seek agreement between speakers. For these states of mind, “two heads are better than one”; there is a general advantage in exhibiting differences between speakers in this respect, so that less well-justified beliefs may be replaced by more well-justified ones. This explains why language has developed a general means of indicating such agreements and disagreements, in the application of the terms ‘true’ and ‘false’ to the associated utterances.

Utterances such as questions, commands and requests, on the other hand, characteristically result from states of mind for which no such reason for unanimity exists. Different speakers can reasonably hold conflicting such states of mind (conflicting in the sense that no one person could hold them concurrently), even if fully acquainted with each other’s viewpoint. Appropriately, ordinary usage does not apply ‘true’ and ‘false’ to the types of utterance which express, or result from, these states of mind.

The claim is thus that there would be a certain value in a general device for merely, in effect, repeating or declining to repeat a previous utterance; this value would lie in brevity and convenience. But the process thereby facilitated would be one of less significance than that which forms a part of a dispute procedure, whereby conflicting beliefs are brought under common scrutiny, to the general advantage of the speakers concerned. So the facilitating of this latter process is the more important task of the terms ‘true’ and ‘false’, and therefore the one which takes priority, in determining their extension to utterances of a particular kind. (1983c, 356–357)

After this, in material that clearly draws on my thesis, the paper goes on to discuss different ways in which two speakers can disagree, based on differences of evidence, differences of inference, and other things. And it defends what is effectively the same criterion for being a *genuine* assertion, or a *genuinely* truth-conditional claim, on which I had relied in my thesis. In the language of FFT, the test is a matter of not permitting No Fault Disagreements.

However, I soon came to doubt whether the No Fault Disagreement criterion could mark a sharp line. A crucial influence was the rule-following considerations, as presented in Kripke's celebrated (1982) book, *Wittgenstein on Rules and Private Language*. I had had the opportunity to discuss this book at ANU in 1983, in a reading group with Philip Pettit, Peter Menzies, and others. I became convinced that the ineliminable possibility that different speakers might diverge from one another, in virtue of possessing different dispositions to 'go on in the same way', was a *universal* potential source of No Fault Disagreements. If so, No Fault Disagreements couldn't possibly mark a boundary between genuinely truth-conditional uses of declarative language and all the rest (except in the trivial sense in which there is a boundary between a null class and everything else).

This doesn't seem to have led me directly to the conclusion I come to in FFT, that there is actually no such boundary to be marked. In a piece presumably written in 1985–86 (Price 1987), I propose instead that genuinely fact-stating claims are those that express genuine beliefs. I had already begun work on FFT at that point, and some of the material in this piece found its way into Chapter 2. The main claim is that the fact-stating/non-fact-stating distinction cannot be grounded on what David Wiggins (1980), following Bernard Williams (1966), had called a *substantial* theory of truth. At the end of the piece I propose the belief-based criterion as an alternative, and refer readers to FFT, under its original planned title *Statement of Fact*, for further discussion.

As I wrote FFT, however, I became convinced that these various attempts to ground the fact-stating/non-fact-stating distinction were taking in each other's washing – that there was no such distinction to be had. Instead, turning back to the No Fault Disagreement criterion, what we had was something that was always a matter of degree, with no pure cases.

In this way, my reliance on No Fault Disagreement as a mark of the kind of non-factuality I had wanted to attribute to probabilistic claims had led me, albeit indirectly, to the thought that *no* uses of assertoric language were fully factual, in the assumed sense. And this seemed to me an appealing option to have on the table, not least in Kripke's own case. I felt that the assessment of Kripke's 'sceptical solution', by Kripke himself amongst others, was insufficiently sensitive to the broader implications of the arguments. To put it crudely, it was less shocking that there were no genuine facts about meaning, if one had in any case had to renounce the conception of genuinely factual language, against which the contrast was supposedly being drawn.

Thus I had become sceptical of an assumption on which I had relied in my thesis, and which was (and is) still very widely taken for granted: namely, the view that there is a division in language, *within* the class of declarative claims, between genuinely factual claims and claims with some other role or status (or, equivalently, a sharp sense/force distinction).<sup>4</sup> I had also become aware of the attractions, as they seemed to me, of a kind of *global* non-factualism, that treated the familiar non-cognitivism of metaethics and probability as merely the tips of an all-encompassing

---

<sup>4</sup> I might have saved myself some time if I had been aware of criticism of this assumption by (Toulmin and Baier, 1952); see fn. 5 below.

iceberg. Moreover, I had come to this position, in part, by thinking about the role of responses such as ‘That’s true’ or ‘That’s false’ in dialogue (both in general, and in what I had previously seen as special cases, such as probability judgements and the artificial query-statements of Q-English).

With this combination, the three main themes of FFT were in sight, from my perspective: rejection of this assumption concerning a division in language, and of the notion of strict or genuine factuality on which it depends; a defence of a kind of global non-factualism; and an account of truth and falsity (and their limits) as norms of disagreement. Let’s now see how those themes develop, in the structure of the book.

## 0.2 Structure and main themes

FFT is in two main parts. Part I (Chapters 2–5) is a sustained argument against the assumption that there is a substantial property of ‘statementhood’, as I put it – i.e., a property of being a ‘factual’ or ‘descriptive’ claim – that some declarative speech acts have and others lack. I was thus rejecting what Robert Kraut (1990), discussing Rorty, was soon to dub ‘the bifurcation thesis’.<sup>5</sup>

Kraut notes the confusing variety of ways in which this thesis tends to be characterised:

The bifurcationist often undertakes the task of determining which of our well formed declarative sentences have truth conditions and which ones, though meaningful, are simply the manifestations of attitudes or the expressions of ‘stances’. He wants to know which of our predicates get at real properties in the world, and which, in contrast, merely manifest aspects of our representational apparatus—‘projections borrowed from our internal sentiments’. On different occasions he articulates his task in different ways; but they all point to some variant of the bifurcation thesis ..., the thesis that some declarative sentences (call them the D sentences)

- describe the world
- ascribe real properties
- are genuinely representational
- are about ‘what’s really out there’
- have determinate truth conditions
- express matters of fact
- limn the true structure of reality

whereas other declarative sentences (call them the E sentences)

- express commitments or attitudes
- manifest a ‘stance’ (praise, condemnation, endorsement, etc.)
- are expressive rather than descriptive
- do not ‘picture’ the world
- lack truth conditions, but possess ‘acceptance conditions’ or ‘assertibility conditions’

---

<sup>5</sup> Shapiro (2021b) points out that the bifurcation thesis is already clearly identified by (Toulmin and Baier, 1952), who call it ‘the Great Divide’. Shapiro notes that Toulmin and Baker are critical of the thesis, taking it to result ‘from conflating several cross-cutting distinctions drawn in ordinary language using terms such as “describe”, “statement”, and “act”’ (Shapiro 2021b, §2); he draws parallels with my criticisms of the use of the bifurcation thesis by ‘local’ expressivists such as Blackburn and Gibbard.

- merely enable us to ‘cope’ with reality
- are true (or false) by convention
- do not express ‘facts of the matter’. (Kraut 1990, 158–159)

As I say, ‘statementhood’ was my term for the property supposedly possessed by what Kraut calls the D sentences. (As noted above, my initial working title for the book was *Statement of Fact* – a choice that reflects the centrality of the investigation of the grounds of the bifurcation thesis in my conception of what the project was about.)

In Part I I set out to argue not quite that there is no such property, but that on the assumption that there is, neither of two main (conflicting) views that required it could defend their contention about where its boundaries lay. Neither view, in other words, could defend its own contention about where the supposed bifurcation falls in language. (In Part II I then aimed to show how we could make sense of the intuitions underlying both views, but without a property of this kind, or any sharp distinction in language.)

What were the two views in question? One was what I was calling non-factualism, i.e., the non-cognitivist-like view that I had defended for probabilistic claims in my thesis. (In FFT I say that I chose not to use the term ‘non-cognitivism’, because I took it to be too tied in people’s minds to the ethical case. These days, of course, ‘expressivism’ is a common label for this kind of view.) Non-factualists maintained that there is a bifurcation *within* the class of declarative utterances – some of them are genuine statements of fact, others have some other kind of meaning.

The second view was what I called ‘the marker view’. This was a position easily found in linguistics texts, and evidently assumed at least implicitly by many mainstream philosophers of language in what I referred to as the Fregean tradition. It holds that the indicative mood itself is a linguistic ‘mark’ that an utterance is a statement, or a factual claim. It recognises no bifurcation *within* the class of indicative utterances, but certainly sees one *between* indicatives and non-indicatives.

Part I aims, as I say, to drive both views around a sceptical circle. In each case, there were various options as to where the required distinction might be grounded. I aimed to show that all of them implicitly assumed that it was already grounded somewhere else. I took Quine’s famous critique of the analytic–synthetic distinction as a model for what I was trying to do. (At some point I was tempted by the thought that statementhood itself was another dogma of empiricism. In retrospect, it seems a bit unfair to try to pin that one on empiricism – it’s a dogma, certainly, but much more widespread.)

More on the internal structure of Part I below. Part II offers the positive proposals of FFT – as I say in the original Introduction, my intention was that it could be read before Part I, by readers primarily interested in the positive story. It can be seen as a defence of what I regarded as correct intuitions *underlying* non-factualism (and to a lesser extent the marker view), via an ambitious two-part route. The first component is a proposal about the dialectical, argument-encouraging role of the norms of truth and falsity. I was offering what I thought of as a novel account of truth, though with a crucial shift of emphasis compared to most rival accounts. As I stressed, I wasn’t proposing a new *analysis* of truth, an account of its nature, or anything of that kind; on the contrary, I was proposing an *explanation* of why creatures like us should have a use for the notion. (This shift from analysis to explanation is central to the book as a whole, and to much of my later work – indeed, it is the key move in what I now characterise as the most interesting form of pragmatism; see, e.g., Price 2017a, and Ch. 11 below.)

The central proposal is the one from (Price 1983c), viz., that the norms of truth and falsity give social creatures like us a useful motivation to resolve disagreements. Truth provides the normative ‘friction’ that makes disagreements matter, as I put it in later work (Price 2003). But the second part of the argument of Part II of FFT is that these dialectical norms allow No Fault Disagreements – to different degrees and for different reasons, in different discourses – in a way that tallies with the intuition that some matters are more ‘factual’ than others. I argue that this is a matter of gradation rather than bifurcation, and that no domain counts as entirely factual, by these lights.

I was thus proposing that we should not think of statementhood, or factuality, as a property *presupposed* by assertoric language (and hence a property of which we need an analysis, if we are to understand assertoric language). Instead we should think of it, to the extent that it is a property at all, as something that *arises within* assertoric practice, something that we explain by calling attention to the relevant features of that practice. (Again, I was relying on the contrast between analysis and explanation.)

These days it feels natural to me to put the point in vocabulary I learnt from Brandom. My view was that the category of factuality is ‘instituted by’ assertoric language, rather than being *presupposed* and *marked* by it (whether perfectly, as on the marker view, or imperfectly, as on orthodox non-factualist views). It was the kind of view to which Brandom alludes in passages like these:

The pragmatist direction of explanation ... seeks to explain how the use of linguistic expressions, or the functional role of intentional states, confers conceptual content on them. (Brandom, 2000, 4)

Starting with an account of what one is doing in making a claim, it seeks to elaborate from it an account of what is said, the content or proposition—something that can be thought of in terms of truth conditions—to which one commits oneself by making a speech act. (2000, 12)

FFT offered that kind of account of factuality. In terms I have used more recently (Price 2017a, 153), my proposal was that factuality and statementhood were explanatorily *downstream* rather than *upstream* of the assertoric mode of speech.

To illustrate this terminology with an example I offered in FFT, the categories of ‘manual tasks’ and ‘manipulation’ are *downstream* rather than *upstream* of the existence of the hand – they evolved with it, as it were, rather than being predefined functions that the hand evolved to perform. As I put it in a later piece:

[O]ur hands serve a huge range of functions, practically all of which can also be served by other means. Although there is no doubt that hands in themselves are theoretically significant objects of study, we miss their true significance if we fail to recognise that the category of manual tasks—of things done by hand—is not a unitary natural kind. If we say that the function of the hand is manipulation, and leave it at that, we miss something very important: we miss the underlying functional diversity. (Price 2004a, 223)

As I note, the point is a kind of dual to one of Wittgenstein’s famous metaphors for the plurality that he takes to underlie the ‘common clothing’ of assertoric language games.

It is like looking into the cabin of a locomotive. We see handles all looking more or less alike. (Naturally, since they are all supposed to be handled.) But one is the handle of a crank which can be moved continuously (it regulates the opening of a valve); another is the handle of a switch, which has only two effective positions, it is either off or on; a third is the handle of a brake-lever, the harder one pulls on it, the harder it brakes; a fourth, the handle of a pump: it has an effect only so long as it is moved to and fro. (Wittgenstein 1968, #12)

In summary, then, the big take-away messages of FFT – both as I intended them then, in all but terminological details, and as I see them now – are three. First, factuality is downstream rather than upstream of a certain linguistic practice (in Brandom’s terms, the game of giving and asking for reasons). Second, truth plays a central normative role in this practice, and is itself best understood (where this means *explained* rather than *analysed*) in terms of this dialectical role. And third, factuality so construed comes by degrees, in the light of the variable prevalence of no-fault disagreements in different discourses – different reasons in different cases, but no pure cases and no sharp bifurcation.

From my present perspective, more than three decades downstream from FFT, this three-part package still seems original, significant, and essentially correct. Yet it remains highly unorthodox. More than three decades downstream, FFT is still swimming firmly against prevailing currents. Partly for that reason, the package it offers is not one that a reader can now find on the shelf in other, perhaps more digestible, versions. That’s why it seems worthwhile to make FFT itself as accessible as possible to a new generation of readers.

### 0.3 Chapter summaries and commentary

In the interests of accessibility, I now offer a kind of extended reading guide for Parts I and II of FFT (which are unchanged in this edition, except for minor typographical and grammatical corrections). This section comprises chapter by chapter summaries, in some cases with additional commentary linking the arguments of the chapter in question to later work in the field. These summaries and commentaries come in variable levels of detail, depending on the density of the original material, and on what I have to say about its connections to more recent work. (In general, my impression is that Part I is both harder to follow than Part II, and more in need of annotation in the light of developments in the field in the past 30 years. So Part I gets a bit more attention here.)

As I said above, the big claim of Part I is that the supposed property of statementhood can’t be grounded in a nontrivial way, either by non-factualists or by the marker view. Wherever either view tries to ground it, they turn out to be vulnerable to opponents who aim to trivialise the notion, by arguing that the proposed grounding does not exclude the possibility that *all* utterances possess the property in question – i.e., that all utterances are statements of fact, according to the criterion proposed. In each case the non-factualist or the marker theorist can only meet the challenge by appealing to some further criterion, which implicitly assumes that the property is grounded somewhere else. (Hence the sceptical circle.)

#### 0.3.1 Chapter 2 – No substantial theory of truth

How are we to distinguish factual and non-factual uses of language? This question was certainly in the air in British philosophy in the 1970s. So, too, were the twin thoughts that truth might be the key to the distinction, and that not just any theory of truth would be up to the job. These issues were so much in the air that they came to be discussed *on the air*. There is a remarkable

discussion between P F Strawson and Gareth Evans, filmed by the BBC for a Third Year Arts course for The Open University in 1973 (Strawson and Evans 1973). Under the title *Truth*, Strawson and Evans discuss the question as to whether the factual/non-factual distinction can be grounded on a theory of truth (inclining to the view that it cannot be). I had seen his discussion at least once in the early 1980s, when I found it by chance in the subject catalogue of the ANU library, and organised a screening for colleagues and graduate students.

In FFT I borrow terminology from Bernard Williams (1966), via David Wiggins (1980), calling a theory of truth *substantial* if it is capable of grounding a distinction between factual and non-factual discourse. Chapter 2 argues that there is no such theory. The main claim is that all proposals of this kind end up taking for granted that the factual/non-factual distinction can actually be grounded somewhere else. In particular, they tend to rely on a psychological grounding, via the idea that factual claims express beliefs, whereas non-factual claims express other sorts of psychological attitudes, such as desires.

A major theme in the chapter, one that now seems a little laboured, is that some accounts of truth are simply too ‘thin’ to provide the required criterion. This point was certainly already in play. As I noted above, Williams (1966) makes it about Strawson’s (1949) ‘performative’ theory of truth. It is also prominent in the discussion between Strawson and Evans mentioned above. (The discussion concludes with a suggestion by Evans that we might appeal to belief, instead of truth, and a reply by Strawson to the effect that that might fare no better, for similar reasons.)

The reason my discussion of this point now seems a little laboured is that – thanks to the influence of Crispin Wright, Paul Horwich, Hartry Field and many others – minimalism about truth has become something of a default position in the last thirty years. At any rate, it is enough of a respectable option to make us immediately suspicious of the idea that it might do useful work in distinguishing different kinds of speech acts.

These days we speak more or less interchangeably of *thick* or *thin*, *inflationary* or *deflationary*, *non-minimal* or *minimal*, or *substantial* or *insubstantial* theories of truth. Two common characterisations of the heavyweight side of each of these distinctions go something like this. One is that truth is a notion that ‘does important work’ in some sort of scientific or philosophical theory, e.g., about language and/or thought. The second is that truth is the kind of thing that ‘has a nature’, and is hence a proper object of the investigation of its nature – of the kind of analysis that seeks to answer the question, ‘What *is* truth?’

This contemporary terminology relates to FFT in two ways. First, Chapter 2 uses ‘substantial’ in a sense that is narrower than, though compatible with, the first way of drawing these distinctions – in effect, a sense that has a *particular* theoretical role for truth in mind (that of distinguishing factual from non-factual uses of language). As I said, the idea that truth is insubstantial in the broader sense is now very familiar. The broader sense implies the narrower sense, of course, and that’s why the argument of Chapter 2 now seems a bit laboured.

The second way in which this contemporary terminology relates to FFT is that the account of truth offered in Part II of FFT is clearly insubstantial in the sense implied by the second distinction above. My proposal about truth in Part II is tied to a rejection of analysis, or of an enquiry into the nature of truth, in favour of what I call explanation.

Another relevant piece of the contemporary landscape, one that emerged within a few years of FFT’s publication, is an argument to the effect that minimal notions of truth make things difficult for non-factualists, non-cognitivists, expressivists, and others of a similar stripe. This

argument can already be discerned in a footnote by John McDowell (1981), as I note in FFT. But it soon became more prominent in versions offered by Boghossian (1990), Humberstone (1991), Wright (1992), and (from a different angle) Jackson, Oppy and Smith (1994).

In one sense, the argument of Chapter 2 can be seen as foreshadowing this argument. I, too, was arguing that truth, minimally construed, is too thin to bear the weight that the non-factualist wants to put on it. But then, as now, I wanted to stop in a completely different place. I wanted to show that non-factualist intuitions could be vindicated in other ways (albeit only in a global form). So my argument in Chapter 2 wasn't an argument *against* non-factualism, as the point is intended by others. It was simply an argument that non-factualism had to be grounded somewhere else.

I shall return to this point in the new Part III (§§10.1–10.4). It relates to a range of more recent discussions, both in my own work and elsewhere, about the implications for expressivism of minimalist approaches to truth.

### 0.3.2 Chapter 3 – Against the marker view

The considerations we have just been discussing are highly relevant to the dialectical role of Chapter 3. The role of this chapter is to undermine an option that might otherwise seem attractive, in the light of some of the arguments of Chapter 2. Suppose we are convinced by Chapter 2 that there's no notion of truth that will give us a bifurcation of the kind the non-factualist seems to require, *within* the class of declarative claims, between genuine statements of fact and the rest. Isn't it natural to conclude, with the marker view, that *all* declarative claims are statements of fact – and that marking that fact is just what the declarative mood is *for*?

The role of Chapter 3 is to deny this easy option to opponents of non-factualism, as well as to the marker view itself. The chapter proceeds by arguing that the marker view can't exclude the possibility that all speech acts are statements 'underneath'. If there were a property of statementhood, of the kind the marker view assumes, the central claim of the view would be unjustifiable.

The argument clearly descends from my imaginary example of Q-English in my PhD thesis, though with an important shift. In my thesis I took it for granted that questions and other non-indicatives don't really have truth conditions. The point of my indicative paraphrases ('It is queried whether P') was supposed to be to show that utterances without truth conditions could masquerade as something else (their true colours being revealed in dispute behaviour, or lack of it).

In FFT, however, it is the assumption that there is any such thing as being *genuinely* truth conditional (or any other useful characterisation of a class of genuinely factual claims) that is under attack. So indicative paraphrases now have a different point. They are offered in an attempt to show that by their own lights, marker theorists cannot exclude the possibility that such paraphrases capture the real 'underlying' logical form of non-indicative utterances. If the argument succeeds it means that there is no easy fallback position, for someone who takes Chapter 2 to be bad news only for non-factualism.

In addition to this dialectical function, Chapter 3 also puts on the table a question to be taken up in later chapters. The argument of this chapter relies on the fact that various linguistic tasks seem to be able to be performed either in an assertoric mode, using an indicative construction, or in a

non-assertoric mode, using a non-indicative. Where this is so, what does the difference amount to? What (if anything) does assertion add?

In Part II my answer is that what the assertoric mode brings with it is participation in the game encouraged by the norms of truth and falsity – again, in Brandom’s terminology, the game of giving and asking for reasons. In Chapter 8 I return briefly to the kind of speech acts typically performed by non-indicatives, proposing that their function provides little scope or need for participation in this game. Otherwise, however, they and the marker view play little further role in the book.

### 0.3.3 Chapter 4 – Non-factualism retreats unsuccessfully to psychology (I)

In Chapters 4 and 5, with the marker view set to one side, my imagined opponents are orthodox non-factualists. Their views all rely on the bifurcation thesis – the assumption that there is a division within the class of declarative utterances between genuine factual claims and the kind of pseudo-statements that require non-factualist treatment. Chapters 4 and 5 argue that this assumption is ungrounded. No adequate account is available of the property that is supposed to be confined to the ‘factual’ side of the line.

Chapter 2 had already argued that this account can’t be grounded in terms of truth. In Chapters 4 and 5 the main argument is in two steps. First, I argue that non-factualists can be forced to fall back to a psychological criterion – to the thesis that genuine statements of fact are those claims whose standard use is to express a *belief*, as opposed to some other sort of psychological attitude (such as a desire or a partial belief). Second, I argue that the required psychological distinction (the belief/non-belief distinction) is itself ungrounded, unless it is able to rely on an independent criterion for factuality, grounded somewhere else.

Chapter 4, in particular, now seems to me to be excessively dense, with a number of threads in play. It will be helpful to mention two major strands, before turning to the internal structure of the chapter. The first is a parallel I mentioned above, between the motivations for non-factualism in the probabilistic case, on the one hand, and the ethical or evaluative case, on the other. It is easy to see why this analogy was salient to me. I came to these issues from an unorthodox direction, via the defence of probabilistic non-factualism in my PhD thesis. I was keen to show – contrary to an orthodoxy that remains puzzlingly widespread even today (more on this in §11.6) – that to think of non-factualism simply as a metaethical doctrine is an absurdly blinkered view. (As I make it clear in FFT, I didn’t regard these two cases as the only possibilities – far from it. But the striking parallels made them a particularly good route to trying to remove the conventional blinkers.)

The second strand running through much of Chapter 4 is a kind of pincer movement, assailing my orthodox non-factualist opponents from two sides. As I have said, the overall strategy of Part I involves a sceptical challenge to these opponents. I argue that they cannot ground the required factual/non-factual bifurcation in a way that supports their contention that there is such a division in real life, within the class of ordinary declarative utterances of a language such as English. But in Chapter 4 this challenge is no longer personified by a single anti-bifurcationist rival, the across-the-board factualist. This rival is joined by another anti-bifurcationist, namely, an across-the-board *non-factualist*. More on the latter character in a moment. At this point the thing to keep in mind is that the orthodox non-factualist needs to steer between two extremes, universal factualism and a kind of universal non-factualism. The chapter’s sceptical challenge exploits opportunities on both sides.

Turning now to the structure of the chapter, §4.1 sets up the probability/ethics parallel, noting analogous issues and options concerning the semantics of what I call *simple probability ascriptions* (SPAs) and *simple ethical ascriptions* (SEAs), respectively. In both cases, the factualist takes these ascriptions to have a distinctive *content*, whereas the non-factualist says that the role of the ethical or probabilistic operator concerned is not to modify *content* but to modify *force*. The factualist thus incurs an obligation to say what the distinctive content is – in effect, to give us an account of what probabilistic or ethical claims are distinctively *about*. I mention some difficulties for that programme, but close with the suggestion that a factualist might seek to avoid them by taking probabilities or moral properties to be primitive and irreducible.

§4.2 discusses the typical metaphysical case for non-factualism. It describes how, especially but not exclusively in the ethical case, non-factualists have claimed that such facts would be ‘queer’, and questioned their epistemology, asking how we could know about them. On behalf of the factualist, I respond that these arguments are not convincing. The factualist can say that the facts in question are just as different as they need to be, charge the non-factualist with an impoverished world view, and help herself to the non-factualist’s account of how we come by such judgements. The section closes with the suggestion that non-factualist’s strongest card is a different one. The objection should be not so much that these facts are ‘queer’, or doubtfully accessible, but that they are *unnecessary* – we can explain the appearances (i.e., our apparent judgements ‘about’ such facts) without them. Metaphysical and explanatory economy thus recommends the non-factualist view.

I think that by contemporary lights, a significant omission from this section is any significant discussion of error theories, or fictionalist theories, of the domains in question. On the face of it, these approaches are equally entitled to whatever credit might be due to antirealism on the basis of these metaphysical arguments, without the additional semantic baggage that comes with the bifurcation thesis. Thus I could have used these options as an additional challenge to the orthodox non-factualist, at this point in my argument. (Had I introduced these characters, however, it would have been a subtle matter to chase them out again, in favour of non-factualism properly constructed. That move requires a kind of metaphysical deflationism that is still very underdeveloped in FFT.)

At any rate, §4.3 introduces a real non-factualist, in the person of Simon Blackburn. I describe the view that Blackburn then called ‘projectivism’, thinking of the Humean metaphor about ‘gilding and staining with a new creation’ (Hume 1998, 163). I quote Blackburn on the economical advantages of this view:

The projective theory intends to ask no more from the world than what we know is there – the ordinary features of things on the basis of which we make decisions about them, like or dislike them, fear them and avoid them, desire them and seek them out. It asks no more than this: a natural world, and patterns of reaction to it. (Blackburn 1984, p. 182)

At this point, the second kind of anti-bifurcationist makes their first appearance, dressed as an idealist:

[U]nless the projectivist can distinguish ... real beliefs [about the world] from ... ‘patterns of reaction’, then projectivism becomes a species of idealism – the view that all ‘facts’ are projections of our reactions to the world. This outcome obliterates the distinction at the very core of the non-factualist’s thesis. In the process, moreover, it overturns the projectivist’s claim to metaphysical economy. The idealist’s world can be as cluttered as

anybody else's: idealists differ from realists not so much as to what exists, as to what its existence amounts to.

To avoid the slide to idealism, the projectivist thus needs a core of genuine facts: a secure foundation, on which the explanation of higher-level appearances can be constructed. Blackburn talks of asking 'no more from the world than we know is there'. But how do we know what is really 'there', and what we put there? [Page ref to come from proofs]

Asking '[w]hat then would it take to arrest the slide to idealism?', I answer as follows:

The one live possibility seems to be to appeal to psychology. If the non-factualist could distinguish the genuine beliefs ... from the various attitudes from which we project, then the barrier to idealism would be readily apparent. Projectivism would be safely contained, its task being to explain the fact that some non-beliefs are treated like beliefs. [Page ref to come from proofs]

Thus the idealist helps to push the non-factualist towards a reliance on a belief/nonbelief distinction, a grounding for the bifurcation thesis that I want to argue is incapable of bearing this kind of weight.

For me, these arguments are the beginnings of what turned out to be a long engagement with Blackburn's work. I was, and am, very sympathetic indeed to Blackburn's programme. But I have long felt that it was internally unstable, unless it became a global rather than a local view. In later work I called this the global challenge to Blackburn's quasi-realism. (See Macarthur & Price 2007, Price 2015b.) These passages in FFT mark the beginnings of that challenge.

However, readers of FFT – especially readers who know my more recent work – might be puzzled by the figure of the idealist, who seems to play the role of a dialectical bogeyman. But even in FFT, I'm presenting myself as a kind of global non-factualist. And these days, contrasting myself to Blackburn, I describe myself as a global expressivist. So isn't the idealist just me?

Later in FFT, in §9.4, I say that idealism is too homogenous, missing the functional distinctions which are the residue of traditional non-factualism that I want to preserve. Another factor may be that at least in places in FFT, I was taking for granted that the projectivist was a metaphysical *anti*-realist, who *denies* the reality of moral properties, probabilities, or whatever. My rather crudely drawn idealist was then someone who made that anti-realism a global view. I didn't yet have clearly in focus an alternative global non-factualist, better characterised as a *deflationary* or *minimalist* realist, rather than as an anti-realist. This kind of non-factualist wants to *set aside* the metaphysician's realist/anti-realist debate, rather than coming down on the 'anti' side of it. (Later, I took to citing the Carnap of 'Empiricism, Semantics and Ontology' (Carnap 1950) as someone who is very clear about this distinction between siding with the anti-realist and rejecting the question; see Price 2009b.)

This metaphysically deflationary position is the one I really want to globalise. I say so late in FFT, when I contrast my view to several other positions, including idealism. But it is much more explicit in my later work.<sup>6</sup> I think it is fair to say that Blackburn himself was in transition about this point in the mid-1980s, when FFT was written. This particularly clear statement of the view is from his remarkable piece 'Morals and Modals', first published in 1987:

---

<sup>6</sup> Including in 'Metaphysical pluralism' (Price 1992a), written shortly after FFT appeared.

What then is the mistake of describing such a philosophy [i.e., Blackburn's quasi-realism] as holding that 'we talk as if there are necessities when really there are none'? It is the failure to notice that the quasi-realist need allow no sense to what follows the 'as if' except one in which it is true. And conversely he need allow no sense to the contrasting proposition in which it in turn is true. He need no more need allow such sense than (say) one holding Locke's theory of colour need accept the view that we talk as if there are colours, when there are actually none. This is doubly incorrect, because nothing in the Lockean view forces us to allow any sense to 'there are colours' except one in which it is true; conversely neither need it permit a sense to 'there are actually none' in which that is true. (Blackburn 1987, 57)

Had I had this deflationary option in plain sight earlier, it would have worked equally well as idealism in the role of the global threat to Blackburn's local view, on the non-factualist side of the dilemma. As things stand, however, it is an option that is poorly understood, even today. Many commentators persist in reading Blackburn as a metaphysical anti-realist. So the idealist version may still be the better one, for the purposes of the argument of §4.3.

§4.4 and §4.5 then turn, as I put it, 'to the second major strand in the contemporary case for ethical and probabilistic non-factualism'. As I say:

This line of argument has a more overt link with psychology than the supposedly metaphysical arguments we have just discussed. ... It focuses on the distinctive functional roles of probabilistic and moral judgements in our mental and behavioural lives. [Page ref to come from proofs]

The argument turns on the fact that in both the probabilistic and evaluative cases, judgements of the kinds in question seem to be linked to behaviour in very distinctive ways. Someone who professes to believe that it is probable that P is typically disposed to prefer a bet that P to a bet that not-P, other things being equal. And someone professes to believe that it is good that P is typically disposed to choose P rather than not-P, other things being equal. Crude as these characterisations are, they show that the factualist has some work to do. What is it about the content of these beliefs that gives them this practical relevance? I call these 'the confidence problem' and 'the approval problem', respectively; and note that the non-factualist avoids them, by taking the judgements in question simply to be expressions of degrees of confidence and desires. Looking ahead to discussion in Chapter 11, let me call these issues *Practical Relevance Constraints*, a term that Brad Weslake and I (Price and Weslake 2010) have used for the corresponding issue in the case of causation. (As we shall see, they represent a very general part of the case for non-factualist views, or their contemporary descendants.)

§4.4 deals with the probability case. As in my thesis, I note that a popular option is to appeal to rationality, defining probability in terms of reasonable or rational credence. I suggest that this simply moves the problem from one place to another, and that a better option is to take the so-called 'downward inference', from beliefs about probability to credences, to be analytic, or definitional. Someone who doesn't make the downward inference is someone who doesn't know what 'probable' means. (I call this, it now seems oddly, the 'contextual' view. The context concerned is ahead of the judgement, in its practical effects, not behind it, in its assumed epistemic environment.)

To this factualist option my non-factualist objects that it involves redundancy, in admitting *two* psychological states: both the belief that it is probable that P, and the corresponding credence or

partial belief that P. Why not ditch the former, and take 'It is probable that P' simply to be an expression of the latter? In response, as I put it, 'the factualist has one more card to play':

The non-factualist's attack has rested on the distinction between the full belief that it is probable that P, on the one hand, and the attitude of strong partial belief that P, on the other. But why shouldn't the factualist simply reject this distinction, claiming that these two mental attitudes are one and the same – that a belief that it is probable that P just *is* a strong degree of confidence that P? Factualism would thereby take on the explanatory advantages of non-factualism, without an embarrassing duplication of psychological attitudes.

In other words, this 'radical' factualism challenges the psychological distinction on which the non-factualist has been relying. ... The prospects for non-factualism again turn on its ability to distinguish genuine beliefs from other psychological attitudes.<sup>7</sup> [Page ref to come from proofs]

Leaving the probabilistic case at this point, §4.5 then steps smartly through the same dialectic on the ethical side. It emphasises the importance for the non-factualist of the Humean principle that 'beliefs by themselves are incapable of motivating action':

Hume's principle is clearly crucial. Someone who rejects it can agree that moral judgements express mental states on the active or motivating side, and yet insist that moral commitments simply are beliefs about moral facts. This will be the ethical equivalent of the radical factualism we have just encountered in the probabilistic case. In both cases the non-factualist has attempted to burden the factualist with a cumbersome and implausible duplication in the mental states associated with utterances whose status is in question (SPAs and SEAs, respectively). The radical factualist responds by accepting the model but denying that it involves any duplication: rather, she suggests, the non-factualist has confused the existence of two ways of describing a single mental state for the existence of two distinct mental states. [Page ref to come from proofs]

After deferring to Chapter 5 the question of the defensibility of Hume's principle and its probabilistic analogue, §4.6 discusses Blackburn's argument that non-factualism makes better sense of the apparent supervenience of moral facts on nonmoral facts. (I argue that the radical factualist can do just as well.) Finally, in the spirit of pushing the analogy between the evaluative and probabilistic cases as far as it will go, §4.7 offers a brief discussion as to whether there is a probabilistic analogue of the supervenience argument.

### 0.3.4 Chapter 5 – Non-factualism retreats unsuccessfully to psychology (II)

Chapter 5 begins by pointing out that the psychological grounding has been proposed for other kinds of non-factualist views, e.g., that of indicative conditionals, as well as in the ethical and probabilistic case. Hence, as I put it,

I think we are ... entitled to conclude both that non-factualism has often in fact put its faith in psychology; and that at least in certain central cases it is bound to do so. Our final sceptical task is to show that this faith is misplaced. Beliefs cannot be adequately distinguished from other propositional attitudes; not at any rate unless we are prepared

---

<sup>7</sup> I was sympathetic to this move on the radical factualist's part. As I noted in §0.1, I had argued in (Price 1983b), against Dummett, that a single utterance might be both a denial that P and an assertion that not-P, even if assertion and denial needed to be understood initially as different kinds of speech act. The radical factualist is generalising that point to the ethical and probabilistic cases.

to rely, as the non-factualist here cannot, on a characterization of this psychological distinction in semantic or metaphysical terms. [Page ref to come from proofs]

§5.1 takes on the Humean orthodoxy. In effect, I accuse the orthodoxy of a failure of imagination. ‘If we begin with an orthodox Humean model of ourselves’, as I put it, ‘we can imagine ... psychological modification[s] that would allow beliefs to motivate actions.’ Something of the flavour of one of these modifications is captured by the following suggestion:

Consider, for example, a classical economic rationalist, whose actions are based entirely on the profit motive. Such an agent forms intentions to perform those actions he believes will maximize his net monetary return. One might object that the rationalist *wants* to make a profit. But why should he? Why should his mind not simply be organized so that believing that an action would maximize his profits normally leads him to form the intention to undertake it (in just the way, and to whatever extent, belief/desire combinations are ordinarily supposed to lead to intentions)? [Page ref to come from proofs]

I also cite approvingly these lines from McDowell:

[T]he idea of the world as motivationally inert is not an independent hard datum. It is simply the metaphysical counterpart of the thesis that states of will and cognitive states are distinct existences; which is exactly what is in question. (1978, p. 19)

The section ends with two points. First, that the Humean criterion was in any case a very parochial place for a non-factualist to try the belief/non-belief distinction – it could only seem to be adequate to someone wearing the kind of blinkers that prevented them from noticing non-ethical non-factualism. And second, that the obvious place to look for a less parochial basis for the distinction was in contemporary functionalist theories of mind:

[The non-factualist’s] mistake may seem to have been to concentrate on too small an aspect of the functional organization of the mind – to concentrate on the initiation of action, to the exclusion of other functions of the mind. As it happens, this is exactly the mistake of which contemporary functionalist accounts of the mind accuse their behaviourist predecessors. In the next section we want to enquire whether these contemporary theories succeed where Hume failed, in distinguishing beliefs from non-beliefs. [Page ref to come from proofs]

Before turning to §5.2’s exploration of this possibility, I want to note some connections between §5.1 and some later literature.

In the year that FFT appeared, David Lewis published a paper (Lewis 1988) arguing against what he termed the Desire as Belief (DAB) thesis, that desires could be regarded as a special kind of belief. It seemed to me that Lewis’s argument proved too much. Simple examples such as FFT’s economic rationalist (see above) showed that the idea of an agent in whom beliefs played the functional role of desires was not incoherent. I wrote a short response with a version of this argument, replacing the economic rationalist with a perfect servant, motivated solely by beliefs about the consequences of his actions for his master’s happiness. I sent it to *Mind*, in the person of Simon Blackburn, then the Editor, as well as to Lewis.<sup>8</sup>

---

<sup>8</sup> Lewis’s response, dated 17 May 1988, is available in Beebe and Fisher (2020), 456–66.

*Mind* turned that first piece down, but by that time I had a different response, arguing that Lewis's formal argument ignored an independently plausible conditionalised version of the DAB thesis. Blackburn liked that one, and it was published as Price (1989). There is now a considerable literature on the topic. It often cites my (1989) paper, but not the discussion in FFT. But FFT still offers a much broader perspective on the issues at stake than can be found elsewhere.

Michael Smith (1987) had also argued against the DAB thesis on the grounds that beliefs and desires have different 'directions of fit'. As I characterise the point in Price (1989):

Beliefs have a 'mind to world' direction of fit; they aim at fitting the world, at being true. Desires on the other hand have a 'world to mind' direction of fit; we aim to change the world to fit our desires, and not vice versa. The idea is due to Anscombe, who uses it in her work on intention to try to distinguish assertoric and non-assertoric forms of language. (1989, 120)

Smith cashes out this metaphor as follows:

[T]he difference between beliefs and desires in terms of direction of fit comes down to a difference between the counterfactual dependence of a belief and a desire that *p*, on a perception that *not p*: roughly, a belief that *p* is a state that tends to go out of existence in the presence of a perception that *not p*, whereas a desire that *p* is a state that tends to endure, disposing the subject in that state to bring it about that *p*. (Smith 1987, 54)

Then a desire that was also a belief would have to have *both* directions of fit, and Smith takes this to be incoherent.

A state with both directions of fit would ... have to be such that *both*, in the presence of [a perception that *not p*] it tends to go out of existence, and, in the presence of such a perception it tends to endure, disposing the subject to bring it about that *p*. (1987, 65)

As I point out, however, this argument overlooks the fact that the proponent of DAB (our radical factualist from §4.5) will not dream of claiming that the desire that *P* is equivalent to a belief *that P*. On the contrary, it will be said to be equivalent to something like the belief *that it is good that P*. And this can be claimed to have world-to-mind direction of fit with respect to the content *It is good that P*, and mind-to-world direction of fit with respect to the embedded content *P*, with no contradiction. (The antecedents of Smith's counterfactuals are a perception that it is not good that *P*, in one case, and a perception that not *P*, in the other – and these are of course quite different circumstances.)

Let's return to Chapter 5. §5.2 argues that when we turn to contemporary functionalist theories of mind, there is simply nothing there to provide the distinction the non-factualist needs, between beliefs and other sorts of propositional attitudes. On the contrary, the entire field simply helps itself to the crude kinds of two-box Humeanism said to be embodied in folk psychology, and there are no resources whatsoever to provide the principled grounding that the non-factualist would require. The discussion is couched in terms that may now seem dated, involving positions such as the Representational Theory of Mind (RTM), as characterised by Jerry Fodor, and rivals such as the Syntactic Theory of Mind (STM), advocated by Stephen Stich. But the conclusion doesn't depend on the details of these views.

At the end of the section externalist views of content make a brief appearance, long enough for it to be pointed out that they cannot provide what the non-factualist is looking for.

In summary, then, we find that if content is individuated in terms of functional role, there is no apparent reason to distinguish more than one type of propositional attitude; and that if it is individuated in terms of external semantic relations, the limits of belief depend on the limits of fact. Neither possibility provides what the non-factualist now needs in order to answer the sceptic or the radical factualist: a psychological criterion with which to distinguish beliefs from other attitudes, and thereby to distinguish fact-stating from non-fact-stating discourse. [Page ref to come from proofs]

With this, the sceptical ‘great circle’ of Part I is complete. I claim to have shown that neither the marker view nor non-factualism is able to ground the fact-stating/non-fact-stating distinction in the way that their positions require.

The final section of Chapter 5 is an outlier, a detailed discussion of some issues relevant to non-factualist views of indicative conditionals. As I say:

There is one other important case in which what can be taken to be functionalist arguments have been widely thought to support a non-factualist conclusion. The case is that of indicative conditionals, and the arguments rest on Ernest Adams’s (1965) hypothesis ... that the ‘assertibility’ of an indicative conditional goes by the conditional probability of the consequent, given the antecedent (rather than by the probability of the corresponding material conditional, say). [Page ref to come from proofs]

In effect, the section defends the possibility of a radical factualist interpretation of indicative conditionals, in the light of Adams’s hypothesis. To do this it needs to respond to arguments due to David Lewis (1976) and to Carlstrom and Hill (1978) that there can be no conditional ‘ $\rightarrow$ ’ such that  $P(A \rightarrow C) = P(C/A)$ . In both cases, my response to these arguments draws on analogies between probabilistic judgements and indexical judgements, specifically with respect to their sensitivity to changing evidence. In the probability case, these are the kinds of considerations that had led me to favour non-factualism about probability in Price (1981) and Price (1983a). In FFT I am still sympathetic to that view, of course, when properly expressed in the form I have in mind for Part II. But in §5.3 I argue that a determined factualist can exploit them to evade the Lewis and Carlstrom and Hill results.

### 0.3.5 Chapter 6 – Explanation versus analysis, and the argumentative role of truth

Chapter 6 is the beginning of Part II, and of the positive face of FFT. It begins (§6.1) with the distinction between analysis and explanation, which I illustrate with the case of social class. I think what appealed to me about this case was that it is one in which we could make sense of the idea that class distinctions reflected ‘natural’ differences between individuals, while at the same time recognising that there’s an alternative: we can treat class, and all the norms that go with it, as evolved social constructs. Seen in the latter way, it seems clear that they call for explanation (of the practices that support them), not analysis (of the properties themselves). This shift was the one I wanted to recommend for linguistic categories, such as truth and factuality.

In one sense, the section now seems to me to make rather heavy weather of ideas I could have borrowed from elsewhere, such as from Hume or Nietzsche. In another sense, the hard work still seems necessary, because it still seems difficult for many philosophers to bring the distinction between analysis and explanation into focus – more on these issues in Part III.

The next section, §6.2, is a dense response to a concern that the explanatory approach to such linguistic categories might seem unacceptable, in virtue of being in tension with a certain conception of what speakers must be held to know, in choosing their words. My proposal wanted to put the underlying function of various linguistic forms ‘in the background’, in a place where they were certainly inaccessible to ordinary speakers (being winkled-out only from the kind of theoretical perspective I claimed to be offering). I felt that I needed to defend the coherence of such an account, given the gap it seemed to open up between what speakers take themselves to be doing in choosing their words, and what the theoretician takes them to be doing. The section includes the analogy I mentioned above, in which I compare the view of my opponents to the suggestion that the hand evolved to enable us to perform manual tasks. Obviously, this suggestion would be missing important underlying diversity in the functions of the hand. Equally obviously, ordinary folk need not be explicitly aware of this diversity, when they use their hands for one thing or another. Both points are relevant to the use I make of the analogy.

§6.3 and §6.4 describe my core ‘functional’ account of the notions of truth and falsity. To summarise it here, I can’t do better than to quote the opening paragraph of Chapter 7:

I have suggested that we try to explain the existence and application of notions of truth and falsity in terms of their functional role. In particular, I suggested that their main function might be to encourage a useful form of linguistic behaviour, namely reasoned argument. It seems that in virtue of their normativity, these notions enable and encourage us to favour opinions with which we agree and to disfavour opinions with which we disagree. Disagreements thus become socially unstable, and we are led to defend our opinions, to argue and to seek agreement. In the long run, I suggested, such disputes can be expected to have a beneficial effect on the behavioural dispositions of individual speakers. They help to ensure that as individuals we hold and act on attitudes that reflect, to some extent, the combined wisdom of our linguistic community. Our behavioural dispositions can thus be tested against those of other speakers, before they are put to use in the world. The guiding principle is that it is better to be criticized for claiming that tigers are harmless than to discover one’s mistake in the flesh. [Page ref to come from proofs]

Finally, §6.5 distinguishes my ‘explanatory’ approach to truth from another idea in the literature at that time, namely, that a realist notion of truth might *do explanatory work*, e.g., in accounting for the success of language users. This proposal had been made by Putnam, building on issues raised by Field:

The *success* of the “language using program” may well depend on the existence of a suitable correspondence between the words of language and things, and between the sentences of language and states of affairs. (Putnam 1978, p. 100)

This section discusses these ideas, and related work by Devitt and McGinn.

As I note, Field himself begins with a perspective that does align with the explanatory programme, in my sense.

Field asks ‘why we need to be able to apply semantic notions like truth and reference to the sentences that people believe and desire’ (1978, p. 48). This question seems to me to involve very much the same explanatory perspective as the approach suggested above. It

sees the task as that of explaining the ordinary use of the notion of truth. ... But in the discussion to which Field's work has given rise, the phrase 'explanatory theory of truth' has come to mean something different. The emphasis has shifted from explanation *of* (the use of) the notion of truth, to explanation *by means of* the notion of truth. [Page ref to come from proofs]

I argue that neither programme succeeds, in the forms present in the literature at that point.

[T]he recent search for an 'explanatory theory of truth', stemming from the work of Field, does nothing to provide an explanation of truth in our sense. Nor, incidentally, does it seem to make any progress towards a theory that would be explanatory in Putnam's sense – that is, in which truth is itself an explanatory notion. As we saw, the purported applications of such a notion seem to need nothing more than the disquotational property of truth. [Page ref to come from proofs]

### 0.3.6 Chapter 7 – Disagreement: how it works and why it matters

Chapter 7 develops the proposal about the argument-encouraging role of the norms of truth and falsity further, and begins its application to (reconstructed) non-factualist cases, via the phenomenon of No Fault Disagreement. §7.1 and §7.2 respond to two potential objections, both of them accusations of circularity in the proposed account. The first is that disagreement itself needs to be characterised in terms of truth and falsity, so cannot be 'prior' to these notions, in the way required. There are subtle issues here, which I took to connect *inter alia* to an understanding of negation. I soon discussed them at greater length in a paper called 'Why "Not"?' (Price 1990).

The second objection is that the worth or value of a belief is simply a matter of its truth or falsity. This threatens to reduce the proposal that the role of these notions is to help us to *improve* our beliefs to, at best, some sort of empty tautology. In effect, my response turns the well-known deficiencies of a success analysis of truth into a plus for my proposal. A belief's contribution to success is a very long term, messy, and holistic matter, very hard to identify with truth. But it is there all the same, to provide the basis of a story about long-run improvement in belief states, of the kind my proposal requires.

In §7.3 I develop the idea that disagreement contributes to long-run success, and note the importance of the fact that in many cases (though of course not all), social creatures like us are 'in the same boat' with our fellows: what works for one of us works for all. It is in these cases especially that the pooling of epistemic resources encouraged by argument is likely to be valuable. The story about truth thus rests on what I term the 'Same Boat Property' (SBP).

§7.4 makes some comparisons between my proposal about truth and the views of Brian Ellis and (very briefly) Jürgen Habermas, before §7.5 turns to some machinery needed to steer the chapter in the direction of No Fault Disagreement. Drawing on material in (Price 1983c), I introduce a distinction between *evidential* and *inferential* sources of disagreement, and note that in typical cases we take for granted that SBP applies to the evidential beliefs and rules of inference on which our judgements rely, as well as to the judgements themselves: 'In general, in other words, we argue as we should if ... the SBP applies as much to evidential beliefs and rules of inference as to the judgements that lead us into conflict.' This is important, because it will allow me to associate NFDs with exceptions to this general rule.

With this machinery in place, §7.6 and §7.7 turn to what I take to be a striking feature of probabilistic judgements (the SPAs of Chapter 4). It is the same fact on which I had relied in my

PhD thesis, in arguing that SPAs are not genuine factual claims: viz., that fact that when a disagreement about such probabilistic matters turns out to have arisen because the speakers concerned have access to different bodies of evidence, our sense that there is a substantial disagreement – that one side must be *at fault* – tends to dissipate. (I say that the dispute ‘evaporates’ in such a case.) I link this feature of probabilistic disputes to the fact, as I put it, that

[p]robabilistic judgement is a species of intrinsically ‘non-final’ judgement. Probability has its origins in situations in which we feel that final commitment is unwarranted. [Page ref to come from proofs]

Such judgements therefore contrast with the normal case, for which I draw on Gilbert Harman’s observation ‘that people find it difficult to accept things only tentatively .... There is a strong tendency quickly to convert such tentative acceptance into full acceptance’ (Harman 1986, p. 49).

§7.8 then responds to the important objection that this feature of probabilistic judgement might be explained by indexicality, introduced via an implicit reference to the evidence on which the judgement in question is based. I argue that this doesn’t capture the actual (fault-presumptive, but fault-cancellable) character of the relevant disagreements – a fact obscured by unresolved ambiguity in terms such as ‘available evidence’. (Again, this argument descends from my thesis.) Finally, §7.9 argues that rival ‘analytic’ views of truth have trouble accommodating No Fault Disagreements about SPAs, in virtue of the equivalence between P and ‘P is true’. When P is a SPA, this equivalence entails that disagreements about the two claims should behave in the same way, but rival views don’t seem to be able to permit this. (I suggest that again, this point may have been obscured by the mistaken assumption that indexicality is at work.)

As Lionel Shapiro (2014) has noted, the issues raised in this chapter have close connections to the issues discussed by the so-called ‘New Relativists’, in work stemming from that of MacFarlane (2003, 2005a, 2005b), among others. I shall return to this in Chapter 13.

### 0.3.7 Chapter 8 – From probability to everything else

Up to this point, FFT’s discussion of No Fault Disagreements, and the idea that they represent cases in which the sense of ‘factuality’ induced by the norms of truth and falsity reaches a limit, has rested on just one kind of example, that of probability judgements (SPAs). Chapter 8 introduces a much broader range of cases, including various kinds of evaluative judgements (§8.1), conditionals (§8.7), cases involving presupposition, and judgements about meaning (§8.9). Fulfilling a promise made earlier, it also returns to the non-indicatives (§8.11). And along the way, via the application to meaning, it makes a case for *global* non-factualism.

Woven into the chapter’s discussion of these cases are several other themes. For one thing, I seem to have been bothered by the possibility that what I was proposing as a substitute for conventional non-factualism – ‘analytic’ non-factualism, as I call it – would be thought to be excessively descriptive. As I put it, ‘the non-factualist traditionally begins with claims about the “real” status of moral judgements, and proceeds to prescriptions concerning the use of moral utterances’. For example, I imagined that the traditional non-factualist might

argue that moral and aesthetic truth is relative – that it is relative, as a matter of fact, and therefore should be treated as relative. It follows from this view that if a community treats moral truths as absolute, its members are mistaken about the nature of moral judgement. [Page ref to come from proofs]

I was concerned that my alternative form of non-factualism would make this impossible:

We have given up the idea that there is an independent standpoint, with respect to which there is anything more to being a statement of fact than being treated as such. We may now describe usage and hence explain it, but there is no longer a place for trying to analyse usage and hence prescribe it. [Page ref to come from proofs]

In §8.2, §8.3 and §8.6, these concerns lead to a discussion of the possibility that the descriptive–explanatory kind of non-factualism I am recommending need not be entirely ‘passive’ in its attitude to linguistic practice. In principle, the illumination we get about our own (present) linguistic practices from this perspective might allow, and recommend, that we seek to modify those practices.

Interspersed with these sections, in §8.4 and §8.5, are two brief sections raising challenges for my proposal, challenges that come into view after the proposed application to evaluative judgements. The first turns on the fact that ‘That’s true’ (said in response to a claim by a previous speaker) is itself an evaluative judgement, on my view. It is natural to wonder whether this doesn’t involve some sort of illicit boot-strapping. (I defer this challenge to Chapter 9.) The second seeks to revive the bifurcation thesis, by proposing that for judgements with world-to-mind direction of fit, truth can be equated with worldly correctness. As I say:

This is an important objection. If sustained, it would suggest that the present account is a form of the projectivist theory we discussed in Chapter 4. For it would seem to admit a distinction between those areas of discourse for which truth can be cashed in terms of a world-imposed notion of correctness, and those in which there is a useful function for a pretence of world-imposed correctness. [Page ref to come from proofs]

In response, I first emphasise what I take to be correct in the objection, in a passage that reads to me now like a manifesto for much of my later work. I say that the objection

is correct in noting a difference, on the proposed account, between the grounds for the application of the notions of truth and falsity in different areas of discourse .... Indeed, I think the explanatory approach should explicitly affirm this possibility. Once we detach the explanation of the use of truth from the search for an analytic base, we should see it as possible, even probable, that across the range of functions served by the assertoric form of language, the advantages of truth-encouraged argument will not be the same in all cases. Certainly, we should expect that these cases have something in common, to explain the fact that they all employ the assertoric form and the notion of truth. But this degree of uniformity is compatible with considerable diversity in its explanatory base. The advantages of argument may vary from case to case.<sup>9</sup> [Page ref to come from proofs]

But I go on to argue that in no case is there the simple equation between truth and some behavioural notion of appropriateness or success, on the prospect of which the objection implicitly relies.

To put the point in terminology I have proposed more recently (Price et al, 2013, Chapter 2, §5), I would say that objection confuses two notions of external constraint on belief and judgement – an ‘i-representational’ notion supplied by the norms of truth and falsity, that works uniformly

---

<sup>9</sup> How does this pluralism about the explanatory base relate to Wright’s well-known alethic pluralism? I discuss the relation of my view to Wright’s in Chapters 13 and 14 below, and at greater length in (Price 2022d).

across the assertoric language game as a whole, and an ‘e-representational’ notion applicable in those discourses whose function lies in part in tracking our physical environments. I now follow Sellars in saying that these two notions (or two corresponding notions of truth, in his discussion; see Sellars 1968, 136) ‘belong in different boxes’.

However, the theme from this chapter that plays the most pivotal role in my later work is the argument for *global* (explanatory) non-factualism, in §8.9. It turns on the underdetermination of meaning by the (finite) basis of evidence available to individual speakers. As I note, this idea plays a part in several well-known discussions in twentieth century philosophy: Quine’s arguments for the indeterminacy of meaning, Dummett’s case for verificationism, and (most saliently of all, for my purposes), Kripke’s discussion of following a rule. Drawing on the lessons of the last case, I argue that there is a kind of No Fault Disagreement to which, in principle, *any assertoric claim whatsoever* might be subject. It is the sort of ‘merely terminological’ disagreement that would arise if two speakers turned out to be using a term in different ways (having extrapolated differently from previous cases). As I put it:

The indeterminacy of meaning thus seems to be the basis for a form of universal non-factualism. That is, it offers a reason for departing from the pattern of usage characteristic of the statement form that may be applicable in any area of discourse. This topic-neutral form of non-factualism should not be confused with the form that Quine himself extracts from the indeterminacy thesis: the view that there are no genuine facts about *meaning*. (Cf. Quine 1960, p. 221.) Ignoring for a moment our scruples about the analytic viewpoint, we might say that the present view, in contrast, is that in virtue of the nature of meaning there can be no genuine facts about anything.

More correctly, the result suggests that the factuality of usage is always a matter of degree. Across the broadly fact-stating part of language – in English, roughly, the indicatives – the appropriateness of the statement form varies from case to case ... From the explanatory point of view there is nothing surprising about this conclusion. Only the analytic presumption leads us to expect a sharp distinction. [Page ref to come from proofs]

### 0.3.8 Chapter 9 – Wrapping up

The final chapter of FFT ties up some loose ends from earlier chapters, attempts to situate the position for which FFT has argued with respect to various other landmarks in the philosophical landscape, and closes with some remarks about the application of the explanatory approach to issues in the philosophy of mind and language more generally.

FFT begins, in a sense, with the conflict between orthodox Fregean semantics, especially as embodied in much Tarski- and Davidson-inspired work in the 1980s, and non-factualism. The former wants to apply a uniform model for the understanding of truth and meaning at least to all the indicative parts of language. The latter wants to identify important distinctions *within* the class of indicative judgements. And they both want to appeal to truth to do the job! (This tension is the central topic in Strawson and Evans (1973), the discussion I mentioned earlier.)

§9.1 describes FFT’s resolution of this tension. It turns on a distinction between the disquotational face of truth, which is what matters for the Fregean semantic programme, and the normative face, which is what matters for non-factualism, treated as FFT proposes.

This distinction between two faces of truth raises the question as to why the two faces attach to the same notion. §9.2 begins with this question, and uses its answer to tie up a concern left over

from Chapter 8, about whether there is some sort of circularity or vulnerability in the fact that truth ascriptions – statements of the form ‘It is true that P’ – are themselves the kind of thing that can be called true or false. It also returns to Wiggins’s ‘marks of truth’, discussed in Chapter 2, and explains why they should characterise truth if the notion has its origins where the account of FFT has proposed.

§9.3 also discusses the truth of truth ascriptions, but now with the aim of showing how FFT’s account avoids a problem facing orthodox non-factualists, on any conventional ‘analytic’ account of the fact-stating/non-fact-stating distinction:

[T]he analytic point of view leaves a non-factualist with the following dilemma. Either ‘P is true’ is sometimes a statement and sometimes not a statement, depending on the status of P itself; or all such truth ascriptions have the same status. Neither choice is particularly appealing. The former entails that truth ascriptions are not the semantically homogeneous class their form suggests. While the latter implies that ‘P is true’ can be a genuine statement even though P itself is not a statement, or *vice versa*; and hence that the equivalence principle fails. From the explanatory perspective, however, the dilemma looks entirely self-inflicted – a product of the assumption that any utterance lies on one side or other of a real semantic divide. [Page ref to come from proofs]

(The discussion links this point to an objection to projectivism raised by Arthur Fine.)

The next three sections, §§9.4–9.6, aim to distinguish the proposal of FFT from three other viewpoints, idealism, instrumentalism, and quietism. In the case of idealism, I begin by noting what I take to be some similarities:

As we saw in Chapter 4, the idealist takes over the projectivist’s goal of accounting for the realist features of linguistic practice but aims to do it quite generally. Our approach thus concurs with idealism in rejecting the projectivist’s bifurcation between real and projected facts, arguing that there is a single appropriate account of truth, applicable wherever truth is. Given that the idealist takes over the projectivist’s constructive programme, we also agree on an explanatory as opposed to an analytic approach to truth, and perhaps on some of the details. [Page ref to come from proofs]

How then does the proposed account differ from idealism? The answer I give is that idealism is too homogenous, missing the distinctions which are the residue of traditional non-factualism:

[T]he present approach concurs with the idealist in seeking to globalize the projectivist’s constructive approach to certain areas of reality – the attempt to ground our conception of the world ‘internally’, so to speak, in our own patterns of reaction to our environment. Unlike the idealist, however, we find room in this for a kind of non-factualism. [Page ref to come from proofs]

It is not now clear to me who, if anyone in particular, I took to embody the kind of idealism I had in mind here. But as I said in §0.3.3, I seem to have felt that it was a view which would globalise the apparent *anti*-realism of projectivism. Whereas my proposal, as I put it, ‘is best described as *functional* or *explanatory* realism. Its goal is to explain our realist practices; as well, of course, as their non-factualist variations.’

I was clearly concerned that this wouldn't be sufficient to gain me admission to the realist club (the philosophical equivalent of the cricket club, in Australia in the 1980s). The section finishes with some somewhat defensive remarks:

In expressing even this much sympathy for idealism I am likely to be accused of denying common sense intuitions about the reality of the external world. So I want to make it clear that my views about most aspects of the external world are the same as those of the most hard-nosed realist. We each hold true much the same class of propositions about how things are. Our professional differences concern at most a tiny fraction of such propositions: those that deal with the explanation of the fact that humans hold such views. For my part, I believe that at least in so far as holding a view about the world depends on our possession of the notion of truth or correct representation, the fact that we do so is explicable in the kind of terms here suggested. Hard-nosed realists may have different views on this topic, or none at all. But this doesn't prevent us agreeing about almost everything else (in other words, about almost everything). [Page ref to come from proofs]

Perhaps in the same spirit, the next section (§9.5) begins by emphasizing 'that functional realism is not a form of instrumentalism'.

Indeed, it is likely to reject instrumentalism on the same grounds as it rejects projectivism. Both these approaches ask us to make a distinction between two kinds of superficially factual discourse: the genuinely factual areas, on the one hand, and the instrumental or projected areas, on the other. The present theory denies that we can draw such a distinction. [Page ref to come from proofs]

However, I go on to argue that my view can avail itself of some of the advantages of instrumentalism, as identified by Arthur Fine.

§9.6 begins with Fine's 'natural ontological attitude' (NOA). As Fine says:

NOA allies itself with what Blackburn ... dismissively calls 'quietism'. Less pejoratively, NOA holds a 'no-theory' conception of truth .... NOA does not think that truth is an explanatory concept, or that there is some general thing that makes truths true. (1986, p. 175)

I side with Blackburn here, against Fine's quietism.

NOA is right, I think, in rejecting attempts to *analyse* truth. And it is right to dismiss recent attempts to make truth itself an explanatory notion. ... NOA is wrong, however, to conclude that there is no further role for a theory of truth. At the very least there remains the not inconsiderable task of accounting for the fact that we have such a notion at all.

The quietist is simply blind to this, a blindness perhaps explained by the tendency on all sides in these discussions to ignore non-descriptive uses of language. From this perspective we are liable to overlook the contrasts in language that draw attention to the remaining explanatory task. To notice the non-indicatives, and the issues raised by the non-factualists, is to notice not only that truth is not a universal notion in ordinary language, but also that there is considerable doubt as to the limits of its application. At this point, I think, to decline to theorize about truth would be an attitude to ontology

with all the naturalness of the ostrich's famed response to challenges of a different kind.  
[Page ref to come from proofs]

The reference to 'ontology' in the final sentence here now seems to me to be not quite right. The interesting questions about the functions of truth aren't ontological questions, but questions about human linguistic behaviour. As an objection to quietism, however, this still seems to me entirely correct. A related objection is that quietism is blind to important and interesting questions about the *differences* between our assertoric discourses. In this respect, too, I side with Blackburn's quasi-realist (against, for example, John McDowell's kind of quietism; see Price 2015a).

Finally, §9.7 offers some reflections on the explanatory approach to mind and language more generally. I mention Dennett's famous work on the intentional stance, and suggest that this can be seen as an example of the explanatory stance at work. But I suggest that the framework of FFT makes a difference to how a view such as Dennett's should be presented:

Dennett presents himself as an instrumentalist about folk psychology. On the present view, however, this reading of such a theory is unnecessary and unfounded. A sharp realist/instrumentalist distinction would call for a sharp factual/non-factual distinction; and that, we argued, is not to be had. This suggests that things are better than they seemed for an account of Dennett's kind. The explanatory perspective doesn't call for ontological apologies – or at least not such profuse ones, the difference between psychology and neuroscience being at most a matter of degree. [Page ref to come from proofs]

The closing point of the chapter is thus that the explanatory approach to truth and factuality turns out to have ramifications in other areas – and that the direction of attack matters.

It thus turns out to have been important to treat truth first. To accept an explanatory account of truth is to constrain one's understanding of any other theory in the same general spirit. Explanatory theories have in the past been crippled by their failure to apply their insights to truth itself (think of the problems of projectivism, for example). The promise of the explanatory approach lies in considerable measure in its ability to explain the terms in which the traditional issues have been phrased. [Page ref to come from proofs]

This, too, was an idea that I was to return to at various points in later work. In Price (2003, §6), for example – a later presentation of the kind of account of truth I had offered in FFT, and before that in Price (1983c) – I respond to the objection that view amounts to a kind of fictionalism about truth. Again, my point is that the proposed account of truth undercuts the terms of traditional metaphysical debates:

In common with other deflationary approaches to truth, the present account not only rejects the idea that there is a substantial metaphysical issue about truth (a substantial issue about the truthmakers of claims about truth, for example). Because it is about truth, it also positively prevents 're-inflation'. In other words, it seems to support a general deflationary attitude to issues of realism and antirealism. If so, then deflationism about truth is not only not to be equated with fictionalism, but tends to undermine the fictional–nonfictional distinction, as applied in the metaphysical realm. [Page ref to come from proofs]

## 0.4 Looking ahead

The next nine chapters are the original material of FFT, with very minor grammatical and typographical changes. The material new in this edition recommences in Chapter 10, at the beginning of Part III. Readers primarily interested in FFT's connections with current debates should feel free to skip ahead to that point, of course.

## 1

**Introduction**

## 1.1 WHY READ THIS BOOK?

Few areas of contemporary philosophy have not been touched by the suggestion that their distinctive problems are in part a product of some sort of misconception about language. One view of this kind has been particularly widespread. It is the claim that in virtue of their grammatical form, certain uses of language are mistakenly construed as descriptive, or fact-stating. Perhaps the most familiar cases arise in ethics. Moral philosophers have often argued that despite the syntactical resemblance that ethical utterances bear to genuinely descriptive uses of language, such expressions should not be interpreted as statements of fact. There are actually several versions of this view, differing as to what they take it that ethical judgements really are. One at least of these is common currency among non-philosophers. Known to philosophers as ‘emotivism’, this version holds that moral claims are ‘value judgements’ – expressions of attitudes of approval, disapproval, or something of the kind. The Prime Minister might say for example that the Soviets have the best possible system of government, with the exception of all the rest. An emotivist would interpret the remark as an expression of the PM’s own political preferences, of her marked antipathy to the Soviet state, rather than as a factual claim in comparative political theory.

We shall need a name for the general philosophical strategy that emotivism here exemplifies. I shall call it ‘non-factualism’. ‘Anti-realism’ would be a little less ugly, and is perhaps more widely used in this context. However, it is widely used for other programmes as well. ‘Non-cognitivism’ suffers from the opposite problem, being largely confined to the ethical case. A non-factualist thus holds that there are certain utterances whose grammatical form is that of statements of fact, but whose linguistic role is actually quite different.

Further philosophical consequences will usually be held to flow from this linguistic insight. If nothing else, it allows the non-factualist to dismiss previous philosophical concerns about the nature of the ‘facts’ in question. An emotivist will say for example that the traditional philosophical problem of the nature of good and evil rests on a misconception about the language of morality. As I said, ethics is by no means the only branch of philosophy whose practitioners have been attracted to non-factualism. Aesthetic judgements, mathematical statements, theoretical sentences in science, Lockean secondary qualities, knowledge claims, meaning ascriptions, indicative and subjunctive conditionals, and probabilistic and modal judgements: all of these topics have been given non-factualist treatments, and readers may well be able to add to the list.

This book is not about any of these topics as such. Its initial concern lies rather at the next level of generality, with the foundations and hence the viability of non-factualism itself. In particular it is concerned with the nature of the distinction between factual and non-factual uses of language. I am going to argue that the view of this distinction normally taken for granted by non-factualists is radically mistaken. Non-factualists have taken the issue to concern the distribution in language of a property that we may call ‘statementhood’. Their claim is then that this property is less extensive than we ordinarily assume. Certain classes of utterance – moral judgements, or whatever – appear to possess it but actually lack it. Once the issue is couched in these terms the important questions become ‘What does the property of statementhood consist in?’ and ‘How do we determine that a particular kind of discourse possesses or lacks the property so construed?’ In other words, non-factualism then calls for an *analysis* of statement of fact –

preferably, of course, an analysis that serves to support the view that there are fewer statements in language than we initially suppose.

The main task of the first part of the book will be to challenge this analytic perspective. I shall argue that of the available ways of cashing out the notion of statement of fact, none provides what the non-factualist evidently requires: a means to exclude the possibility that there are actually more statements than the non-factualist holds there to be. Indeed, we shall see that none of the available analyses can exclude the possibility that *all* linguistic utterances are fact-stating.

More in a moment on the structure of this argument. I want to make it clear first of all that the book is not against non-factualism as such. On the contrary, I am going to defend a version of non-factualism, albeit an unorthodox one, and argue that it vindicates the non-factualists' concern with language as a route to enlightenment on a range of philosophical topics. At the very least it vindicates the intuition that there is something unusual about the language of morality, probability, conditionals and a range of other topics with which non-factualists have been concerned.

In particular, therefore, the book is not aligned with the influential view that in virtue of some rather simplistic mistakes about language, non-factualism itself is misconceived. As I shall explain, this view stems particularly from the Fregean tradition in modern philosophy of language. Like non-factualism itself, this tradition is committed, by and large, to the existence in language of a distinction between fact-stating or assertoric discourse and discourse of other kinds. The conflict arises as to where this distinction should be drawn. A non-factualist wants to narrow the class of genuine statements. A Fregean, on the other hand, is inclined to do exactly the opposite – to construe the bounds of assertoric discourse as widely as possible. We shall see that this inclination is grounded particularly in the role that truth plays in Fregean linguistic theory. Its effect is that non-factualism, the epitome of philosophy in the linguistic style, comes ironically to be rejected by the philosophy of language itself.

In criticizing the usual presentation of non-factualism, how then do I fail to align myself with its Fregean critics? By rejecting the principle that non-factualists and Fregeans have in common, namely that there is a real property of statementhood, or some such, whose distribution in language can be a matter for dispute. As I said, I argue in part I that however we try to characterize this supposed property, we are left with no basis on which to show that it applies to some utterances but not to others. In particular, we cannot exclude the possibility that all utterances are statements: an admission that would be as unacceptable to (most) Fregeans as it would be to non-factualists.

I shall describe the conflict between non-factualism and Fregean semantics in more detail in the next two sections. Its effect, I think, has been to stunt the development of an adequate theory of the linguistic notions on which non-factualism depends. Both sides suffer as a result. For its part, non-factualism has not died out. It remains popular in traditional strongholds, such as ethics and aesthetics, and it continues to win converts elsewhere. As always, however, it lacks adequate conceptual foundations. On the other side, I think that contemporary philosophy of language has suffered from a failure to take seriously the non-factualist's concerns. The notions non-factualism appeals to – factuality, statement, truth, and so on – are central to an understanding of language. By focusing on some of the hard cases, non-factualism offers a particularly fruitful perspective on these notions. To dismiss non-factualism is to miss this perspective.

The book thus approaches the philosophy of language by way of linguistic philosophy. The needs of non-factualism, as viewed in the light of part I, lead in part II to a novel account of

truth. One of the advantages of this account is its ability to deal with these difficult cases, but its interest and application is more general. Its novelty, by contemporary standards, lies in the fact that it is explanatory rather than analytic. By this I mean that the question it addresses is not so much ‘What is truth?’ as ‘What use is the concept of truth, to creatures like us?’ In other words, it seeks to explain our possession of the notion of truth in terms of its function in a linguistic community. It is a genealogical theory. It asks why we should ever have developed such a notion.

I shall have more to say later on about the distinction between analytic and explanatory theories. In case naturalists in the audience are already tempted to leave, however, I emphasize now that the proposed account will be thoroughly naturalistic. For any concept in common use there are (at least) two possible naturalistic accounts of how it got there. One approach is to analyse the concept concerned in naturalistic terms, and hence explain its use in terms of some general account of our reaction to our natural environment. But another – often overlooked, I think – is to find some non-referential function for the concept in the lives of natural creatures such as ourselves. The latter approach is most clearly appropriate in cases in which we have reason to deny referents to the concepts concerned – religious concepts, or those of certain former scientific theories, for example. It is not excluded in other cases, however, and the proposed account of truth will be of this kind.

In view of the wide importance of the notion of truth in contemporary philosophy, this may seem a case in which a solution outgrows the problem from which it stems. Some readers will inevitably feel that in grounding such an account of truth on the needs of non-factualism, I use a small tail to wag a very large dog. Others, more favourably disposed to the general approach of part II, will feel that the dog is quite capable of wagging itself.

To the former objection I would reply that if the argument of part I goes through then it is not simply non-factualists who need a new approach to truth and related notions. Their Fregean opponents should be equally concerned by the failure of the usual ways of construing assertoric discourse. That said, non-factualism provides a valuable guide to and constraint on alternative approaches. One of the advantages of the account of truth proposed in part II is that it does make very good sense of non-factualist intuitions. Moreover, non-factualist concerns are a powerful antidote to another non-analytic approach to truth which has had significant support in recent years. This approach reacts to the perceived failure of attempts to analyse truth by dismissing the problem, recommending instead that we change the subject and stop worrying about truth. In drawing our attention to the limits of truth, non-factualism does much to counter this unhealthy temptation to let a sleeping dog lie.

I have more sympathy with the latter objection – the view that the proposed account of truth might be motivated without reference to the needs of non-factualism. I would simply encourage the reader whose interest is more in the dog than the tail to feel free to begin at part II. Chapter 6, sections 1 to 5 of chapter 7 and chapter 9 deal largely with the general account of truth. They make little reference to non-factualism and the concerns of part I, and might be read independently. The remainder of chapter 7 and chapter 8 apply the general account to the non-factualist cases, and might be read next. Part I would then recommend itself to the reader who found the orientation implausible, but was curious to see where I had gone astray. This approach to the book aside, I have starred a number of sections to indicate that they may be omitted without serious discontinuity.

Who then might profitably read this book? First, I think, anyone with an interest in non-factualism in any of the areas in which it has been applied. I imagine that most such readers will take it that they have a reasonably clear conception of what distinguishes fact-stating discourse,

and therefore as to why a non-factualist interpretation might be judged appropriate in any particular case. Part I aims to shake such confidence, and part II to show what might replace it. Secondly, I recommend the book to those with an active disinterest in non-factualism – those in particular who take its misconceptions to have been exposed by contemporary philosophy of language. Here, too, part I aims to undermine such prejudices. And thirdly and no doubt most numerous, I recommend it to anyone with an interest in truth. The account of truth here offered is novel and potentially important. The route will be unfamiliar to most readers, but I think the result will repay their efforts.

The remainder of this chapter is in four sections. The next, section 1.2, outlines the historical background to the conflict between non-factualism and contemporary philosophy of language. In section 1.3 I explain how the conflict can be thought of as a disagreement concerning the extension of the property of statementhood in language, and therefore how I propose to attack the common assumption that there is such a property. Section 1.4 is then a brief guide to the argument of part I, and section 1.5 some remarks about orientation of the alternative programme to be developed in part II.

## 1.2 NON-FACTUALISM AND FREGEAN SEMANTICS

With the benefit of hindsight, at any rate, it is clear that non-factualism predates the linguistic movement of the mid-twentieth century. In ethics it traces its origins at least to Hume, who may be regarded as an early emotivist. Hume seems to have taken a similar view of causal judgements: namely that they do not refer to a realm of causal facts, but rather express our expectations, formed on the basis of observed regularities. But the great popularity of non-factualism in this century – indeed the recognition that Hume should be seen in these terms – seems largely associated with the linguistic tum. In those halcyon days after the war the non-factualist techniques were applied to a wider and wider range of traditional philosophical topics.

In the 1960s, however, non-factualism acquired a reputation as a cheap approach to subtle philosophical problems: too quick to claim new territory, and too slow to justify its existing claims. The reputation is not entirely undeserved. The non-factualists of the 1950s do seem to have paid too little attention to the foundations of their own position. But it also owes something to the fact that the approach had the misfortune to coincide with what, from the non-factualist's point of view, was a particularly inhospitable period in the philosophical study of language itself. For non-factualism is at odds with a view that at this period philosophers of language had increasing reason to uphold – the view that assertion or statement of fact is the central and primary use of language, rather than one use among others of equal significance.

At first sight, indeed, a reader might gain the impression that there is no place for the notion of a non-fact-stating use of language in contemporary philosophy of language. Writers will refer, for example, to the task of specifying the truth conditions of 'the sentences of a language', as if there were no such thing as a sentence without truth conditions – as an utterance whose point is not to state that certain conditions obtain. As it stands, however, this may be said to reflect a benign idealization, not inconsistent with a more pluralistic view of natural language. The real basis of the doctrine of the primacy of assertion lies deeper, I think, in the Fregean use of the notion of truth.

Once truth is part of a philosopher's repertoire, there are at least two routes to the primacy of assertion. One turns on the fact that a systematic approach to the theory and taxonomy of language will need to distinguish the 'genuine' uses of languages, to which the theory should be expected to apply, from various sorts of pseudo-linguistic noise – for example the kind of noises

we make for dog rousing, baby settling and the like. A natural suggestion is that the genuine uses of language are the sentential utterances. (Thus Davidson 1984, p. 60: ‘In defining sentencehood what we capture, roughly, is the idea of an independently meaningful expression.’) In order to apply this suggestion we have to know what to count as a sentence. If the notion of truth is already at hand, it may suggest the following solution: a sentence is the bearer of a truth value. The significant utterances are those that are capable of being true or false: in other words (or so it seems), the assertions, or statements.

In practice, this route to the primacy of assertion is overshadowed by another. The notion of truth not only promises to delineate the raw material for a systematic theory of language; it also offers an approach to the problem of meaning – an approach developed since the 1960s in the influential work of Davidson and his followers. (See particularly Davidson 1984, essays 1 to 5; and Evans and McDowell (eds) 1976.) Davidson proposes that a systematic theory of meaning for a language – a theory that relates meaning to syntactic structure, in such a way as to show how the meanings of complex sentences depend on those of their constituents – should be a systematic theory of the truth conditions of the sentences of the language concerned. He suggests also that this truth theory be modelled on Tarski’s definition of truth for formal languages. Such a theory is bound to accord a central place to those linguistic expressions that can be said to have truth values (and hence truth conditions). At first sight, indeed, there seems to be room for nothing else.

It may seem then that whichever route it takes to the doctrine that assertion is the primary use of language, a Fregean theory of language should be embarrassed by the existence of utterances such as questions and commands. These would appear to be complete, independent, meaningful uses of language, and yet not to be assertions – to lack truth values, for example. However, there are at least two ways in which a Fregean theory might deal with these cases. One way, obviously, would be to claim that questions, commands and the like do have truth values after all. Questions and commands would thus be construed as special kinds of assertions – perhaps as reflexive assertions about their speakers. As we shall see in chapter 3, this option has been explored by some of the proponents of the truth-conditional approach to general semantics.

On the whole, however, advocates of Fregean general semantics have preferred a less radical solution, turning on the Fregean distinction between sense and force. Roughly, the idea is to confine the significance of the notion of a truth condition to the determination of a certain core factor in the meaning of an utterance (its *sense*); and to regard the complete meaning of an utterance as a product of this factor and another (its *force*). Thus, the meaning of the question ‘Has the dog been fed?’ will be explained as a product of the interrogative force (here indicated by the interrogative mood) and of a sense that the sentence shares with its assertoric transform, ‘The dog has been fed.’

This solution enables a Fregean theory of meaning to accommodate non-assertoric utterances. Indeed, in principle it allows for the assertoric mode of utterance to be treated as simply one among many. Each mode calls for an account of its peculiar *pragmatic* significance. The assertoric mode might be said to be distinguished by the convention that in using it, speakers aim to utter true sentences; the imperative, by the convention that speakers utter sentences they require their audience to make true; and so on. In principle, the list could be very long indeed.

In practice, however, the theorist has a strong incentive to peg the boundaries of the class of assertions as widely as possible, at the expense of non-assertoric varieties of force – to expand the semantic component of a theory of meaning at the expense of the pragmatic. Faced with a new linguistic expression, the theorist has to decide whether its contribution to the meaning of a

sentence consists in a modification of sense or an effect on force. The former answer has the advantage, for a Fregean, that it brings the case within the scope of the core component of the general theory of meaning. It avoids the need for a further addition to the supplementary theory of force. The latter answer, on the other hand, is a threat to the simplicity and comprehensiveness of the Fregean approach.

Most seriously, however, the latter answer would undermine the Fregean's claim to provide a systematic account of meaning. As Davidson emphasizes, the great advantage of the truth-theoretic approach is that it promises to account for the finite-based 'creativity' of language use – the ability of speakers to use and interpret a potentially infinite number of novel utterances. It does so by linking syntactic structure to conditions for truth. The significance of sentence-forming components is explained in terms of their contribution to the truth conditions of sentences in which they occur. On this view then, the open-endedness of language simply mirrors the open-endedness of possible truth conditions. This elegant solution would be crippled by the admission of a dimension of linguistic creativity not matched to a range of truth conditions. If sentences expressing moral judgements do not have truth values, for example, then the infinity of novel utterances we can generate from moral sentences (using the ordinary logical connectives, say) cannot be matched to a corresponding infinity of truth conditions; and our capacity to understand such sentences cannot be explained in terms of such a correspondence.

So if creativity is to be confined to the semantic part of a theory of meaning, non-assertoric expressions can be tolerated only on the condition that they do not proliferate. Like questions and commands, non-assertoric expressions must not significantly combine with each other (or indeed with assertoric sentences) to generate further such expressions. However, this condition clearly excludes the cases of interest to non-factualists: expressions that behave like fact-stating sentences, at least to the extent of participating in the usual modes of supra-sentential complexity. Truth-conditional general semantics thus has a powerful incentive to oppose non-factualism in general, and to dispute its claims in individual cases. The same applies, incidentally, if truth is replaced with a notion such as 'assertibility'; the resulting theory will have the same reason to confine creativity to what it regards as the non-pragmatic component of a theory of meaning.

As a distinctive product of the linguistic movement in philosophy, non-factualism was thus overwhelmed by developments in the philosophy of language itself: by the growing interest in the project of a general theory of meaning, and the growing popularity of a Fregean approach to that project. This conflict has not had the attention it deserves, I think. Proponents of truth-conditional general semantics, and indeed of other Fregean approaches to a general theory of meaning, have shown a degree of professional blindness to the attractions of non-factualist treatments of particular topics; while on their side, non-factualists often seem to have been unaware of the problem of reconciling their theories with a workable approach to general semantics.

There are signs of renewed interest in non-factualism, of a new respectability. At least one source of this interest lies within the philosophy of language itself: in recent interest in Wittgenstein's later views on meaning, and in particular in the suggestion (notably by Kripke 1982) that the notion of meaning itself might require a non-factualist treatment. However, it seems to me that without proper attention to its foundations, and to its place in a general semantic programme, the approach will prove no more durable at the end of the century than it was in the middle. This book, then, is in part an investigation of these foundations, in the light of the conflict just described.

### 1.3 THE NATURE OF STATEMENT: CONFLICTING CONSTRAINTS AND A SCEPTICAL PROGRAMME

As I noted earlier, some Fregeans think that all discourse is assertoric (and hence, in effect, that all language is fact-stating); but most prefer to distinguish assertions from utterances such as questions and commands. The more common view is thus that the fact-stating/non-fact-stating distinction roughly corresponds to the English indicative/non-indicative distinction.

A major advantage of this view has been thought to be that it gives us an account of grammatical mood. Mood is a striking feature of natural languages, whose existence we might expect a theory of language to be able to explain. It has been widely assumed that an adequate explanation would come in two parts. The first would identify some property characteristic of just those speech acts for which we do employ the indicative mood; while the second would consist in the claim that, by convention, the indicative mood is used to ‘mark’ the performance of a linguistic act with this particular property. This view of mood – the *marker view*, as I shall call it – thus takes the indicative mood to be the conventional sign of a certain property, or factor, in the meaning of an utterance. The following remarks are representative; they come from a recent text on mood by the linguist F. R. Palmer.

It is undoubtedly the case that most, perhaps all, languages have a clear way of indicating that the speaker is making a statement that he believes to be true. This is what may be called the Declarative – the grammatical form that is typically used for such statements ... In languages that have systems of mood, the indicative ... typically indicates a declarative. (Palmer 1986, pp. 26–7)

Where a language has an indicative and an imperative mood, these are the formal grammatical markers associated with the notions of statement and [command], though there is no exact one-to-one correspondence. (1986, p. 24)

As Palmer notes, the marker view needs to be qualified, in practice, to cope with various additional grammatical devices. In English, for example, stress and inflexion enable us to do many things with indicatives normally done with non-indicatives, and *vice versa*. However, a proponent of the marker view will take this to mean that in practice the marks are just more complicated than the initial account recognized. For present purposes we may ignore such refinements. As long as we don't criticize the marker view on the grounds that it misses these complexities, we won't go wrong in criticizing the simpler version.

In drawing the bounds of assertoric discourse at the limits of the indicative mood, most Fregeans thus align themselves with the marker view; and marker theorists will often characterize their position in terms of the Fregean force categories. The views thus have a common interest in the elucidation of these categories; in particular, in an account of assertion in terms of which the status of a given utterance can be established. For both views want to defend the claim that the non-indicatives are not assertions; and to be able to show, *pace* the non-factualist, that all indicatives are assertions.

The marker view will be the second front for the argument of the next four chapters. Our main target will be non-factualism, as usually presented. However, I want to show that the untenability of this position can be traced to its dependence on the analytic presupposition – on the view that there is a real property of statementhood, about whose distribution in language we are liable to be misled. I think that expressed in other terms, non-factualism contains important insights about language. So I shall need to exclude an alternative response to the argument – that of

rejecting non-factualism, while retaining its usual analytic footings. The marker view makes this response particularly appealing. In other words, we have to show that the failure of orthodox non-factualism is not attributable to the success of the marker view.

Fortunately, the same style of argument works on both fronts. In effect, indeed, we shall be able to encompass both targets in a single advance, arguing first that the available analyses do not enable the non-factualist to hold a line short of the marker view itself; and secondly that they do not enable the marker theorist to hold the line at the indicatives (and so to exclude the possibility that all discourse is factual discourse).

I emphasize that the conclusion we are aiming for is not that all discourse is factual discourse; merely that neither non-factualists nor marker theorists can exclude this possibility. The argument will thus be a sceptical critique of the analytic assumption these views have in common – a demonstration that in neither case can analysis serve its intended purpose, in justifying a particular division of language into fact-stating and non-fact-stating parts.

#### 1.4 GUIDE TO PART I

The structure of the next four chapters is as follows. In view of the importance of the notion of truth in general semantics and to the proposal of part II, chapter 2 treats non-factualists and marker theorists in tandem, examining the suggestion that assertions be distinguished in virtue of their connection with the properties of truth and falsity. Both sides have found this idea appealing. I argue, however, that without a prior understanding of the nature of assertion, we cannot restrict the notion of truth in such a way that not every utterance has the appropriate connection with these properties.

With truth out of the way, it is then convenient to separate our targets. Chapter 3 concentrates on the marker view, arguing by way of *reductio* that if there were an underlying property of the kind the view requires – a property marked by the indicative mood – we could not exclude its possession by non-indicatives. This shows, in effect, that the line cannot be held at the limits of the indicative mood. If all indicatives are fact-stating, then there is no way to exclude the possibility that non-indicatives are also fact-stating. The remaining task is then to show that the line cannot be held within the indicative mood – that of the various possible accounts of the nature of statement, none suffices to show that some indicatives are not statements of fact.

In chapters 4 and 5, with this end in mind, we take the side of the non-factualist's opponent in particular cases. Chapter 4 argues that in two important and arguably representative cases (probability and moral judgement) such an opponent forces the non-factualist to rely on a psychological characterization of statement of fact – on the claim, essentially, that statement is the expression of belief. Chapter 5 then examines some possible bases for the required distinction between beliefs and other propositional attitudes. We find that the attempt to make good this distinction leads the non-factualist once more to appeal to truth – in effect, to the proposal rejected in chapter 2. The chapter concludes with a look at some recent arguments that seem to favour non-factualism about conditionals – confirming that here, too, non-factualism lacks the basis it requires.

#### 1.5 THE EXPLANATORY TURN

One option at this point would be to blame the problem for the failure of the analytic solution – to conclude that there is simply nothing of philosophical interest to be said about a distinction in language of the kind non-factualism requires. Such 'quietism' has an evident appeal: to dismiss a

problem is to save oneself the trouble of finding a solution. It is the relief the lapsed theist might feel in being able to put theological problems aside. But of course, this advantage is like the claimed psychological benefits of religious faith itself. It is a motive for wanting a particular conclusion to be true, but not a reason for believing it so. In the present case there seem to be two good reasons to remain at least open-minded. Quietism cannot explain the significance of mood, or the continuing attraction of non-factualist approaches to particular philosophical topics.

Part II therefore explores a different response. This concentrates on the assumption that non-factualism needs to be couched as a claim concerning the distribution of a certain property in language. I think that the prevalence of this assumption among non-factualists is particularly a product of their native tendency to set aside linguistic appearances. Non-factualists want to say that appearances are deceptive, that assertions cannot readily be distinguished in virtue of their syntactical form. This is needed to explain how more traditional philosophical theories can have been led astray. And there must be something in the picture for these theories to have been wrong about – in other words, it seems, some underlying property that syntax can mislead us about. Ask what property this is, and one is committed to what I have called the analytic programme.

The solution, I think, is to recognize that there is room for deceptiveness at the syntactical level itself. An expression that behaves as an assertion in most respects and circumstances might nonetheless behave differently in minor respects or unusual circumstances. If the differences are then overshadowed by the likenesses, it is easy to see how interpreters might be led astray. With the deceptiveness of linguistic appearances explained in these terms, non-factualism doesn't need to appeal to linguistic categories underlying usage. So the non-factualist no longer owes us an analysis of statementhood. Non-factualism simply doesn't need to appeal to any such property. True, it ought to explain the syntactical anomalies to which it does now appeal – the features of usage that distinguish moral judgements, probabilistic claims, or whatever. As we shall see, however, these anomalies can be explained without being taken to reflect any underlying semantic distinction.

What are these anomalies in usage, and what is the standard pattern from which they depart? We shall see that both involve the notions of truth and falsity. Dealing first with the ordinary use of these notions, I shall propose that we explain it in functional and evolutionary terms. The main idea will be that in virtue of the normativity of truth and falsity, speakers are encouraged to resolve disagreements. This has long term survival advantages. On the whole, it improves the behavioural commitments with which language users meet the world. As I said earlier, this is not intended to be an analysis of truth. It is a genealogical theory, an explanation as to why a language community might come to possess such a notion. It is comparable for example to a biological view of the origins of morality – the view say that human societies tend to evolve such a value system in virtue of the communal advantages that flow from its regulative effects.

Returning to non-factualism, we shall see that this account of truth leaves room for certain borderline cases – cases in which the use of truth and falsity departs from the standard pattern. These variations turn out to be explicable in similar terms to the standard pattern itself. Moreover, they appear to correlate well with the cases in which non-factualism has seemed attractive. The explanatory perspective thus provides a sympathetic reconstruction of non-factualist intuitions – while vindicating, in a sense, the popular conception that truth is at the basis of the fact-stating/non-fact-stating distinction. It also finds a place for the non-indicatives.

It will be necessary to defend this account against the charge that it merely offers another analysis of the nature of statement. An obvious source of such a charge will be the fact that the account professes to explain variations in a pattern of usage across language. Hence it needs apparently to appeal to some more basic differences between utterances, or possible uses of language. Could not these differences themselves be made the basis of a characterization of fact-stating discourse? The reply to this will turn particularly on holistic considerations – the crucial point being that the relevant underlying difference need not show up in the individual case.

We began with the conflict between non-factualism and the Fregean programme for general semantics over the scope of the fact-stating/non-fact-stating distinction. In the final chapter I shall indicate how this conflict is resolved by the proposed treatment of truth. I shall locate the proposed account in relation to some recent discussions of realism and objectivity. And I shall finish with some comments on the significance of the explanatory perspective in other areas – particularly its relevance to the problem of the status of folk psychology.

**PART I**

**A Sceptic's Guide to the Matter of Fact**

## 2

**The Place of Truth**

Truth has often seemed to be the key to an analysis of fact-stating discourse. Marker theorists and non-factualists have both been tempted by the idea that an utterance is genuinely assertoric or fact-stating just in case it stands in a particular relation to the properties of truth and falsity – for example, just in case it bears a truth value, or is capable of doing so. In this chapter I want to show that truth cannot be invoked in this way.

The argument will turn on the demands placed on truth by the requirement that it provide the base property for an *analysis* of the notion of statement of fact. I cannot emphasize too strongly that I am not suggesting that truth has no role to play in an adequate account of fact-stating discourse. On the contrary, truth will be central to the theory to be developed in part II. But that theory will not be an analysis of statement of fact. Instead it will offer a non-reductive explanation of the characteristic marks in usage of the distinction between fact-stating and non-fact-stating discourse – including, perhaps most characteristically, the use of the notions of truth and falsity.

The present chapter goes as follows. I begin by clarifying the demands on the notion of truth that flow from its attempted use by marker theorists and non-factualists to elucidate the notion of assertion. With reference to familiar theories of truth, I then illustrate some of the ways in which a theory of truth can fail to meet these demands (fail to provide a *substantial* theory of truth, as we shall say). I offer some general arguments against the possibility of a substantial theory of truth, and conclude by examining a specific proposal: the approach suggested by David Wiggins in a paper called ‘What would be a substantial theory of truth?’ (1980). Wiggins’s approach – I think the most explicit attempt in the literature to tailor a theory of truth to these purposes – has several features of interest, from our present point of view. For one thing it is grounded in the truth-theoretic approach to general semantics, whose constitutional aversion to non-factualism we noted in chapter 1. Wiggins is aware of this tension, observing that the notion of truth he derives from truth-theoretic considerations – a notion applicable to the indicatives as a whole – leaves little scope for the more restricted notion the non-factualist requires. In showing that such a notion of truth cannot in fact be counted substantial, we shall defend non-factualism against this objection and clear the ground for its promised reconciliation with general semantics. For another thing, Wiggins’s theory sets useful targets for any alternative account of truth. Wiggins identifies properties or ‘marks’ that are plausible characteristics of any satisfactory account of truth. We shall see that his own derivations of these marks are illicit, given the intended role of a substantial theory of truth. But the marks themselves will set a standard for our own explanatory theory of truth. We shall want to be able to explain why truth has these particular characteristics.

**2.1 TWO WAYS OF APPEALING TO TRUTH**

In attempting to characterize assertion in terms of truth, marker theorists and non-factualists must both aim for more than a merely descriptive account of a distinction found in usage. Neither side can simply observe that in practice we apply the notions of truth and falsity selectively – that only certain kinds of utterance are treated as capable of possessing these properties. In the marker theorist’s case, this is because the marker theory itself is an attempt to explain a distinction we observe in usage. It may be illuminating to describe this distinction further – to note that in practice the use of the indicative mood correlates with that of ‘true’,

‘false’ and related terms. But to gloss the problem in this way is not to provide a solution. The ordinary use of truth is as much in need of explanation as the indicative mood itself.

In the non-factualist’s case a merely descriptive characterization would be inadequate for a different reason. The non-factualist’s interest in usage is not so much to explain it as to escape it. Non-factualism urges us to draw distinctions that are missed by ordinary usage. Its characteristic claim is that we are misled by usage into mistaking the real nature of a certain kind of discourse (typically with philosophical consequences). True, the non-factualist might say that we are misled only by insufficient attention to usage – that if we concentrate on the details, we do find evidence of the distinctions in question. But evidence of what? Even in this case the non-factualist assumes that the usage distinctions in question reflect the deeper distinctions on which the philosophical moral depends. The point is made by Michael Dummett. In his (1959) paper on truth Dummett considers the suggestion that an appeal to the use of ‘true’ and ‘false’ can settle non-factualist claims, for example about the status of ethical statements. ‘At one time’, he says,

it was usual to say that we do not call ethical statements ‘true’ or ‘false’, and from this many consequences for ethics were held to flow. But the question is not whether these words are in practice applied to ethical statements, but whether, if they were so applied, the point of doing so would be the same as the point of applying them to statements of other kinds, and, if not, in what ways it would be different. (1959, p. 144)

Non-factualists and marker theorists should therefore not be construed, in trying to characterize assertion in terms of truth, as merely labelling a distinction they claim to find in ordinary usage. Their appeal is not to the ordinary use of ‘true’, ‘false’ and related terms, but to the properties of truth and falsity. Their common claim is that an assertion (or statement of fact, or whatever) is an utterance that stands to these properties in some distinctive way.

At this point the possible views begin to diverge, in two senses. First, the non-factualist and the marker theorist clearly apply this characterization in different ways. The marker theorist takes standing in the specified way to truth and falsity to be the property conventionally marked by the indicative mood (and as we shall see, offers parallel accounts of the other moods); whereas the non-factualist claims that not all indicatives stand in this way to truth and falsity. Secondly, there are a number of views of the precise role of the properties of truth and falsity in a characterization of this general kind. Is it that assertions *have* these properties, that they are *capable* of having them, that they are governed by the convention that we *try to ensure* that they have one of them (i.e. truth), or some yet further connection?

These various elaborations have this much in common, however: they all depend on truth and falsity, and hence on the availability, at least in principle, of some account of the nature of these properties. The point is a trivial one. Unless the property of truth is subject to some theoretical constraints there is nothing to exclude its application not merely to the *utterances* the non-factualist or marker theorist wants to regard as non-fact-stating, but to any object whatsoever. Without a theory of truth, it is not simply that we cannot exclude the possibility that imperatives or ethical judgements have truth values (or stand in some other important relation to truth); we cannot exclude the corresponding possibility for trees, rocks and clusters of galaxies. (It is no good appealing to usage. We have seen that non-factualists and marker theorists both need more than that.)

Of course, not just any theory of truth will do. Again, one only needs to consider some trivial possibilities. Let us say that if a theory of truth can be invoked to characterize fact-stating

discourse in the way that either non-factualists or marker theorists require, then it is a *substantial* theory of truth. (I think the term was first used in this way by Bernard Williams 1973, p. 202). I want to show that there can be no such theory.

## 2.2 TWO WAYS FOR TRUTH TO BE INSUBSTANTIAL

One of the most explicit versions of a truth-based formulation of the marker view is that of Stenius, who says that ‘we may ... formulate the rule for the indicative in the following way: *Produce a sentence in the indicative mood only if its sentence-radical is true* (1967, pp. 267–8). By ‘sentence-radical’ Stenius, following Wittgenstein, means a sentence considered as devoid of assertoric force. Thus:

Consider the following sentences:

- (1) You live here now.
- (2) Live (you) here now!
- (3) Do you live here now?

These three sentences have something in common, which ... I call the *sentence-radical*; what is different in each of them I call the *modal element*. The sentence-radical signifies the *descriptive content* of the sentence, the modal element signifies its *mood*. (1967, p. 254)

Stenius suggests the following rules for the use of the imperative and interrogative moods: ‘React to a sentence in the imperative mood by making the sentence-radical true’, and ‘Answer the question by “yes” or “no”, according as its sentence-radical is true or false.’

Let us focus on the possible justification for a theory of this kind. If we want to appeal to a semantic property to explain the significance of the indicative mood, we need to be sure that we can tell whether an utterance has that property, without appealing to its grammatical mood. A theory such as Stenius’s cannot simply take for granted that certain sentence-forming expressions are mood indicators, and others are modifiers of descriptive content. Rather it needs to justify the particular dividing line it claims to find in language.

This point is liable to be overlooked, if – in the grip of the orthodox view – we fail to notice that there are alternatives. But consider negation, for example. Why not think of ‘It is not the case that ...’ as a mood indicator? It will be governed, of course, by this rule: Produce a sentence in the negative mood only if its sentence radical is *false*. Or on the other side, why not think of the imperative as a content modifier (so that imperative utterances have assertoric force)? We cannot rule out these possibilities on the grounds that negatives but not imperatives are indicatives, if what we want to explain is the significance of the indicative construction itself.

These considerations place the following constraint on a substantial theory of truth. It must not take for granted that only indicative utterances have truth values, or stand in the requisite relation to the properties of truth and falsity. Yet at the same time it must exclude the extension of these properties to non-indicatives. (The non-factualist will want the bounds to be narrower still.) We shall see that there is a tendency to avoid the latter mistake at the cost of the former.

The second major constraint on a substantial theory of truth is merely an instance of a general requirement for acceptable analysis: an account of a base property cannot itself depend on the very notion that property is intended to elucidate. A substantial theory of truth cannot rely on the notion of assertion, for example. In particular, it cannot rely on the otherwise appealing

intuition that a sentence is true just in case it would be correct to assert it. For that would reduce the supposedly distinctive claim that assertions are governed by the convention that speakers aim to speak the truth, to the truism that assertions are governed by the convention that speakers aim to make them correctly. Any convention-governed activity is governed by that convention.

Stenius is bothered by this sort of point, and suggests that

there are two concepts of truth: (i) the concept of truth ... which refers to the sentence-radical – I shall call this kind of truth *descriptive truth*; (ii) a concept of truth referring to a sentence in the indicative – I shall call this kind of truth *modal truth*.

When stating that somebody ‘speaks the truth’ one does not mean that he is producing a true sentence-radical – for one does not say that somebody is speaking the truth if he presents a true sentence-radical in the imperative or interrogative mood .... The expression ‘speaking the truth’ thus refers to the modal concept of truth, and means that one is following the rule for the indicative correctly. This moral rule is often motivated by the argument that by not following it one is undermining the communicative function of language, which morals must safeguard. (1967, p. 268)

As it happens, of course, to use the indicative correctly is to utter a sentence radical which is true in the descriptive sense; and so the two kinds of truth seem co-extensive. On Stenius’s view this seems to be a kind of accident, stemming from the rule which happens to govern the indicative mood. But there is a danger here: the more the modal notion of truth is detached from the actual use of the indicative mood, the more it needs to be explained why the modal notion should not extend to other moods. Why, for example, should we not say that someone ‘acts the truth’ when he or she obeys a command, thus following the rule for the imperative correctly? The modal notion of truth thus floats free of the indicative mood, while we are still in need of a substantial theory of the descriptive notion of truth.

In summary then, there are two main ways in which a theory of truth can fail to be substantial. It can ‘get the grammar wrong’, either illicitly assuming that only indicatives stand appropriately to truth or failing to exclude the possibility that non-indicatives so stand. Or it can appeal to the sorts of notions it is supposed to elucidate (or, not quite such a serious sin, to notions that would do a better job by themselves). We shall see that the first disability is endemic to the proposal to analyse assertion in terms of truth, though it is liable to be disguised by the second.

It is easy to find both kinds of insubstantiality in popular accounts of truth. This is not an objection to the accounts concerned, of course. It simply means that they cannot be applied, as in general they were not intended to be applied, in the special role of a substantial theory.

Thus traditional correspondence and coherence accounts of truth are likely to embody the second form of insubstantiality. To explain truth in terms of ‘correspondence to the facts’ is to take for granted the very notion we are supposing that the non-factualist wants to elucidate (and so restrict) by means of a substantial theory of truth. For example, it makes the issue as to whether there are genuine moral truths – whether moral claims can properly be said to be true or false – depend on that of whether there are moral facts. If we could already settle the latter issue, we wouldn’t need to characterize fact-stating discourse in terms of truth. Similarly, it is unlikely to be helpful to try to explain truth in terms of coherence with a system of beliefs. In the moral case, that leads us back to the question whether moral attitudes are genuine beliefs. But given the notion of belief, we would be able to use it to characterize assertion (in terms of expression of belief), and the excursion through truth would be redundant.

The first kind of insubstantiality is likely to be displayed by any account of truth that trades on the equivalence property of truth: on the fact that for any ('appropriate') sentence P, the sentences P and 'P is true' are in some sense equivalent in meaning. The most familiar such accounts are the 'redundancy' and 'disquotational' theories of truth (Ramsey 1978, chapter 2; Quine, 1970, chapter 1), for which this property characterizes truth. A less familiar case is Strawson's (1949) 'performative' account of truth. This turns on the application we make of the equivalence property – on the fact that by saying 'That's true' we can endorse an utterance made by another speaker. The novelty of the approach lies in taking this endorsing or confirmatory speech function to be the primary characteristic of truth; or as the view would think more accurate, of the use of the term 'true'. (This view is also known, rather aptly, as the 'amen' theory of truth.)

The insubstantiality of such theories stems from their reliance on the assumption that the equivalence property involves a restriction of truth to indicative sentences. It is not in fact the case that for *any* sentence P, P and 'P is true' are equivalent in meaning. Manifestly, 'Feed the dog' does not mean the same as "'Feed the dog' is true.' Hence the restriction to 'appropriate' sentences. But which are the appropriate sentences? Why, those that grammar permits us to substitute for P in the construction 'It is true that P'; in other words (for our purposes), the indicatives. So the equivalence property distinguishes the indicatives only by fiat. A theory of truth which rests on the equivalence property either fails to pick out the indicatives (or any subclass thereof) at all, or does so only because the restriction of truth to such a class of utterances is assumed from the start. But if the latter, then we can again invoke the sceptical challenge: Why this restriction? Such a notion of truth does not therefore provide the base property that is supposed to be characteristic of assertion and marked by the indicative mood.

The fact that Strawson's theory is insubstantial in this way is noted by Bernard Williams, in the paper in which he introduces the idea of a substantial theory of truth. Williams says that he doesn't see 'how on such a theory it could be more than an accident of language that "is true" signified agreement with *assertions* rather than agreement with anything else' (1973, p. 203). In practice we cannot endorse an apt question or an appropriate command by calling it true. Of course, we can do it in less direct ways, but so we could in all the cases in which we can just say 'That's true.' The amen theory does not explain why we should be allowed the conversational convenience in some cases but not in others.

The above examples suggest that a substantial theory of truth would be an unusual theory, by existing standards. The common approaches seem to be insubstantial in one or both of the ways we identified. Could there be such a theory? I think we can show that there could not.

### 2.3 TWO ARGUMENTS THAT TRUTH CANNOT BE SUBSTANTIAL

The first argument generalizes the sceptical challenge. Marker theorists and non-factualists both claim to find a division in language between fact-stating and non-fact-stating uses. A substantial theory of truth would permit a particular account of this distinction. It would allow the theorist to say that the fact-stating uses are distinctive in standing in a certain (yet to be specified) relationship to the properties of truth and falsity. Given such a proposal, the sceptic challenges the theorist's warrant for assigning any given utterance to one side of the division or other.

For simplicity let us focus on the claim that marker theorists and non-factualists will have in common, that non-indicatives do not stand in the designated way to truth. How could such a claim be justified? Consider an example. On what grounds do we say that sentences such as those in this paragraph do not have the property supposedly marked by (or confined to some

subclass of) the indicatives? In other words, how do we know that these sentences are not fact-stating?

Ordinarily we might appeal to their mood, but what is now at issue is our right to the kind of theory that would allow such an appeal. Perhaps then we should appeal to the intuition that in uttering such sentences we do not express beliefs, make any claims about the world, and the like. However, to depend on claims of this sort is to render our theory of truth insubstantial in the second of the two ways we distinguished earlier. A substantial theory of truth cannot rely on alternative characterizations of fact-stating discourse.

The last alternative seems to be to appeal to ordinary usage – to the fact that the sentences concerned are not ordinarily ascribed truth values. We have noted, however, that in distinguishing fact-stating from non-fact-stating discourse, marker theorists and non-factualists both want more than a merely descriptive characterization of ordinary usage. Among other things a substantial theory of truth is supposed to provide an analytically grounded explanation of usage. But it seems that there can be no evidence for such a theory, other than the merely linguistic phenomena it is intended to explain.

It might be objected that this is a common predicament for scientific theory. It seems that the grounds for acceptance of a theory often consist simply in the fact that it offers the best available explanation of a range of phenomena. Such cases lack a crucial feature of the present linguistic case, however. In trying to account for distinctions in usage in terms of the fact-stating/non-fact-stating distinction, marker theorists and non-factualists both seem to have regarded the relevant features of usage as responsive to an underlying distinction. The marker theory, in particular, seems to involve that the choice of grammar be guided by at least an implicit recognition of underlying linguistic categories. The sceptical argument challenges the possibility of such recognition (given, in the present case, that these categories are supposed to rest on a substantial theory of truth). There is no parallel to this recognitional requirement in the scientific cases, and hence no easy answer to the sceptic by analogy with those cases.

The role of implicit recognition of the underlying categories is well illustrated by Stenius's version of a truth-based marker theory. We saw earlier that if assertions are to be characterized in terms of the convention that they aim at truth, truth must be considered a property of the members of a class of unasserted sentences: Stenius's sentence radicals. Making an assertion is then to be thought of as uttering a sentence radical, while indicating by means of its (indicative) mood that one takes it to be true. There are parallel conventions for the other moods. However, such conventions require that speakers be able to distinguish the sentence radicals in advance of the choice of mood. The sceptic argues that this is impossible (at least in terms of a substantial theory of truth). Our only guide to the limits of the class of sentence radicals is the indicative mood itself. So the convention that supposedly governs that mood turns out to presuppose an understanding of the significance of mood.

The second and perhaps more familiar argument against the possibility of a substantial theory of truth also turns on the fact that such a theory makes truth conceptually prior to assertion. If fact-stating constructions are characterized as sentences that bear a certain relationship to the properties of truth and falsity then the distinction between truth and falsity comes into question. For example, couldn't we imagine a group of speakers who followed the convention that they aimed to utter *false* sentences. This convention seems no less useful than the ordinary one for purposes of communication. If one of these speakers says 'I've made the tea', then knowing the convention we know what he means. He means that he has not made the tea. However, what he *said* was supposed to be false. (We were imagining that he follows the convention of trying to

utter false sentences.) But the truth value of a sentence depends on what it means. If he meant that he had not made the tea, then if we agree we should say that the sentence he uttered was true, not false. ‘For a sentence is true if things are as we say they are by means of it; and if by “*p*” we mean “ $\sim p$ ”, and things are as we mean them, then “*p*” is on the new interpretation true and not false.’ The sentence comes from the *Tractatus* (4.062). Dummett (1973, p. 318) refers to it in support of the contention that assertion cannot be analysed in terms of truth. Indeed, in his (1959) paper on truth Dummett criticizes ‘Frege’s theory of truth and falsity as the references of sentences’ on the grounds that it ‘leaves ... quite out of account’ that

it is part of the concept of *truth* that we aim at making true *statements* .... Frege indeed tried to bring it in afterwards, in his theory of assertion – but too late; for the sense of the sentence is not given in advance of our going in for the activity of asserting, since otherwise there could be people who expressed the same thoughts but went in instead for denying them. (1959, p. 143, my italics)

The two arguments of this section show the absurdity of Stenius’s ‘descriptive’ notion of truth – of the idea that sentences could be considered to have truth values (or indeed meanings) in advance of an understanding of their conventional linguistic role. On the contrary: a sentence can be taken to have a significant truth value only in virtue of the understanding that its conventional use is such as to admit a standard of right and wrong, of correctness and incorrectness. (This is a minimum requirement; presumably not just any standard will do.) The true sentences are those whose utterance would be correct, according to this standard. In other words (now taking for granted what we are trying to explain), a true sentence is true in virtue of the fact that it would be correct to assert it – which reduces the claim that in making assertions we aim to speak the truth, to the truism that in making assertions, we aim to do so correctly.

Dummett criticizes Stenius in similar terms, but seems to think that the point concerns only the distinction between truth and falsity:

Stenius insists that the ‘sentence-radical’... itself has a ‘*directed*’ sense, independently of the assertion which can be expressed by attaching the assertion sign to it. Admittedly, the ‘sentence-radical’ has ‘true/false poles’, in that its sense consists in the distinction between conditions under which it is true and conditions under which it is false: but we can determine the ‘direction’ of this sense ... only by considering its use when assertoric force is attached to it, because it is only by reference to its role in assertions that we can identify one set of conditions as those in which it is *true* and the other as those in which it is *false*. (1973, p. 327)

However, the sense of a sentence radical seems equally dependent on its role in assertion. Dummett’s own views on truth and sense suggest as much: if ‘the notion of truth has first to be explained’ in terms of ‘a distinction between correct and incorrect assertions’ (1978, p. 20), and ‘sense consists in a distinction between conditions under which [a sentence radical] is true and conditions under which it is false’ (1973, p. 327), then a grasp of sense already depends on a conception of potential assertoric use.

#### 2.4 TRUTH THEORIES: INSUBSTANTIAL THEORIES OF TRUTH

Let us turn now to the promised example: David Wiggins’s (1980) attempt to construct a substantial theory of truth. ‘What would be a substantial theory of truth?’ is Wiggins’s contribution to a *Festschrift* for P. F. Strawson. He begins, appropriately, with some remarks about Strawson’s views on truth. In particular, he claims (1980, pp. 190–1) to find in Strawson’s

(1950) article on truth the view that ‘of proper concern to the theory of truth ‘ is ‘the task of elucidating the nature of a certain type of communication’: the ‘empirically informative’ or ‘fact-stating’ use. He thus counts Strawson as an early advocate of the need for what Williams was to call a substantial theory of truth: ‘I do not doubt that what [Williams] felt he needed was the very same thing that Strawson had envisaged in 1950’ (Wiggins 1980, p. 192).

Strawson does refer to the task just mentioned. However, when he asks ‘why should the problem of Truth ... be seen as this problem of elucidating the fact-stating type of discourse?’ his ‘answer is that it shouldn’t be .... The problem about the use of “true” is to see how this word fits into [the fact-stating] frame of discourse. The surest route to the wrong answer is to confuse this problem with the question: What type of discourse is this?’ (1950, pp. 142–3). At that time, at any rate (there is no hint of it in his reply to Wiggins in van Straaten 1980, pp. 294–6), Strawson’s view thus seems to have been that the problem of truth arises only *within* the fact-stating form of discourse, and hence depends on, or assumes, an understanding of that form of discourse.

Wiggins also claims Dummett as an early advocate of Williams’s project, quoting from Dummett’s (1959) paper on truth. But we have seen that in that paper and elsewhere Dummett rejects the idea that assertion can be explained in terms of truth. He again emphasizes the point in a postscript to the (1958) paper, first published in 1972:

What has to be added to a truth definition for the sentences of a language, if the notion of truth is to be explained, is a description of the linguistic activity of making assertions; and this is a task of enormous complexity. What we can say is that any such account of what assertion is must introduce a distinction between correct and incorrect assertions, and that it is in terms of that distinction that the notion of truth has first to be explained. (Dummett 1978, p. 20)

We shall see in chapter 3 that Dummett prefers a quite different account of the distinctive features of assertion.

History aside, Wiggins’s approach is to consider what must be true of truth if the predicate ‘is true’ is to figure in a truth theory adequate for radical interpretation. More precisely, the idea is to look at theories in the style of a Tarski truth theory for a language L, yielding equivalences of the form

(2.1) S is  $\varphi$  if and only if P.

Such a theory is to be regarded as specifying the meaning of the L-sentence S. There are many such theories, but some of them make better sense than others of the speakers of L. The constraints on  $\varphi$  imposed by the anthropological constraint of ‘making sense’ are the marks of the notion of truth.

I am going to argue that the distinction that a substantial theory of truth is supposed to help us with – between assertoric or fact-stating uses of language, and all the rest – is already built into the truth-theoretic approach to meaning, as Wiggins employs it. Wiggins gets out only what Davidson and others have already built in. The project of radical interpretation, even when constrained by the structural insights we have from Tarski, does not in itself yield either the Fregean distinction between sense and force, or (what would come to the same thing) a substantial theory of truth. These have to come from somewhere else.

Wiggins's strategy does not consist simply in a Tarskian definition of truth. Tarski's approach takes for granted the notion of translation. His 'condition of adequacy' is that a truth theory should yield sentences of the form

(2.2) S is true if and only if P

where P is a translation in the metalanguage of the object language sentence to which the description S refers. Whereas for Davidson and the truth theorists who follow him, the priorities are in a sense reversed. The T-sentences are now expected to yield interpretations of the sentences of the object language, and the theory is constrained, roughly speaking, by the requirement that it make good anthropological sense of the speakers of the object language. Moreover, it is not crucial to this programme (as it was to Tarski) that the predicate the theory invokes be the predicate 'is true'. John McDowell has repeatedly emphasized this point. McDowell describes what he calls 'the best version of the truth-conditional proposal ... along these lines':

We may reasonably set ourselves the ideal of constructing, as a component of a complete theory of meaning for a language, a sub-theory which is to serve to specify the contents of (for instance, and surely centrally) assertions which could be made by uttering the language's indicative sentences .... [A] direct assault on that task would be to look for a sub-theory which generates, on the basis of structure in the object-language sentences, a theorem, for every appropriate sentence, of this form: 's can be used to assert that p'. Now there is a truistic connection between the content of an assertion and a familiar notion of truth ...; the connection guarantees, as the merest platitude, that a correct specification of what can be asserted, by the assertoric utterance of a sentence, cannot but be a specification of a condition under which the sentence is true. A radical proposal at this point would be as follows: as long as the ends of the theorems (think of them as having the form 's... P') are so related that, whatever the theorems actually say, we can use them as if they said something of the form 's can be used to assert that p', it does not actually matter if we write, between those ends, something else which yields a truth in the same circumstances; our platitude guarantees that 'is true if and only if' fits that bill, and this gives a more tractable target than that of the direct assault. (1981, pp. 228–9)

McDowell here takes for granted the association between assertoric force and the indicative mood. In speaking of the indicative sentences of an arbitrary language, he clearly has in mind the assertoric sentences. The idea is thus that the truth theory (or rather, what McDowell says may as well be regarded as a truth theory) yields an account of the descriptive content, or sense, of what are effectively the sentence radicals of the language in question. For a complete theory of meaning, we need to add to this content-specifying sub-theory a theory of force – an account of the various different sorts of things that speakers of the language do with sentence contents (including, of course, asserting them).

As we noted in chapter 1, the extent of this additional theory of force depends on how many varieties of force there are reckoned to be in the object language in question. This suggests that other things being equal it would be better to treat differences of meaning as differences of sense than as differences of force; for this way they are incorporated into the theory of sense, and don't have to be dealt with later. Every effort should be made to handle a difference of meaning *within* the sort of sub-theory that McDowell describes, before consigning it to the assortment of leftovers which will have to be dealt with when that sub-theory is complete. Of course, if we had some criterion for whether a difference of meaning was a matter of sense or a matter of force, there would be no question of *trying* to accommodate it in the sub-theory; it would be in or out,

according to what the criterion told us. But at the moment we have no criterion. A substantial theory of truth is supposed to give us one.

Moreover, McDowell's care in distinguishing a theory whose primary role is to state the contents of sentences of an object language from a theory of truth, removes what may have seemed a barrier to accommodating certain sorts of differences of meaning within a theory of this general form. The standard 'filling' for a T-sentence – 'is true if and only if – suggests that what goes into the blanks must at the left hand end designate the sort of object language sentence to which the predicate 'true' applies; and at the right hand end be the sort of metalanguage sentence that can form the antecedent or consequent of a material conditional. But McDowell's insistence that the standard filling is not obligatory should alert us to the possibility of a filling that does the same work, without imposing such a grammatical restriction; and that hence enables the same process of meaning specification to extend to what are traditionally regarded as matters of force.

McDowell suggests the filling

(2.3) ... can be used to assert that ....

Of course, he has already restricted the theory to assertions – to 'the language's indicative sentences'. Except in that it actually hints at this restriction, 2.3 is no different from

(2.4) ... can be used to say that ....

However, 2.3 and 2.4 might still seem to restrict the kinds of sentences we can use to fill the blanks. On the right-hand side, in particular, we are still effectively restricted to indicative sentences; for English grammar won't allow anything else to fill the gap in this sort of 'that' clause.

At this point, however, we can appeal to Davidson's own (1968) 'paratactic' analysis of this kind of grammatical construction. On this analysis the logical form of the sentence 'Ramsey said that truth is redundant' is

(2.5) Truth is redundant. Ramsey said that.

This allows that the clause which appears after 'that' in the original sentence is *used*, rather than *mentioned*, and hence makes sense of the intuition that in translating this remark about Ramsey into another language, we should not leave this clause in English. (We are not *quoting* Ramsey.)

The present relevance of this proposal is that it frees us from grammatical restrictions on the nature of 'that ...' clauses.

(2.6) Make truth redundant. Ramsey said that.

makes as much sense as 2.5. So indeed does

(2.7) Is truth redundant? Ramsey said that.

Of course, in this case we would be inclined to say 'Ramsey *asked* that', but this is incidental. The important point is that just as in 2.5, the imperative and interrogative sentences in 2.6 and 2.7 are used rather than mentioned. We are not reporting Ramsey's exact words. We are *samesaying*, as Davidson (1968, p. 140) puts it, in our own words.

Applying the same analysis to 2.4, the ‘filler’ we derived from McDowell, we find that the theorems of resulting sub-theory can be regarded as having the form

(2.8) S can be used to say this: P,

and that there is now nothing in the form of these theorems to prevent the sub-theory from extending to the non-indicative sentences of the object language. If the object language contains imperatives, for example, there is now nothing to prevent the sub-theory giving us their meaning in the same way it does for indicatives: by *using* a metalanguage sentence, and telling us that this sentence says the same thing as the given sentence of the object language. Note that the fact that the metalanguage sentence is still used rather than mentioned means that the theory is still distinguished from a translation manual in the way that truth theorists have been keen to stress. (See, for example, Evans and McDowell 1976, p. x.)

It now appears that the preferred form for what was supposed to be a theory of the *content* of the sentences of the object language, makes no distinction between differences in content (or sense) and differences in force. The tasks initially consigned to the separate theory of force are in fact taken care of by the so-called sub-theory. The significance of the force indicators of the object language – its imperative and interrogative moods, for example – is explained in the same way as that of sense-modifying expressions: by the use of corresponding constructions in the metalanguage. The supplementary theory, originally thought of as a theory of force, survives, at most, in a force-neutral account of the activity of *saying*: a mere gloss to the main theory, whose main task would be to distinguish the full-blooded use of sentences from other sorts of appearances in writing and in speech.

This need not mean that there is no longer a place for the notion of force; only that the distinction between force and sense need not be reflected in the radical interpreter’s methodology, as usually supposed. As long as we draw such distinctions in the metalanguage, the process of interpretation will draw corresponding distinctions in the object language. But the process makes no intrinsic demands on such distinctions – which suggests that we can’t found an elucidation of the fact-stating/non-fact-stating distinction (or, what is supposed to come to the same thing, a substantial theory of truth) on the requirements of radical interpretation.

If this conclusion seems surprising, I think the impression stems from a distorted view of the aim of radical interpretation. Radical interpretation is most naturally seen as the programme of systematic translation of the sayings of others into the home language. ‘Sayings’ here means sayings of all kinds – as many kinds as we need, to understand what our subjects are saying. From this point of view the introduction of a Fregean sense/force distinction looks rather odd – a case of changing horses in mid-stream. So from the radical interpreter’s natural perspective there is an odd discontinuity in a truth-theoretic theory of meaning. We have seen however that truth-theoretic general semantics shows a constitutional tendency to concentrate on assertoric uses of language. This emphasis obscures the discontinuity, because it obscures the fact that a truth-theoretic theory of meaning leaves us some way short of the interpreter’s goal. *Saying* comes to be equated with *asserting*, and the radical interpreter’s programme comes mistakenly to be identified with the task of specifying the contents of assertoric utterances.

I want to mention two objections to the above conclusion – two considerations that might be thought to show that a radical interpreter does have methodological grounds to distinguish indicatives from non-indicatives. (I discuss these and other objections at greater length in Price 1987.) The first stems from a popular view of the empirical grounds of a theory of radical interpretation: the view that ‘the evidential base ... will consist of facts about the circumstances

under which speakers hold sentences of their language to be true' (Davidson 1974, p. 320). It might be claimed that the radical interpreter's theory will need to embody a sub-theory of the kind McDowell describes, in order properly to reflect its empirical base.

The objection is easily turned, however. The available 'facts about circumstances under which speakers hold sentences of their language to be true' cannot exceed the available facts about circumstances under which speakers are prepared to utter these sentences, with serious intent. Adding the truth predicate doesn't help to distinguish the literal utterances from all the rest; and doesn't, for example, help us to decide whether reluctance to make an utterance is due to politeness or disbelief. The subject who is too polite to say, 'You linguists all look the same to me', will be too polite to assent to the suggestion that we linguists all look the same to him.

On the other hand, to insist on framing the evidence in terms of 'holding true' is to restrict one's gaze, quite unnecessarily, to those sentences that can be held true. It is not clear that such a restriction is possible – it depends on being able to distinguish (what will be interpreted as) the indicative sentences at a very early stage. But even if it is possible, there can be nothing to be gained by it – far better to collect facts about the circumstances in which sentences of all kinds are seriously uttered. The facts about 'holding true' can then be distilled, if necessary, when we've distinguished the indicatives – and the residue will do us nothing but good. (The interpreter's standard question will thus be, not 'Would you assent to this: P?', but 'Would it be appropriate to say this: P?')

Secondly, the truth theorist might object that the justification for restricting one's initial meaning-specifying theory to the indicatives lies in the advantages of expressing its theorems in the predicative form exemplified in 2.2. This point is best coupled with another. The main thing that recommends a truth-theoretic approach to a theory of meaning is its ability to deliver plausible recursive clauses for at least some of the sentence-forming operations of natural language; while at the same time keeping faith with the intuition that an adequate theory of meaning will give us the meanings of expressions of the object language by using, rather than mentioning, expressions of the metalanguage. It might be thought that these virtues justify what otherwise seems a needless complexity: the restriction of the meaning-specifying theory to indicative sentences, and the consequent introduction of a separate theory of force.

This argument depends on the assumption that it is impossible to express the theorems of our extended sub-theory in the predicative form of 2.1. In other words, it turns on the impossibility of a predicate  $\varphi$  such that for *any* sentence P, 'P is  $\varphi$ ' can be used to say that P – i.e. such that

(2.9) 'P is  $\varphi$ ' can be used to say this: P.

However, in a sense we have already encountered such a predicate, in Strawson's amen theory of truth. We saw that the amen theory does not explain why 'true' should not be used to endorse any sort of utterance whatsoever. It is thus insubstantial in the first of the ways we distinguished in section 2.2.

We noted that the same can be said of any theory of truth that relies on the equivalence property – on the fact that for any (appropriate) sentence P, P and 'P is true' are in some sense equivalent in meaning. However, it is precisely this feature of truth that is exploited by the present objection. For imagine that we could generalize truth to the non-indicatives, its significance being given by the generalized equivalence principle:

(2.10) For all sentences P, P is true if and only if P.

The theorems of our general meaning-specifying theory could then be written in the predicative form of 2.2, and the advantages of such a form could not be the grounds for a discontinuity in a programme of radical interpretation. Our generalized truth theory would do what an orthodox truth theory does, but do it more generally. Certainly, we could carve the orthodox theory out of the general one; but only if we already knew what to cut away.

So the objector needs a reason why we shouldn't generalize the equivalence principle, and hence apply 'true' to non-indicatives. As we saw, however, such a reason will call for an independent characterization of the indicatives, in order to account for the required restriction. This defeats the avowed aim of a substantial theory of truth. There is surely some reason why we don't use 'true' in this way. But unless we know more about truth than the equivalence property tells us, it is bound to be a reason to which a substantial theory of truth is not allowed to appeal.

In a footnote to the passage quoted above, McDowell remarks that 'it is a philosophical issue whether there are respectable purposes for which a stronger notion of truth is required' – a stronger notion, in other words, than that guaranteed by the 'platitude' that '... is true if and only if ...' is a suitable filling for the theorems of the meaning-specifying sub-theory. McDowell notes that non-factualists will want a stronger notion, but says that he is 'inclined to suppose that this is a matter not so much of an alternative notion of truth as of a characteristically philosophical misconception of the only notion of truth we really have: one which the platitude in fact suffices to determine' (1981, p. 229, n. 9). We find, however, that the platitude does not account for the ordinary use of truth. The truth-theoretic approach to truth is insubstantial in the first of the ways we earlier distinguished. It picks out the indicatives (or any smaller class) only to the extent that it takes for granted that such utterances are the only bearers of truth. So McDowell's platitude does not suffice to determine the fact that truth is not applied to the non-indicatives; and the non-factualist is not alone in wanting something stronger.

## 2.5 THE MARKS OF TRUTH

Wiggins claims to derive certain 'marks of a theory of truth' from the anthropological constraints on a Davidsonian programme of radical interpretation. However, we have seen that that programme does not yield the kind of distinction that a substantial theory of truth is supposed to underpin. If we already have the distinction, interpretation will reflect it; but it needs to come from somewhere else. What then of Wiggins's marks of truth? Are they properties of truth that, like the equivalence property, could equally extend to non-indicatives? Or do they depend on some prior conception of the nature of fact-stating discourse – on the sort of conception that a substantial theory of truth is supposed to provide? The answer, I think, is the latter. Wiggins allows himself the notions of assertion and belief.

Thus 'the first mark of assertibility (and so of truth)' consists in 'assertibility's being *the primary dimension of assessment for sentences*, or that property which sentences have *normally* to be construed as aiming to enjoy'. If there were no such norm, Wiggins says, 'interpretation would either be impossible ... or unjustifiable' (1980, p. 205). Wiggins's point seems to be that unless we take utterance to be constrained by such a norm, we shall be unable to take seriously anything that our speakers say. But only the implicit assumption that by 'utterance' we mean 'assertion' can justify Wiggins's way of stating the constraint. Certainly, we need to assume that by and large our subjects are speaking seriously – that they mean what they say – but only in the case of assertions does this amount to the assumption that they are aiming to speak the truth.

As for Wiggins's remaining marks of truth, they make essential use of the claim that 'language ... *par excellence* ... is the vehicle for the communication and expression of belief' (1980, p. 205). Thus:

‘The argument for [the third mark] ... is similar to the argument for the second mark in turning upon the notion of belief (pp. 208–9); ‘... the fourth mark [is] simply... a condensation of the first, second and third ...’ (p. 211); and as for the fifth mark, ‘it is a norm of rational belief that one is *committed* to conform one’s beliefs to this requirement; and the interpretation of beliefs as beliefs has to see belief as *answerable* to it’ (p. 212).

As a result, Wiggins’s theory of truth is insubstantial in the second of the two senses of section 2.2. Given the notions of belief and expression of belief on which Wiggins here relies, we could quite well characterize the fact-stating use of language in terms of expression of belief. The basic distinction would be that between those utterances whose standard function is to express a belief and those that do something else. Far from looking for a substantial theory of truth, we ought then to expect our account of truth to flow from the same source as our theory of assertion. The obvious approach would be to regard truth as primarily a property of belief, and thence a derivative property of expressions of belief. The correctness of an assertion would thus be a matter of the correctness of the belief it would normally express.

The argument of this chapter should alert us to a potential hazard for an account of this kind. We have seen that assertion cannot be characterized in terms of truth. On the contrary, an account of truth and falsity as linguistic properties seems bound to depend on an elucidation of the fact-stating type of discourse. In offering such an elucidation we must be careful not to depend on a linguistic notion of truth. In particular, therefore, we cannot rely on a characterization of belief that itself depends on a notion of truth as a property of linguistic entities. This notion should arise from, not underpin, our psychological theory. If we want to analyse the fact-stating/non-fact-stating distinction in terms of a distinction between beliefs and other sorts of psychological state, we must be prepared temporarily to forego the linguistic notion of truth. This point will be very much in evidence in chapter 5, where we shall be looking in more detail at attempts to find a psychological basis for the notion of statement of fact.

## 3

**The Insignificance of Mood**

Marker theorists and non-factualists agree that there is a non-trivial distinction between fact-stating and non-fact-stating uses of natural language. They differ, however, as to where this distinction should be drawn. We have seen that if the marker theorist tries to settle the issue by appealing to truth, the attempt backfires. A notion of truth that secures the indicatives tends also to encompass the non-indicatives. We shall see now that this is the inevitable predicament of the marker view. No analysis of fact-stating discourse that includes the indicatives can exclude the non-indicatives. This will enable us to put the marker theory and the non-indicatives to one side. We shall be able to concentrate, in other words, on whether there is a distinction to be drawn within the indicatives.

Our goal is thus a generalized sceptical challenge to the marker view – an argument that non-indicatives cannot be shown not to be assertions. Since this would be entailed by a demonstration that non-indicatives are assertions, our course runs parallel to that of the ‘universal factualist’, who holds that *all* discourse is fact-stating. However, we’ll see that the path leads not to universal factualism but to our own less ambitious destination. It does not establish that non-indicatives are assertions; but it does provide a plausible *reduction* of any purported demonstration that non-indicatives are not assertions.

It will be helpful to have in mind some of the characteristics that have been thought to distinguish the non-indicatives, other than their standing to the properties of truth and falsity. So let us again refer to the views of Dummett, whom we have seen to be a critic of the attempt to analyse assertion in terms of truth. In his discussion of assertion in *Frege: Philosophy of Language*, Dummett mentions four main alternative characteristics. He says that assertions are the expressions of mental states with inner consequences (so that we can make sense of an internal assertion – a judgement – but not for example an internal command); are not typically associated with definite consequences (as are questions and commands); can be used to deceive; and can be used (perhaps stripped of assertoric force) as components of complex sentences. (Dummett 1973, chapter 10).

Not all of these features could plausibly be offered by way of analysis of assertoric discourse (or are, I think, so intended by Dummett). The last property is clearly grammar-dependent; and hence, as we saw in chapter 2, no use to the marker theorist, whose aim is to explain the grammar. At best this criterion amounts to a further characterization – a rather obvious one – of the distinction in usage that calls for explanation.

The suggestion that only assertions can be used to deceive also seems an unlikely basis for an analysis of assertion. If true, it surely rests on some more fundamental property of assertion. However, it proves a useful focus for our present concerns. In asking how non-indicatives can be used to deceive, we shall be led to a notion of the informational content that such an utterance could be considered to convey. From there it will be a short step to the claim that a non-indicative can be construed as an assertion with such a content – or rather, cannot be shown not to be such an assertion.

The attempt to make sense of non-indicative deception will lead us in particular to the proposal that non-indicatives are equivalent to explicit performatives (these being themselves assertoric). This proposal – advocated in the interests of general semantics by David Lewis (1972) – has been criticized on a number of grounds. I shall defend it against the better known objections, but

argue that it is nevertheless untenable. Once recognized, however, the main problem is easily solved. I shall argue that the improved account meets our purposes (if not the universal factualist's); and confirm that in so far as they need to do so, the assertoric paraphrases suggested by the account do indeed have the features Dummett takes to distinguish the non-indicatives.

### 3.1 NON-INDICATIVE DECEPTION

The universal factualist has two possible responses to the claim that assertions are distinctive in allowing deception: to argue that some assertions (and in particular non-indicative assertions) cannot be used to deceive; or to try to show that non-indicatives can be used to deceive. Let us begin with the latter possibility.

In support of the claim that assertions can be characterized in this way, Dummett argues that non-indicatives show no parallel to Moore's paradox. He concedes that 'the asking of a question is in a sense the expression of a desire to know the answer: so one may ask a question to deceive someone into thinking that one wants to know the answer.' But he counters that this

is not ... part of the convention which is learned when one learns to use and answer questions: it merely results from the fact that a desire to know the answer provides the usual motive for asking a question. This is shown by the fact ... that there is nothing self-defeating in asking a question and at the same time letting it be seen that one has no interest in the answer (if, e.g., one is an official whose duty involves the posing of certain questions); whereas the making of an assertion while letting it be seen that one does not believe it to be true is intrinsically self-defeating. (1973, p. 356)

At best, however, this shows us that a question does not express an interest in knowing its answer. It does not show that there is not some other attitude whose disavowal will frustrate an interrogative speech act. And on the face of it, any linguistic act can be defeated by an appropriately chosen attitude report. One simply needs to add that the utterance was not intended seriously, or sincerely. This takes advantage of the most general linguistic convention of all, namely that one means what one says. There may be various ways of signifying that a previous utterance should not be taken as governed by this convention, but there is always at least one way: thanks to Quine, one need only add that the utterance should be counted a mention rather than a use. True, not all effects of an utterance can be cancelled in this way. Some effects are insensitive to a speaker's intentions: for example, effects on voice-operated machinery. Others are immediate, so that the disavowal after the fact is bound to come too late. (Think of crying 'Wolf!' – or for that matter, of the effects of 'Think of a wolf!') Clearly, however, such cases do not distinguish the non-indicatives.

Does the possibility of insincerity allow non-indicatives to be used to deceive? It might be objected that pretending to sincerity is not the same as lying. The liar pretends to sincerity, but not all pretenders to sincerity are liars. A person who utters an insincere non-indicative misleads his audience about his own sincerity, but does not *lie*. Whereas a person who lies misleads her audience in two ways: about her own sincerity, and (if she is believed, at least) about the facts. To lie is to try to mislead in the second sense. In the case of commands, for example, the claimed contrast would thus be that a person who obeys a command is in no way at fault for doing so, even if it turns out that the command was insincere. Whereas a person who unknowingly accepts an insincere assertion may end up, not at fault for accepting it – that was the appropriate thing to do – but at fault in possessing a false belief. It is a virtue of this difference, it seems, that lying misleads about more than a speaker's sincerity.

However, it is easy to see how obedience to insincere commands might be inappropriate. In general obedience will involve some disadvantage, or behavioural cost; but this will be regarded as preferable to the disadvantages of disobedience – in other words, to being punished for failure to comply. However, the relevant social conventions might well recognize the possibility of insincere commands, and administer punishment accordingly. Disobedience to an insincere command would thus go unpunished. Hence we would have a worldly basis for the state of being misled by an insincere command, in the consequent performance of an unnecessary and otherwise undesirable action. Moreover, we could match such a state to a false belief. We could say that someone wrongly believed that he was *required* to perform some action. This suggests that we might interpret a command as an assertion with content of this belief – for example, that we equate ‘Fill in this form’ with ‘You are required to fill in this form.’

To forestall an objection, let me clarify the role of the doubtful (and certainly contingent) proposition that insincere commands do not establish real obligations – that a person is not at fault for failing to comply with what turns out to be an insincere imperative. The suggestion does not require that actual commands conform to this convention. The point is rather that since the giving of commands clearly could be constrained in this way, we can make sense of a worldly basis for imperative deception. In terms of that basis we can then explain, if necessary, the fact that some or all commands do not permit this sort of deception. I shall come back to this point. But briefly: we can say that some or all actual commands are governed by the convention that they be self-justifying – the convention that their own utterance be sufficient to establish the requirement that makes their suggested paraphrase true.

There would be other objections to any serious proposal to paraphrase imperatives as reports of obligations. As yet, for example, such a report does not capture the first-person character of a command. Perhaps the paraphrase could be refined to cope with such objections, but for present purposes there is a more serious concern. Even if reports of requirements can be made to do duty for imperatives, we shall be left with the other non-indicatives. Insincere commands may mislead us about our obligations, but what about insincere questions, requests, wishes, and so on? A case by case argument is no use to us. We need a general guarantee that some such transformation will always be available. I want to show that the marker theory gives us such a guarantee. This will deliver the promised *reductio*. It will show that from the marker theorist’s point of view, universal factualism remains a viable alternative.

### 3.2 HOW TO CONSTRUE THINGS IN WORDS

The general strategy relies on some familiar points about linguistic communication. Utterances have various effects on the social environment of speakers and audiences. Among these effects are some that depend on linguistic understanding – on the fact that audiences are able to interpret what speakers say. Effects of this kind are convention governed. They rely on the meaning conventions in force in the community in question. Interpretation of an utterance can be thought of, in part, as a process of acquiring beliefs about these convention-governed effects. These beliefs are not often explicit; but it would seem that any of them could be elicited by an appropriate enquiry, at least from a subject with the necessary conceptual resources. Learning to understand a language seems in large part a matter of learning to form such beliefs in the right circumstances.

I take this much to be uncontroversial. The difficulties arise when we try to fill in the details. In particular, the interpretation of a single utterance will in general give rise to a string of beliefs. The proper characterization of some of these beliefs is far from obvious. Interpretation presumably involves the application of the interpreter’s knowledge of meaning, and hence the

proper form of interpretative beliefs will depend on that of a theory of meaning for the language in question. A theory of meaning will bring with it an account of the corresponding interpretative beliefs.

The marker theory, for example, claims that mood distinctions mark significant meaning categories. If so, then the process of interpretation should recognize the factor of meaning marked by mood. Hence on this theory we might think of the main elements of the interpretative string as the following four beliefs: that someone has spoken (rather than simply made a noise); that a certain speech act (a command, an assertion, or whatever) has been performed; that the speech act concerned has a particular content; and that one is oneself thereby affected in some way (for example in being thereby obliged to perform some action).

The present significance of these points lies in the fact that beliefs can be acquired not only directly, in virtue of perceived features of one's environment; but also indirectly, as a result of an assertion by somebody else. So if the interpretation and hence the effects of an utterance always depend on chains of belief, we might achieve the same final effects in a different way: that is, by simply *telling* a person what they would otherwise come to believe at some point in the chain of interpretation. From the marker theorist's point of view this is what happens when we replace a command with an assertion of the existence of a corresponding requirement. In both cases we normally cause our audience to believe that there is such a requirement, but the initial pathways are different. In one case the belief concerned is inferred from the observation that we have issued a command; in the other it flows in the normal way from the observation that we have made an assertion – from the fact that we have *said* that there exists the requirement in question. But note that not only do the pathways converge, in giving rise to the same belief; but also the effects of initial insincerity may be the same, in making this belief false.

The marker theory suggests a simple and elegant way to adapt this construction to the general case. We want to break into a chain of interpretative beliefs, so as to assert what a hearer would otherwise infer. It doesn't seem to matter where we break in, as long as we can characterize the information the new assertion needs to convey. Given the marker theory, the obvious target is the point at which the mood itself is first deciphered. At this point the interpreter comes to believe that he is faced with a question, a request, or whatever. Why not replace the original utterance with an assertion to this effect?

The suggestion is thus that we exploit the marker theorist's account of the significance of mood. That account entails that the process of interpretation of an utterance requires, *inter alia*, a belief that a speech act of a certain kind has taken place – the kind being the meaning category that the marker theorist associates with the mood of the utterance in question. The marker theorist takes this belief to determine the further effects of the utterance concerned (or those, at any rate, that depend on its mood). The suggestion is that we imagine the same belief to be transmitted in another way: not inferred from mood, but reported in content. Non-indicatives are thus to be paraphrased in terms of explicit performatives (these being themselves regarded as assertoric). 'Fill in the form' will now be read not as 'You are required to fill in the form', but simply as 'You are commanded to fill in the form'; or more explicitly, as '*I* command you to fill in the form.'

In this way, the marker theory seems to provide the universal factualist with the material for an assertoric interpretation of the non-indicatives. It allows the factualist to propose that non-indicatives are merely 'disguised' reports of the occurrence of the very speech acts the marker theory claims to identify. To admit such readings would be to abandon the claim that the indicative mood is the mark of assertoric force. But how is the marker theorist to exclude them, given the intuitive appeal of the interpretative model from which they stem?

### 3.3 PERFORMATIVES: TOO EXPLICIT FOR THE FACTUALIST

The marker theorist might well follow J. L. Austin (1962, pp. 5–6, for example) in denying that explicit performatives are fact-stating – as of course the above suggestion requires, if it is to give us *assertoric* as opposed to merely *indicative*, substitutes for non-indicative utterances. In our terms, Austin was a non-factualist: he took it as a discovery about language that it contains this large class of non-assertoric indicative sentences. On Austin's view, the expression 'I command ...' is not used to describe one's own behaviour – to *assert* that one commands – but simply to command; similarly with 'I request ...', 'I promise ...', and many others.

Some of Austin's critics, however, though agreeing that explicit performatives are used for these purposes, have argued that such utterances are also assertions. One such critic is David Lewis, whose interest in explicit performatives is much the same as ours: faced with the task of incorporating non-indicatives into a general theory of meaning, Lewis recommends that they be treated as disguised performatives, and that these be regarded as assertoric, or declarative. In Lewis's words, the result is that

the distinction between declarative and non-declarative sentences becomes a purely syntactic, surface distinction. The only distinction among meanings is the distinction between those sentential meanings that can only be represented by declarative sentences and those that can be represented either by suitable declarative sentences (performatives) or by non-declarative paraphrases thereof. (1972, p. 208)

Lewis's proposal has been criticized on several grounds. We shall see that although the critics have mostly missed the mark, there is a valid objection. However, the problem stems from the unnecessarily Austinian character of the performative approach. Cured of its fixation with speech acts, the proposal turns out to be just what we want.

Lewis's critics have concentrated on the claim that explicit performatives are declarative; and in particular on the consequence that an utterance of an explicit performative must, as McGinn puts it,

be tantamount to the performance of *two* speech acts, where the (felicitous) performance of one (saying or asserting) is sufficient for the performance of the other (commanding, etc.) .... So in uttering a performative one both performs a certain speech act and declares that one does. (1977, p. 305)

McGinn's own objection to this consequence of Lewis's account is that

no such duality is apparent in respect of straight nonindicatives .... 'I command you to shut the door' mentions me and commanding, but 'Shut the door' manifestly does not. So they are simply not equivalent in meaning. (1977, p. 305)

In advance of a semantic analysis of non-indicatives, however, I think this is premature. The significance of my utterance of 'Shut the door' clearly depends in some way on the fact that it was I who said it (you may decline to obey, for example, for precisely that reason). Is it obvious that this significance is not the result of an implicit reference to myself? Similarly, one of the effects of this utterance is surely that you come to believe that I have ordered you to shut the door. Is it clear that this should not be explained by saying that I have implicitly stated that I have so commanded you? (Compare: 'Bring me pig meat' mentions pigs and meat. 'Bring me pork' manifestly does not. So they are simply not equivalent in meaning.)

The best attempt to make an objection out of the duality of the explicit performative seems to be that of Stephen Schiffer. Schiffer takes over from Austin the principle that ‘ordinarily one uses the explicit performative formula to make explicit the precise act one is performing in the issuing of one’s utterance’ (1972, p. 107). He then points out that if we assume that performatives have assertoric force, this principle leads to a vicious regress. To make explicit the fact that ‘I order you to shut the door’ performs an assertoric speech act, we would have to say ‘I assert that I order you to shut the door.’ And here the same problem occurs again: to make explicit the assertoric force of this new assertion, we would need a yet longer explicit performative.

Schiffer’s solution is to say that although performatives are constative, they are ordinarily used without assertoric force. This will not do for Lewis, who wants to claim, in effect, that we give commands by asserting that we are doing so. If ‘I order you to shut the door’ did not have assertoric force, you could justify your inactivity with the remark that I had not told you that there was a command to obey – that you had understood that I was merely entertaining the thought that I was thereby ordering you to shut the door. But as Kent Bach (1975) points out, a more plausible solution to Schiffer’s regress is simply to reject Austin’s view that performatives make explicit the precise nature of the speech act they perform. Instead, we can say that they make explicit one speech act in performing another.

The real difficulty for Lewis’s account seems to me to stem not from the dual role it accords to the explicit performative, but from the fact that the roles are not independent. Lewis wants a semantic analysis of non-indicatives. In effect, his account is meant to be a guide to the interpretation of non-indicative utterances. It tells us, first, that they have the internal structure of a particular kind of assertion: namely an explicit performative. The structure thus looks like this:

(3.1) I perform the speech act A.

So the account tells us that if we receive a non-indicative linguistic ‘package’, its outer layer is to be treated as an assertion that we are being offered another package, another speech act. To understand the original, we have to unpack the package it claims to contain (unless we do so, after all, there will be nothing to distinguish one non-indicative from another). But what is this new package? Why, the very one we started with. Offered a command, for example, we are told to treat it as the assertion that we are being offered a command. Obliging we refer to our account of commands, to understand what it is that the assertion says we have been offered. But it is a command, and so the account tells us that it is equivalent to an explicit performative – an assertion that we have been offered a command. The chain of interpretation thus has a loop in it.

This is not Schiffer’s objection. Schiffer tries to build a regress outwards, working on the assertoric part of an explicit performative. This objection goes inwards, working on the speech act that the assertoric part reports. The problem stems from the attempt to analyse the speech acts performed by the non-indicatives in terms that refer to these speech acts themselves – that is, as reports of the performance of these speech acts. As Charles Hamblin puts it,

the so-called reduction ... [is] a logician’s blind alley; ... it is not really a reduction at all. If someone did not know what an order was, and needed to be told what meaning to attach to a given one *O*, it would be no help at all to tell him that it was a locution that, uttered by me to you, meant

I order you to carry out order *O*,

since he wouldn't understand this any better; *O* remains embedded in it, unanalysed and un glossed. (1987, p. 135)

The marker theorist can thus object that the universal factualist has not offered a coherent assertoric analysis of the non-indicatives. An analysis of meaning is required to account for linguistic understanding. The explicit performative account cannot do this, because the performative refers to the very linguistic act whose analysis is in question. The factualist must do more than show that for any speech act conventionally performed by non-indicative means, we can find an indicative utterance that could be taken to perform the given speech act. Speech acts being conventional, any indicative utterance (indeed, any public action) can be made to do such additional duty. But the trick requires a prior conception of the speech act we propose to add to the conventional significance of the chosen indicative. So it cannot provide an alternative analysis of the speech act concerned. It is as if we had tried to show that pocket knives can be thought of as potato peelers, by producing an object that physically combines the two sorts of implement. What we need to show is rather that an arbitrary pocket knife can already be thought of as a potato peeler (by showing, of course, that it can be grasped as such).

### 3.4 SPEECH ACTS AND SPEECH EFFECTS

The problem for the performative approach to an assertoric treatment of non-indicatives is thus that if performatives are thought of as reporting speech acts, the resulting analysis is viciously self-referential. The key to a solution, I think, is to appreciate that what such an assertion reports does not have to be a speech act at all. It is again crucial that we are dealing with conventional acts. We need the fact that if a social state of affairs is actually the conventional effect of an utterance, then it could instead be taken to be the conventional effect of almost any action (or indeed of other sorts of event: there are societies in which conventional obligations are taken to be dependent on random physical processes, for example).

Thus if convention enables a speaker to impose a requirement by saying 'Please leave the room', it could quite well enable the same speaker to impose the same requirement by other means – including, I think, a simple mental decision that the person concerned shall be required to leave the room. It might be objected that imposing the requirement in this way would enable people to become subject to requirements about which they had no way of knowing. This is true, I think, but hardly problematic. There are other sources of obligations to which individuals may unknowingly be subject. (The possibility is reflected in the principle that ignorance of the law is no excuse.) In any case, it could easily be avoided by treating such requirements as conditional on publicity.

We are liable to be misled here by the very view we are attempting to undermine. If we begin with the idea that a command is a type of speech act – different from, but on a par with, the act of assertion – then of course we read 'I command ...', if as a report at all, as a report of the occurrence of such a speech act. But if we are not already committed to a non-assertoric view of imperatives, we don't have the same reason to think of commands as a category of speech acts. And hence we are free to think of 'I command ...' as a report of a conventionally generated state of affairs: that very state of affairs that on the orthodox view results from the corresponding imperative speech act.

Looking at things from the orthodox point of view, we can thus separate the *effect* of a speech act – in the case of a command, for example, the requirement or obligation it establishes – from the speech act which normally produces that effect. Not only does the existence of the effect not depend on the occurrence of such a speech act; but we are free to characterize the effect in terms

that do not refer to such a speech act. And hence we can *report* the effect, without explicitly (due to our description of the effect) or implicitly (due to its implied causal history) reporting another speech act. So in place of a speech act that produces a conventional effect – and in the process, so to speak, alerts an audience to this effect – we can always imagine an assertion that there has been such an effect, understood as produced by the speaker by some other conventional means. We thus replace a speech act with a report, not of the occurrence of the speech act itself, but of the social state that speech act would normally produce – roughly, by a report of the relevant change that would have been made to an audience’s social environment by a performance of the original speech act.

We thus aim to intersect the chain of interpretation where recognition of the nature of a speech act gives way to response. We want to report just those effects of an utterance that a speaker is normally able to produce at will, without depending on audience compliance (except with respect to the relevant meaning conventions). Even at this level, of course, there are cases in which a speaker will fail to produce an intended effect. Lack of authority will vitiate a command, for example. (Austin 1962 describes many ways in which an intended speech act may be ‘unhappy’, as he puts it.) But these will simply be the cases in which the relevant reports are false.

I think this proposal gives us the indicative paraphrases we have been looking for. We have shown, in effect, that for each category of non-assertoric speech acts the marker theorist claims to identify, we could construct a class of indicative utterances that would serve the same linguistic function. These would be indicatives in the ordinary sense. At worst we would have to coin some new descriptive terms to refer to the typical effects of the non-indicative speech acts we were concerned to replace. The marker theorist cannot object to this procedure. Pressed to explain our terminology we can appeal to the marker theorist’s own account of the social significance of the force distinctions supposedly marked by the grammatical moods. In distinguishing questions, commands and other non-assertoric speech acts, the marker theorist incurs an obligation to elucidate these categories. Such an elucidation is bound to appeal to supposedly distinctive features of the social role of each of these forms of communication, and in particular to their distinctive effects on conventional relations between hearers and speakers. These effects (or some of them) provide the subject matter for the factualist’s assertoric paraphrases.

In summary then, the argument shows that if there were non-assertoric speech acts of the kind the marker view takes to be associated with the non-indicative moods, the function of these speech acts could quite well be performed by utterances that the marker view would clearly take to be assertoric. On what grounds then can the marker theorist maintain that non-indicatives are not simply ‘disguised’ assertions?

I emphasize again that the argument does not show that non-indicatives are assertions. It merely undermines the marker theorist’s grounds for claiming that non-indicatives are not assertions. So it supports the universal factualist only to extent that it counts against a contrary view. Indeed, I think its effect is to call into question the enterprise that universal factualism and the marker view have in common, of seeking to discern the ‘true’ logical form of natural utterances. For it suggests that the categories both views employ are sufficiently flexible to prevent any authoritative judgement on how they should be applied to natural language.

### 3.5 EXPLAINING THE APPEARANCES

The marker theorist’s best hope of a defence would be to find some significant characteristic of non-indicative speech acts not shared by the claimed indicative paraphrases. It is a slim hope,

however. For one thing, the argument exploits the marker theory itself, in conjunction with a simple and plausible model of linguistic interpretation. For another, the factualist's indicative paraphrases seem to account in one of three ways for the evidence that appears initially to favour the marker view. The supposed characteristics of the non-indicatives divide into three classes. Some supposedly distinctive properties of non-indicatives turn out to be shared by their indicative transforms. Some properties supposedly lacked by the non-indicatives turn out to be applicable after all, in ways determined by the indicative transforms. And some properties, as we have seen, turn out to be distinctive but merely grammatical – the kind of thing that the marker theory claims to explain.

All three possibilities are exhibited by the range of characteristics that Dummett takes to distinguish the non-indicatives. As we saw, Dummett says that assertions are not typically associated with definite consequences; are the expressions of mental states with inner consequences (so that we can make sense of an internal assertion – a judgement – but not for example an internal command); can be used to deceive; and can be used as components of complex sentences. Taking these in order, the impression that non-indicatives have definite consequences is confirmed and explained by the subject matter of their claimed indicative paraphrases. These are assertions about the social environment of a speaker and a recipient – in general, about features of this environment of immediate relevance to the recipient's actions. The indicative substitute for a command, for example, informs its recipient of a social obligation (or of a state of affairs of which such an obligation is normally an immediate consequence). So it is hardly surprising that such utterances have consequences which are unusually definite, by the standards of assertions in general.

The apparent lack of 'internal' correlates of non-indicative speech acts falls into the second of the above categories: the indicative transforms provide such internal correlates. Thus if we think of commands as assertions of the form 'You are required to do A', then their internalizations will take the form 'X is required to do A', where X will normally be a term that refers to someone other than the speaker. It is true that once this reference is made explicit, we can consider the case in which it is a self-reference. Externally or internally, judgements of the form 'I am required to do A' are perfectly respectable. Indeed, they have just the form in which the recipient of the indicative substitute for a command will report his or her newly acquired belief. Whether they have another use, as the transforms of self-commands, depends on extraneous factors: on whether such requirements can be self-imposed; and on whether, if so, there is ever a need to report them.

Similarly, the indicative transforms provide a sense in which commands can be said to express mental states with separate internal consequences. If commands express beliefs about requirements, for example, then such beliefs will have other effects: believing that you are required to complete the form, I take steps to ensure that if you do not do so (having been properly informed of the requirement), appropriate measures will be taken. It does not count against this view that the belief may be the result of my own decision to impose such a requirement, or that the requirement may be conditional on your actually receiving news of it.

The supposed impossibility of non-indicative deception also falls into the second category. We have already seen that one can be misled by an insincere command just as by an insincere report of a corresponding requirement. We noted that the assertoric model can cope with the possibility that disobedience might be regarded as punishable, irrespective of the commander's intentions. The basis of such requirements is bound to be a matter of social convention. One relevant convention might be that in the absence of explicit indications to the contrary, any utterance of a command or its indicative substitute shall count as establishing the requirement that makes that

substitute true. In limited ways this sort of convention governs many assertions. Think, for example, of ‘This is a stick-up’, uttered in a bank. Deceit is not always easy.

This leaves the suggestion that non-indicatives do not occur as sub-sentential components. As noted earlier, this is a grammatical restriction, and not therefore a feature of non-indicatives to which the marker view is entitled to appeal. As such, moreover, it collapses in the face of indicative substitutes for non-indicative speech acts. For example: ‘If you are required to complete the form, you are required to do so in black ink.’ The fact that we do not say ‘If complete the form then complete it in black ink’ is on this view no more than a simple consequence of the rules of non-indicative grammar. This does not mean that it is insignificant, but that its significance is just that of broader question as to the reason for the existence of the non-indicative moods.

We shall return briefly to this question in chapter 8. By then we shall have an alternative to the marker theorist’s claim that the moods are the conventional marks of underlying semantic categories. We have argued here that such a claim cannot be justified. Indeed, it follows from the argument that the convention the marker theory calls for would be unworkable. *Qua* speakers, just as *qua* theorists, we lack the mood-independent access to the claimed underlying categories that the convention would require.

In the main, however, the rest of the book will be concerned with the indicative mood and its internal structure. The aim of this and the previous chapter has been partly to sanction this concern, by showing that the issues the non-factualist raises within the indicative mood are not to be swept aside in the marker theory’s grand account of the moods as a whole. More importantly, however, it has been to show that there is no refuge in the marker view, in the face of a sceptical challenge to non-factualism. Such a challenge will be the major task of the next two chapters. We may now appreciate that what the sceptic challenges the non-factualist to exclude is not merely the marker view. That outcome, if not as philosophically illuminating as non-factualism is thought to be, would at least be familiar. Given that the marker theory is itself vulnerable to sceptical challenge, however, the sceptic’s alternative is not so tame. The threat to non-factualism is rather the bleak possibility that there is no way to exclude universal factualism – no analysis of statement of fact that entitles us to draw non-trivial limits to fact-stating discourse. We have seen that for non-factualists, as for marker theorists, to admit this possibility is to admit defeat. Appealing as it may seem to general semantics, universal factualism cannot support the distinctive claims of either doctrine. It cannot explain mood in terms of the limits of statementhood, and it cannot attribute philosophical errors to the misattribution of statementhood.

## 4

**A Retreat to the Head**

In this chapter and the next our goal is to embarrass the non-factualist with an abundance of facts. More precisely, we want to show that the non-factualist's best arguments are inconclusive, not excluding factualist interpretations of the topics in question. So again, our intentions are mainly sceptical: we want to show not that there are more facts than the non-factualist supposes, but that the non-factualist cannot establish that there are not such facts.

Non-factualism is more a species of philosophical theory than a theory in its own right. Since we cannot examine every member of the species, I propose to concentrate on three of its major manifestations: the non-factualist treatments of ethical judgements, probability, and (in chapter 5) indicative conditionals. Among the advantages of this selection is the fact that on the face of it, they have little to do with each other. This will add weight to the claim that the striking parallels which emerge, between probability and ethics in particular, reflect a characteristic pattern of non-factualist argument.

Broadly, the plan of these two chapters is as follows. In this chapter we want to show that non-factualists can be forced to depend on a psychological distinction – in effect, on the claim that genuine statements of fact are the expressions of genuine beliefs. In chapter 5 we shall then challenge their entitlement to this distinction. We shall argue that beliefs cannot be distinguished from the relevant non-beliefs without appealing – illicitly, so far as the non-factualist is concerned – either to truth or to existence of appropriate facts in the world.

First, however, a point of clarification. Such a case against non-factualism may seem superfluous, given the argument of chapter 3. We criticized the marker view on the grounds that by the standards of that view, any convention-governed speech act can be construed as an assertion. The marker view requires that a successful speech act convey, in some sense, the information marked by mood. Consequently, the marker theorist cannot exclude the possibility that the relevant feature is *reported*, rather than merely *marked*. This argument may seem to count as much against non-factualism as against the marker theory. If no utterance can be shown not to be an assertion, then in particular ethical judgements, conditionals, and the like cannot be shown to be non-fact-stating.

However, I think non-factualism fares better than the marker theory. For one thing, it is not clear that the non-factualist need require that the fact-stating/non-fact-stating distinction be marked in speech, or accessible in principle to ordinary speakers. If not, then the distinction need not be manifest in the chain of interpretation, in the way that the marker theory evidently requires. Since the *reductio* argument of chapter 3 depends on this feature of the marker view – the sceptic proposes an alternative chain of interpretation with the same final effect – non-factualism may well be immune.

For another thing, the contents of the assertoric paraphrases on which the *reductio* relies are themselves closely based on speech act categories. The idea was to map force distinctions onto a class of content distinctions (roughly, distinctions within the set of assertions about certain conventionally generated states of affairs). This means that the taxonomy of non-assertoric speech acts that the non-factualist claims to find within the indicative mood would survive the adoption of this sort of universal factualism – albeit on the basis of content distinctions rather than force distinctions. So the non-factualist's case could simply be rephrased. The question would be, not whether a moral judgement should be regarded as an assertion at all, but whether

it should be regarded, at face value, as an assertion about a realm of moral facts. The alternative would depend on the version of moral non-factualism on offer. It might for example be an assertion to the effect, roughly, that the speaker has expressed an evaluation of a certain kind.

The non-factualist could thus adapt to the possibility of universal factualism, by presenting the issue in terms of the *contents* of disputed utterances: granted that moral judgements, like everything else, are statements of fact, are they really about *moral* facts, or rather about some utterance-related conventional state? The non-factualist would still claim to show that certain classes of utterances are not assertions with the contents they seem at first sight to have. The negative claim will remain unchanged, though the positive claim will now be that the utterances concerned are assertions with some other content, rather than that they are not assertions at all. However, the negative thesis is the basis of the non-factualist's 'resolutions' of philosophical problems in various areas. So the approach seems essentially intact.

Moreover, our present interest, here and in chapter 5, is in criticizing the non-factualist's own case for the negative claim. The issue of what would follow from accepting this thesis will thus be largely irrelevant. It is true that in principle we might want to argue against the thesis by *reductio*, and hence be interested in its consequences. But in fact we shall concentrate our attack on the presuppositions of the non-factualist's typical grounds for accepting the thesis. The marker view's problems are thus of no immediate relevance. They stand in the background – a warning that we would be unwise to take the downfall of the non-factualists as a vindication of the marker theory.

#### 4.1 PROBABILITIES AND VALUES: THE NON-FACTUALIST INTERPRETATION

I want to exhibit some similarities between the usual grounds for non-factualism about ethical judgements and about probability. In particular, I want to show to show that for essentially parallel reasons, non-factualists in both areas can be forced to appeal to psychology. It will be helpful to begin with a characterization of what is at issue in each case. These characterizations are far from comprehensive, and perhaps exaggerate the initial formal similarity between probabilistic and ethical non-factualism. They will do for our purposes, however; and are not, I think, an artificial source of the striking parallels we find later on.

Let us begin with the probabilistic case. In English a probabilistic expression often has the syntactical form of a sentential operator: operators such as 'It is probable that ...', 'There is a 60 per cent chance that ...', and 'Probably, ...' itself, for example. Let us call a sentence formed by attaching such an operator to a suitable sub-sentence a *simple probability ascription*, or *SPA*. I have argued elsewhere (Price 1983a) that SPAs correspond to the so-called *single case* applications of probability, much discussed in the philosophical literature. For the present, however, the relevant question is whether a SPA should be construed as a statement of fact. Does a probabilistic operator modify the *content* or the *force* of a sentence to which it is attached?

Most philosophical accounts of probability take for granted the factualist answer. In effect, they are attempts to characterize the distinctive content of a SPA – to tell us what SPAs are characteristically *about*. There are several candidates: class ratios (finite or infinite, actual or hypothetical), logical relations, propensities, objective chances or probabilities, degrees of rational belief, and even a speaker's own degrees of confidence. As the list suggests, there is a wide range of opinion on what should be expected of an account of probability (and especially on the extent to which one should aim to reduce that notion to others).

The non-factualist, however, says that all these views are misconceived. SPAs have no common content. Literally, there is no more something that all sentences of the form ‘Probably P’ are about, than there is something that all sentences of the form ‘Is it the case that P?’ are about. ‘Probably P’ is about what P is about. It differs in that whereas an utterance of P is normally an assertion, ‘Probably P’ is normally a different sort of speech act – what may be labelled a *qualified* or *partial* assertion.

As in many cases, the linguistic distinction on which the non-factualist here relies is conveniently elucidated in psychological terms. It assumes a familiar model of assertion, belief and relevant action: an assertion that Q is typically an expression of a (full) belief that Q; and such a belief is typically displayed in a person’s acting as if Q whenever, in the light of her desires, she thinks it makes a difference to the outcome of her actions whether it is the case that Q. The factualist subsumes SPAs under such a model, regarding them as assertions that express (full) beliefs with some characteristic content. The non-factualist, on the other hand, takes SPAs to express ‘partial’ beliefs, or degrees of confidence. ‘Probably P’ is thus the characteristic expression not of the full belief that it is probable that P, but of a strong partial belief that P. A partial assertion (hence the term) is thus the expression of a partial belief. Moreover, because full beliefs are the limiting cases of partial beliefs, an ordinary or ‘full’ assertion is a special sort of partial assertion.

For the moment, not wishing to prejudge the question of the priority of psychological over metaphysical presentations of non-factualism, we should treat this version as merely expository. In particular, we want to hold open the possibility that the belief/non-belief distinction might need to be founded on the availability of relevant domains of facts in the world.

Turning now to the ethical case, we can characterize the issue in similar terms. For present purposes we can confine our attention to a particularly simple sort of ethical judgement: those of such forms as ‘It is a good thing that ...’, ‘It is disgraceful that ...’, and so on, in which the evaluative expression can be regarded as a sentential operator. Following the pattern of the probabilistic case, I shall call this a *simple ethical ascription*, or *SEA*.

The competing views are then as follows: the factualist holds that SEAs are assertions, and hence that ethical sentential operators are content modifiers; the non-factualist that the role of the sentential operator is rather to signal a non-assertoric force. A psychological elucidation again suggests itself. The factualist says that SEAs express ethical beliefs – beliefs about moral facts or states of affairs. Whereas the non-factualist takes SEAs to express some sort of distinctive moral attitude – approval or disapproval, say – towards the non-moral state of affairs referred to within the scope of the operator concerned.

Not all the arguments for non-factualism about SPAs run parallel to corresponding arguments about SEAs. For example, a probabilistic non-factualist would do well to help himself to some of the objections his factualist opponents raise against one another. Thus, he might well endorse the standard criticism of (most) frequency accounts of probability, concerning single case probabilities (the so called ‘problem of the reference class’ – see Mellor 1971, chapter 3, for example).

Similarly, the non-factualist will do well to concur with several serious objections to naïve subjectivism – the view that SPAs report a speaker’s own degrees of confidence. One such objection is the phenomenological one: in considering the probability of snow this afternoon, I don’t consider my own degree of belief in snow this afternoon. How could I, since the purpose of deliberation is to lead me to some such degree of belief? Another objection points to ordinary disagreements about SPAs. Clearly these are not disagreements about what each of the

participants currently believes. The argument gets its point from the fact that the participants both know what they each believe, and therefore know that they disagree. (In this case, clearly, there is a parallel argument against ethical subjectivism.)

However, arguments of this sort merely encourage the factualist to retreat to a stronger position. The strongest is the simplest: renouncing the attempt to provide a *reductive* interpretation of SPAs, it settles on the claim that SPAs are about, simply, probabilities. Probabilities are objective features of world, and hence a respectable subject matter for statements of fact.

It is easy to imagine a corresponding factualist position in the ethical case. And in the face of doctrines as bare as these, the non-factualist's options are rather limited. There is little for an objection to latch on to – and little, similarly, to distinguish the probabilistic and ethical cases. I think that in each case the non-factualist has two main avenues to explore. The first develops a metaphysical argument, while the second appeals to the functional role of ethical or probabilistic judgement in our lives. I want to show that the probabilistic and ethical cases do indeed run parallel here, and that in both cases both avenues lead to the head.

#### 4.2 ODD FACTS: THE ARGUMENT FROM QUEERNESS

In the ethical case, more perhaps than in any other, non-factualists have been prepared to appeal to the apparent peculiarity of facts of the disputed kind: 'If there were objective values, then they would be entities or qualities of a very strange sort, utterly different from anything else in the universe' (Mackie 1977, p. 38). Mackie calls this 'the argument from queerness'. It is often linked to an epistemological point. Non-factualists argue that if there were such facts, then 'if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else' (Mackie 1977, p. 38).

As we shall see, the non-factualist's metaphysical case does not rest here. And just as well, for there is nothing here to trouble a committed factualist. With a touch of irony, the factualist can object that queerness is in the eye of the beholder. Moral facts seem queer to a non-factualist because he already has a view of the world that excludes them. It is true that they differ from say physical facts; but this is no more than a truism. Facts about different sorts of things are bound to differ – for how else could they be about different sorts of things?

This leaves the epistemological point. But here the factualist will point out that her opponent agrees that we do not make ethical judgements (or whatever sort of judgement is at issue) at random. On any account, there is some process that leads us to make some such judgements and not others. Why not regard this process, whatever it is, as the so-called moral intuition? Once again (the factualist will say) the non-factualist's mistake is to start with some preconceived notion – this time of what a belief-forming process should be like – and then to dismiss the moral case on the grounds that it does not conform to this model.

The non-factualist may feel that these replies do not do justice to the argument from queerness and its epistemological shadow. But I think the feeling stems from the fact that the non-factualist has a stronger metaphysical argument in mind. The trouble with moral facts, the non-factualist suspects, is not that they are odd but that they are unnecessary. Indeed, we are entitled to regard them as odd because they are unnecessary. The non-factualist thinks that we can explain our *talk* of values on a much sparser metaphysical basis.

Similarly in the probabilistic case. Here the argument from queerness consists in an appeal to a conception of the world in which there seems no place for objective probabilities. The non-

factualist tells us that the world is the realm of the actual, the domain of physical things. Everything is determinate, every point either occupied or unoccupied, every particle in its place, and every proposition either true or false. Where, then, is the room for probability?

As it stands, this is no more convincing than its ethical counterpart. True, it presents non-factualism as a corollary of an attractive Humean metaphysical minimalism. But the blessed poor notwithstanding, economy is not a virtue in itself; it needs to keep company with sufficiency. So a proper appeal to metaphysical minimalism is at best one side of a more complex argument. The non-factualist has to show that probabilistic facts are unnecessary – that we can account for our talk of probability without construing it as talk about a distinctive feature of the world. Only thus will non-factualism be shown to have the conjunctive virtue of economical sufficiency.

### 4.3 PROJECTIVISM

Like the metaphysical position it serves, the argument for sufficiency is attributable to Hume. It turns on the claim that in addition to belief itself, there are other kinds of attitude one may hold to things in the world. There may be partial beliefs, and perhaps valuations and expectations of various kinds. However, we are inclined to treat these attitudes as if they were genuine beliefs. We thus mistake for a difference in the world what is really a difference in the head. Such a mistake may have pragmatic advantages – for example, it may enable us to improve our attitudes by reasoning and argument ‘about the facts’. But metaphysically speaking, it is a mistake all the same: the ‘facts’ we seem to find in the world about probabilities (or whatever) are actually the images or ‘projections’ of our own attitudes. We project these attitudes onto the world, and hence come to see it as containing corresponding realms of facts.

Perhaps the leading contemporary proponent of this Humean approach is Simon Blackburn, who calls the view ‘projectivism’. He emphasizes its economical advantages:

The projective theory intends to ask no more from the world than what we know is there – the ordinary features of things on the basis of which we make decisions about them, like or dislike them, fear them and avoid them, desire them and seek them out. It asks no more than this: a natural world, and patterns of reaction to it. By contrast a theory assimilating moral understanding to perception demands more of the world. Perception is a causal process: we perceive those features of things which are responsible for our experiences. It is uneconomical to postulate both a feature of things (the values they have) *and* a mechanism (intuition) by which we are happily aware of it. (1984, p. 182)

Projectivism thus claims much the same advantage as buying furniture in kit form. By doing some of the work ourselves, we reduce our initial commitment – financial or ontological, as the case may be. But in both cases it is possible to overdo things. If a tree is a furniture kit then kit builders become cabinet-makers. And unless the projectivist can distinguish the real beliefs from the ‘patterns of reaction’, then projectivism becomes a species of idealism – the view that all ‘facts’ are projections of our reactions to the world. This outcome obliterates the distinction at the very core of the non-factualist’s thesis. In the process, moreover, it overturns the projectivist’s claim to metaphysical economy. The idealist’s world can be as cluttered as anybody else’s: idealists differ from realists not so much as to what exists, as to what its existence amounts to.

To avoid the slide to idealism, the projectivist thus needs a core of genuine facts: a secure foundation, on which the explanation of higher-level appearances can be constructed. Blackburn talks of asking ‘no more from the world than we know is there’. But how do we know what is

really ‘there’, and what we put there? This issue has an important corollary: it means that projectivism itself is no guide to the facts. To arrest the slide to idealism, the projectivist needs to distinguish the genuine facts from the realm of ‘apparent’ facts for which the projective account is appropriate. So the possibility of a projective account of an area of discourse cannot in itself provide the motive for a non-factualist construal of that topic. To be sure, the projective account pays off a debt that the non-factualist incurs in urging such an account: it meets the obligation to explain the appearances. But the net effect is to balance the books. There is no profit in projective explanation.

The projectivist might object that this ignores the cost of the raw materials. Since Hume, projectivists have been motivated by metaphysical economy – by the suggestion that if we take ontological costs into account, then projectivism comes out ahead. But we can now see that the projectivist’s deal is too good to be true. If we can choose our raw materials, and less is better, then the best choice is none at all: that is, idealism. But idealism defeats non-factualism, so the projectivist cannot really accept that raw materials are a matter of choice. In other words, the economy projectivism offers us is ours already – or so the honest projectivist must believe. Like a good accountant, projectivism may show that we are better off than we thought, but it cannot take the credit for making us so.

What then would it take to arrest the slide to idealism? Have we perhaps underestimated the role of other metaphysical considerations, such as the argument from queerness? We shall see below that Blackburn himself offers another metaphysical argument for a non-factualist interpretation of moral judgement. These arguments bear on the wrong side of the issue, however. They lever the projectivist cart into motion, by rejecting factualist accounts of moral usage. But the non-factualist’s problem is to stop the cart, before it bears him to idealism. He needs an argument to the effect that certain topics *are* factual – and the better projectivism is made out to be at explaining factual appearances, the less likely it is that there can be any such argument.

The one live possibility seems to be to appeal to psychology. If the non-factualist could distinguish the genuine beliefs – in the head, as it were – from the various attitudes from which we project, then the barrier to idealism would be readily apparent. Projectivism would be safely contained, its task being to explain the fact that some non-beliefs are treated like beliefs. We shall return to the question of the availability of such a psychological distinction in chapter 5. But even if the non-factualist is entitled to such a distinction, the fact remains that projectivism provides no grounds for non-factualism. Projectivism is powerful medicine. Unless the need that it serves is grounded elsewhere, however, its effect is to make us idealists – which for the non-factualist counts as an overdose.

Let us turn then to the second major strand in the contemporary case for ethical and probabilistic non-factualism. This line of argument has a more overt link with psychology than the supposedly metaphysical arguments we have just discussed. Indeed, it might be thought to provide the psychological basis the projectivist seems to require. It focuses on the distinctive functional roles of probabilistic and moral judgements in our mental and behavioural lives – their significance in thought, speech and action. The notion of functional role might seem to provide the grounds for a distinction between beliefs and other propositional attitudes. I want to show that in appealing to aspects of the use of probabilistic and ethical judgements, non-factualists have taken for granted that the required psychological distinction can be characterized in these terms.

#### 4.4 FUNCTIONAL ROLE (I): PROBABILITY

Probability provides an important mode of qualification of ordinary empirical judgements. Empirical evidence is rarely conclusive. Ordinary judgements are rarely felt to be certainly true, to be justified beyond a shadow of a doubt. An interpretation of probabilistic sentential operators needs to be sensitive to the fact that such operators are used to register this feature of judgement, to indicate the *degree* to which a judgement is felt to be justified. True, we often make judgements with no such qualification, even if we are less than certain that they are true. But we are quick to forgo this convenience if the circumstances call for precision – if there is a lot at stake, say. Often then we resort to SPAs.

The fact that probability plays this role raises two sorts of problem for the factualist. For one thing it threatens a vicious regress: if probabilistic qualification is a more or less universal feature of accurate factual judgements, and probabilistic judgements are themselves factual, then accuracy seems to demand that in place of ‘It will rain’ we say ‘It will probably rain’; that in place of that we should say ‘It is probable that it will probably rain’; and so on. The regress is vicious because an interest in accuracy – an interest in asserting only what we are *justified* in asserting – will prevent us from resting content with any particular stage.

The factualist has a couple of options here. For a start, she could say that probabilistic judgements are a special case, and can be justified with certainty. The main problem with this is that, notoriously, the statistical rules of inference that are held to justify probabilistic conclusions are themselves probabilistic. A given range of frequency data will make it *likely*, but certainly not certain, that a particular probability has a specified value. (This is the source of concern about the circularity of statistical inference; see for example Braithwaite 1953, chapters 5 and 6.)

A much more promising suggestion relies on what we may call the *limiting case principle*: the principle that as the probability concerned becomes sufficiently high, a probabilistic judgement becomes ‘equivalent’ in some useful sense to the corresponding unqualified judgement. Given this principle, and the fact that in practice our first level probability judgements are usually themselves justified to a sufficiently high probability to allow us to apply it, it will follow that in practice we rarely introduce second (or higher) level probabilities.

The limiting case principle is so natural that it is easy to overlook its importance, and to fail to notice that for the factualist its justification is far from obvious – indeed, its justification is the second and more fundamental difficulty that arises, for the factualist, from the role of probability as a qualifier of ordinary empirical judgements. Clearly there is some important connection between the meanings of ‘It will rain’ and ‘It will probably rain’. Yet for the factualist these statements have different contents. The problem is to explain the connection between these contents, in such a way as to account for the limiting case principle, and hence make sense of the use of probabilistic qualifiers of ordinary judgements.

The non-factualist might explain these things as follows: probabilistic judgements express degrees of confidence, or degrees of partial belief. So ‘It will rain’ and ‘It will probably rain’ simply express different degrees of belief in the same proposition. And a series of judgements of increasingly strong probability that P, with limit 1, is thus associated with a series of increasingly strong partial beliefs that P, whose limit is the full belief that P. A judgement of high probability is practically ‘equivalent’ to the corresponding unqualified judgement in the sense that, and in so far as, a high degree of confidence in a given matter is practically equivalent to certainty. This explanation turns on the notion of the degree of a belief. Hence it is useless to the factualist, who is committed to explaining the use of probabilistic operators in terms not of differences of

degree but of differences in content. The factualist seems to require an (increasing) similarity of content between the belief(s) that it is (increasingly) probable that P and the belief that P.

This problem stems from the role of probability in judgement. A related problem arises from its role in decision making. Factualists and non-factualists agree that probabilistic judgements have certain typical behavioural manifestations, at least in certain straightforward cases, such as the choice of betting odds. For present purposes we can assume that both sides accept that Bayesian decision theory gives an approximate even if very idealized description of this process. The non-factualist again has a ready explanation: this behaviour is simply the characteristic (non-linguistic) behavioural display of partial belief. What Bayesian decision theory describes, in effect, is the functional role of a partial belief, in terms of its content and degree. (The more familiar story about the functional role of a full belief falls out as a special case.)

For the factualist, however, there is a problem: Why should the full belief that it is probable that P have the functional role that the non-factualist accords to a partial belief? Why should beliefs about probability be reflected in betting behaviour, for example? After all, the factualist will expect another sort of manifestation of the belief that it is probable that P: the sort of behaviour standardly associated with a full belief, given in this case a content of the form ‘It is probable that P’.

On the face of it, the factualist’s problem is thus to explain why probabilistic judgements should be accompanied by ‘corresponding’ degrees of confidence: for example, to explain why the belief that it is probably going to rain should be associated with a high degree of confidence, or strong partial belief, that it is going to rain. I shall call this *the confidence problem*. Note that a solution to the confidence problem would not only allow the factualist to explain the practical manifestations of probabilistic beliefs; it would also justify the limiting case principle, and hence explain probability’s role as a qualifier of empirical judgements.

Why should someone who believes that it is probable that P be confident that P? Factualists who notice the difficulty are inclined to appeal to rationality at this point. (See, for example, Braithwaite 1966, a paper entitled ‘Why is it reasonable to base a betting rate upon an estimate of chance?’) There are two difficulties with this move. First, the notion of rationality it depends on is problematic, to say the least. Its natural interpretations either themselves invoke probability, or cognate notions – ‘rational’ means ‘most likely to be successful’, for example – or they invoke some notion such as hypothetical limiting frequency, already invoked by certain of the more reductive factualist accounts of probability. Either way, the confidence problem remains: why should a belief about probability, or about hypothetical long run frequency, have a bearing on degrees of confidence, and hence on action in the single case. It is doubtful whether an account of the relevant notion of rationality can escape circularity of one of these two kinds. Notoriously, deductive inferences from probabilistic premisses yield probabilistic conclusions; while the non-deductive inference to degrees of confidence is what needs justification.

Secondly, it is far from clear that the appeal to rationality makes the connection strong enough to do justice to our intuitions. Faced with subjects who regularly failed to match their degrees of confidence, and hence their actions, to their claimed estimates of ‘probability’, I think that we would be inclined to say not that they were irrational, but that they did not mean what we do by ‘probability’ and related terms.

These factors have led some writers (e.g., Mellor 1971, p. 67) to the view that the inference from a full belief about a probability to the corresponding partial belief – let us call it *the downward inference* – is in some sense analytic. The simplest such account, taking to heart the intuition on

which the earlier suggestion was based, is what may be called the *rationalist* proposal: probabilities simply are degrees of rational belief, so that ‘It is probable that P’ simply means ‘It is reasonable to be confident that P.’ There is then no question of believing that it is probable that P, and yet doubting that it is reasonable to be confident that P.

However, it is important to realize that there is still a question as to why people who believe that it is reasonable to be confident that P are (on the whole) *actually* confident that P. What is it about the content of this belief (about what it is rational to believe) that leads people to adopt the corresponding degree of confidence? Of course, this is simply the original confidence problem, dressed up in new clothes – clothes it acquires in the rationalist’s paraphrase of a SPA.

At this point I think the rationalist has two possible moves. One is to attempt to explicate the notion of rationality, so as to display a decision procedure that justifies the (newly construed) downward inference, and in which ordinary self-interested speakers can plausibly be held to indulge. But this faces the very problems that, as we saw, plague attempts to show that the belief that it is probable that P makes it reasonable to be confident that P.

As in that case, the objection is perhaps not conclusive. But it does encourage another approach: to say that the use of the downward rule is constitutive of a grasp of the meaning of the relevant rationality ascriptions. A person will thus be said not to know what ‘It is reasonable to be confident that P’ means, unless in general when they profess to believe it, they are confident that P. Let us call this *contextual* rationalism: it takes the meaning of the term ‘rational’ to be bestowed by its inferential context.

The rationalist’s appeal to rationality in accounting for the meaning of SPAs may now seem redundant. Why not simply be a contextualist about probability itself – saying that the habit of making the downward inference is (partially, at any rate) constitutive of a grasp of the meaning of SPAs themselves – without making a detour via ascriptions of rationality? Either way, however, it can be seen that the confidence problem, concerning the typical *use* of the factualist’s full beliefs about probabilities, provides a strong motive for a particular account of their content: it favours the contextual theory. In effect, the factualist has been forced to retreat from ambitious reductive accounts of the meaning of SPAs, to the more modest and hence less exposed contextual position.

At this point, it may seem that the factualist has really conceded everything that was originally distinctive about her position. The factualist claims to take SPAs to express full beliefs about probabilities; and in response to the confidence problem, has taken pains to establish that in virtue of the content of beliefs about probabilities, such beliefs are typically accompanied by the corresponding partial beliefs. In effect, then, the factualist has acknowledged that the use of SPAs – indeed, any manifestation of this sort of single case belief about probabilities – is also an indication of a speaker’s possession of a corresponding partial belief. In consequence, the factualist seems vulnerable to a charge of theoretical redundancy. Beliefs about probability now seem idle cogs. Except in the factualist’s account of probabilistic assertion, all the functional work is done by the corresponding partial beliefs. So why not abandon that account of assertion? Why not acknowledge that SPAs are products of partial beliefs?

The factualist thus appears to face a dilemma. If there are full beliefs about probabilities, then the question arises as to their connection with degrees of confidence. If the connection is too loose, then beliefs about probabilities will not have the behavioural significance that everyone is agreed that they should have, if there are such things. If it is too strong, then full beliefs about

probabilities come to seem redundant – dispensable in favour of their constant companions, the corresponding degrees of confidence.

However, the factualist has one more card to play. The non-factualist's attack has rested on the distinction between the full belief that it is probable that P, on the one hand, and the attitude of strong partial belief that P, on the other. But why shouldn't the factualist simply reject this distinction, claiming that these two mental attitudes are one and the same – that a belief that it is probable that P just *is* a strong degree of confidence that P? Factualism would thereby take on the explanatory advantages of non-factualism, without an embarrassing duplication of psychological attitudes.

In other words, this 'radical' factualism challenges the psychological distinction on which the non-factualist has been relying. The non-factualist charges that factualist interpretations of probability are unable to make adequate sense of the functional role of probabilistic judgement – the role of what the non-factualist calls partial belief. The radical factualist responds by helping herself to the non-factualist's functional account; but insisting that it be regarded as a characterization of the distinctive content of probabilistic beliefs. The prospects for non-factualism again turn on its ability to distinguish genuine beliefs from other psychological attitudes.

#### 4.5 FUNCTIONAL ROLE (II): VALUES

The confidence problem has an analogue in the ethical case. Indeed, it is perhaps the most venerable argument for a non-factualist treatment of moral judgements, tracing its origins at least to Hume. It relies on the principle that, as Hume puts it, 'reason alone can never be a motive to any action of the will; ... it can never oppose passion in the direction of the will' (*Treatise*, Book II, Part III, Section III; p. 413 in Hume 1978). In more modern terms, Hume's view is thus that beliefs by themselves are incapable of motivating action. Beliefs need to be supplemented by some desire or 'pro-attitude'. (Since precision won't matter, and I have a con-attitude to 'pro-attitude', I'll continue to talk of 'desires'.)

The non-factualist wants to show that SEAs do not express beliefs. Given Hume's principle, it is enough to show that the attitudes expressed in moral judgements are capable of motivating action; for this puts them on the desire or non-belief side of Hume's distinction. And the motivating character of moral judgement (or indeed of evaluative judgement in general) has usually seemed self-evident.

The non-factualist thus offers the following argument: moral commitments – the attitudes expressed in moral judgements – have a distinctive conceptual role. Briefly, they motivate actions. Hume's principle tells us that no belief, of itself, could fulfil this role. And hence the factualist, who takes SEAs to express beliefs, has to resort to the claim that such beliefs are 'accompanied by' attitudes that can play this role. The factualist faces the moral equivalent of the confidence problem – we might call it *the approval problem* – that of explaining, for example, why the belief that it is a good thing that it is raining should be associated with an attitude of approval towards the fact that it is raining. On the face of it, the non-factualist will suggest, it would seem that 'if moral commitments express beliefs that certain truth conditions are met, then they could apparently co-exist with any kind of attitude to things meeting the truth conditions' (Blackburn 1984, p. 188).

Hume's principle is clearly crucial. Someone who rejects it can agree that moral judgements express mental states on the active or motivating side, and yet insist that moral commitments

simply are beliefs about moral facts. This will be the ethical equivalent of the radical factualism we have just encountered in the probabilistic case. In both cases the non-factualist has attempted to burden the factualist with a cumbersome and implausible duplication in the mental states associated with utterances whose status is in question (SPAs and SEAs, respectively). The radical factualist responds by accepting the model but denying that it involves any duplication: rather, she suggests, the non-factualist has confused the existence of two ways of describing a single mental state for the existence of two distinct mental states.

We shall return to Hume's principle in chapter 5. Before that, however, I want to show the relevance of these issues of functional role to another prominent metaphysical argument for ethical non-factualism. I want to show that the argument cuts no ice with the radical factualist; and also, incidentally, that it has an analogue in the probabilistic case.

#### 4.6 THE SUPERVENIENCE ARGUMENT

Moral properties appear to supervene on natural or non-moral properties: it seems uncontroversial that if two things are taken to be the same in all non-moral respects, then they should not be distinguished in moral terms. The problem is to explain this evident constraint on moral judgement. Simon Blackburn has argued that only a non-factualist has an adequate explanation. I want to show that the radical factualist does just as well.

Blackburn first presents the argument in a paper entitled 'Moral realism' (1971). He returns to it, changing the emphasis to some extent, in (1984) and (1985). Blackburn says, for example, that the explanation of supervenience is

especially hard for the realist [our factualist]. For he has the conception of an actual [moral] state of affairs, which might or might not distribute in a particular way across the [non-moral] states. Supervenience then becomes a mysterious fact, and one which he will have no explanation of (or no right to rely upon). (1984, p. 185)

As far as realism can show us, it could be true that the moral floats quite free of the natural. (1984, p. 221)

Blackburn wants to show that the non-factualist (or anti-realist) can explain the supervenience of the moral on the non-moral. In his first version of the argument he suggests that a

natural way of explaining the supervenience of moral properties would, for an anti-realist, be something like this. There can be no question that we often choose, admire, commend, desire, objects because of their naturalistic properties. Now it is not possible to hold an attitude to a thing because of its possessing certain properties and, at the same time, not hold that attitude to a thing which is believed to have the same properties. The non-existence of the attitude in the second case shows that it is not because of the shared properties that I hold it in the first case. (1971, p. 114)

Blackburn acknowledges that the factualist 'need not refrain from talking about moral attitudes': on the contrary, 'it is just that [the factualist] thinks this is a less clear alternative to talking about moral beliefs' (1971, p. 115). But he thinks that 'if we rephrase the preceding paragraph in terms of moral belief, it is obviously insufficient to explain supervenience' (p. 115). For the claim would be that

it is not possible to hold a moral belief about a thing, believe a second to be exactly alike in all naturalistic respects, yet at the same time not hold the belief about the other thing. But this doesn't explain supervenience at all: it merely shows the realist putting conditions upon what can be *believed* to be the truth, not upon what *is* the truth. Our belief, he is saying, has to be consistent across naturalistic similarities – but this is no explanation of why, on his theory, the truth has to be. (p. 115)

Blackburn's claim thus seems to be that the factualist has trouble justifying moral consistency – justifying the policy of applying the same moral predicates in the same non-moral circumstances. The non-factualist is better placed not because he can justify such a policy, but because he doesn't need to. The need for justification stems from the factualist's conception of truth – in particular, from the principle that the truth is independent of what it is possible for us to believe. Both sides have a more or less causal explanation of the fact that we follow the consistency principle: if moral attitudes are the effects of non-moral attitudes, consistency in practice is a consequence of the relevant causal laws. For the factualist, however, *de facto* fidelity is not enough. Consistency needs to be earned.

So the non-factualist escapes the supervenience problem – as construed in terms of consistency, at least – by rejecting the framework within which it arises. The problem arises because we talk of independent moral truths; and the non-factualist is free to reject that sort of talk.

Not all non-factualists will want to exercise this freedom, however. In particular the projectivist, far from rejecting factual talk, wants to account for it in non-factual terms. Blackburn repeatedly stresses this aspect of projectivism, regarding its explanatory potential as a major virtue. In particular, he thinks that the projectivist can explain our entitlement to doubt our own opinions, and to distinguish between our believing that something is the case and its actually being the case (Blackburn 1984, p. 219, for example). If so, then the projectivist is equally vulnerable to the charge that the causal explanation of consistency simply puts 'conditions on what can be believed to be the truth, not upon what is the truth'.

Whether the projectivist has an answer to this charge will no doubt depend on the details of the constructive account of truth, in terms of which he proposes to explain the factual appearances. But if the projectivist can secure supervenience by this means then what is to stop a factualist from doing the same? In the (1971) version of the argument, Blackburn takes it for granted that factualism arrives with its own account of truth. His target is the traditional realist, whose notion of truth is anchored to a conception of a 'real' external world. For present purposes, however – as to a large extent in Blackburn (1984) – the projectivist's more threatening opponent is the idealist. As noted earlier, idealism should here be counted a version of factualism; for it rejects the non-factualist's attempt to divide indicative discourse into fact-stating and non-fact-stating parts. And a factualist of this sort will have no qualms about helping herself to the projectivist's account of moral supervenience. Thus we find once again that the non-factualist's main problem is to stop the projectivist bandwagon – and that the better projectivism starts, the harder it is to stop.

Perhaps for this reason, the consistency issue gets less attention in Blackburn's more recent presentations of the supervenience argument. Of course, no version of the argument can hope to exclude idealism. As an argument for non-factualism, it simply bears on the wrong side of the issue. But at least the damage might be controlled: we might hope to present the argument in such a way that it favours projectivism about ethics without equally favouring projectivism about everything. Alternatively (though this would be less to Blackburn's own tastes) we might hope for a version of the argument that would favour a non-projectivist non-factualism about ethical

discourse. It turns out, however, that the prospects of either course depend on our ability to refute the radical factualist.

Blackburn now offers the projectivist the following explanation of moral supervenience:

When we announce the [moral] commitments we are projecting, we are neither reacting to a given distribution of [moral] properties, nor speculating about one. So the supervenience can be explained in terms of the constraints upon proper projection. Our purpose in projecting value predicates may demand that we respect supervenience. If we allowed ourselves a system (shmoralizing) which was like ordinary evaluative practice, but subject to no such constraint, then it would allow us to treat naturally identical cases in morally different ways. This could be good shmoralizing. But that would unfit shmoralizing from being any kind of guide to practical decision-making (a thing could be properly deemed shbetter than another although it shared with it all the features relevant to choice or desirability). (1984, p. 186)

This might be taken as a pragmatic justification of the consistency principle – as if Blackburn were suggesting that in view of the role of moral judgements in influencing our practical decisions, it would be unwise, though not impossible, not to make one's moral judgements in accordance with that principle. But there is a more interesting and I think more accurate interpretation. We have noted that moral beliefs, indeed evaluative beliefs in general, appear to play a special role in guiding choice. A (positive) value judgement about a state of affairs is a *prima facie* reason for choosing that state of affairs over others not so valued. As Blackburn says, 'It seems to be a conceptual truth that to regard something as good is to feel a pull towards promoting or choosing it, or towards wanting other people to feel the pull towards promoting or choosing it' (1984, p. 188). To value A higher than B is to be disposed, other things being equal, to choose A in preference to B.

In order to explain supervenience, we need to add to this the plausible claim that in so far as they affect us, the outcomes of actions are completely determined by their natural or non-evaluative consequences. The value judgement guides our choice of destination, but the success or failure of the expedition depends on what we actually find when we get there. (I am simplifying here. In practice, the route from valuation to natural consequences of choice may be multi-staged: as, for example, when we judge that it is better to take the high road than the low road, because the scenery is more attractive on the high road.) This means that to imagine a case of moral difference without natural difference would be to imagine on the one hand that choice A would be preferable to choice B; but on the other that since the two choices have the same natural consequences, it would make no difference which we choose. This conceptual conflict seems a plausible basis for our belief in moral supervenience – indeed, for the supervenience of the evaluative on the non-evaluative in general.

If this explanation is on the right lines, then the supervenience of the evaluative on the non-evaluative depends on the principle that outcomes are judged, in the end, by their natural consequences. This principle might conceivably be challenged. On theological grounds, for example, someone might say that the outcome of an action can have an intrinsic value that is independent of its natural consequences (at least in this world). But this sort of challenge in fact works in favour of the above explanation. To admit intrinsic values of this kind would surely be to deny supervenience – to allow values to 'float free' of the natural, in Blackburn's phrase. However, the above explanation of supervenience turns on nothing more than the action-guiding role of evaluative attitudes. We already know that the factualist has a problem explaining this – it is the approval problem, the evaluative analogue of the confidence problem. But we saw

that if the factualist is prepared to claim that evaluative attitudes simply are beliefs, the problem seems to be dealt with. And in the present context, the same identification will entitle the factualist to this explanation of evaluative supervenience. So it seems that the supervenience problem is not an objection to the radical factualism we have encountered in this chapter: a factualism which accepts the non-factualist's positive account of the mental states characteristically expressed in evaluative judgements, but insists that the mental states so described are themselves the beliefs whose existence the projectivist denies.

In his most recent discussion of the supervenience argument, Blackburn says that the explanation of supervenience depends

crucially upon the role of moralising being to guide desires and choices amongst the natural features of the world. [But] if, as a realist ought to say, its role is to describe further, moral aspects of [reality], there is no explanation at all of why it is constitutive of competence as a moralist to obey the [supervenience] constraint. (1985, pp. 56- 7)

Our radical factualist agrees that the explanation of supervenience lies in the fact that moralizing has the former role; agrees also that it has the latter role; but denies that the two roles are incompatible.

#### \*4.7 SUPERVENIENT PROBABILITY

Blackburn presents the supervenience argument as a problem for the moral realist. In view however of the parallels we have already noted between the ethical and probabilistic cases, it is natural to wonder whether there an analogous argument in the probabilistic case. Do probabilities supervene on non-probabilistic characteristics of things? I want to show that there is a parallel here, that again the best explanation turns on the action-guiding role of the relevant sort of attitude, but that here also the radical factualist has an answer.

Suppose then that we accept that there is a certain probability of a specified outcome to a given event – let us say an  $n\%$  chance that a given experiment will result in the emission of an electron. Do (or should) we allow that, with all non-probabilistic facts held fixed, there could have been some different chance of this result? The answer depends in part on our view of chance. If we think of the chance in terms of the actual frequency of such outcomes in similar experiments, or some such, the appropriate answer seems to be 'yes'. If we are talking about actual class ratios in finite classes, we have no difficulty accepting that those ratios could have been different, even if other aspects of the world had been the same. The possibility is illustrated by the man who broke the bank at wherever it was: because his fortunes depended on indeterminate events, he could equally well have lost his shirt.

Partly for this reason, however, actual frequencies have never provided a plausible approach to single case probabilities. It is more natural to suppose that the chance that an experiment will have a certain result is a further fact, over and above the facts about frequencies in actual similar cases. Could then this further fact vary independently of the other properties of the experiment – including, now, the actual relative frequencies themselves?

I think that we are inclined to deny that chances could vary in this way. The reason is nicely analogous to the above explanation of the supervenience of the moral on the non-moral. Suppose for example that two experimental set-ups are identical in (what we take to be) relevant non-probabilistic respects; and, most importantly, that they yield the same relative frequencies of corresponding outcomes. Let us try to imagine that the two set-ups in fact have different

chances of some outcome – electron emission, say. In set-up A there is a high chance of emission; in set-up B, a low chance. The problem with such a supposition stems from the role of probabilistic beliefs in decision making. Roughly speaking, the higher the probability of an outcome, the more reasonable it is to bet on it, at any given odds. This means that on the given assumptions, we shall be inclined to prefer a bet on electron emission in set-up A to the same bet in set-up B. However, since there is no difference in the relative frequency of such an outcome in the two cases, a gambler who consistently bets on electron emission will do no better in case A than in case B. So there is no reason to prefer one bet to the other. This contradiction seems to show that chances do supervene – in the same sense as in the ethical case – on non-chance factors.

As it stands, however, the argument is vulnerable to a serious objection: namely that the fact that the two cases show the same relative frequency of electron emission does not entail that the same betting policy will have the same outcome – let alone that a *single* bet would have the same outcome in both cases. For individual outcomes may differ, even if the overall frequencies are the same. Of course, if we could take for granted that identical relative frequencies entail identical probabilities, then we could say that the *probability* of success is the same for both bets (either in the single case or in the long run). But to take this for granted is to assume what we are trying to establish: the supervenience of probabilities on the non-probabilistic.

To escape this objection we have to strengthen our premisses. We need to suppose that set-ups A and B are non-probabilistically identical in such a way as to justify the following claim: Any trial on either set-up *would have had* the same outcome if performed on the other. It then follows that in any given case (and at any given odds), a bet on electron emission would have the same result on either set-up. And this conflicts with our preference for a bet on A, stemming from the assumption that there is a higher chance of electron emission in A than in B.

Thus, on sufficiently strong assumptions, we find that single case probabilities do supervene on non-probabilistic features of the world. As in the moral case, the argument makes essential use of a special association between a certain sort of mental attitude – here (what the factualist will regard as) a belief with probabilistic content – and choice behaviour. We saw that this association may itself seem problematic for the factualist: it is the basis of the confidence problem. However, the radical factualist's reply was that the so-called 'association' is actually an identity. We now find that in the probabilistic as in the ethical case, such a strategy is not to be dislodged by the observation that the disputed 'facts' appear to supervene on facts of other kinds. The radical factualist explains this observation as well as anybody.

## 5

**Fact and Psychological Function**

We found that in the ethical and probabilistic cases, non-factualism can be driven to seek its foundations in the head – to try to analyse in psychological terms the semantic distinction between fact-stating and non-fact-stating discourse. Statements thus come to be characterized as (what are normally) expressions of belief; non-statements – or those that the non-factualist is concerned with, at least – as the (normal) expressions of other sorts of propositional attitude. And this appeal to psychology is no mere expository gloss. In order to defeat the radical factualist, the non-factualist must claim to be able to distinguish beliefs from other attitudes, in advance of settling the semantic issue.

We have not shown that the non-factualist's dependence on psychology goes beyond the ethical and probabilistic cases. All the same, it is clear that non-factualists about various other topics do in fact appeal to psychology. One important case is the non-factualist approach to indicative conditionals. This relies on the observation (originally due to Ernest Adams 1965) that the subjective assertibility of an indicative conditional goes by the subjective conditional probability of its consequent given its antecedent; rather than say by the absolute probability of the material conditional. The view thus involves that indicative conditionals are typically expressions of high subjective conditional probabilities; and – crucially, if the account is to be non-factualist – that these cannot be construed as a kind of belief. Another case is the so-called 'epistemic' treatment of possibility and necessity, and the related treatment of laws of nature as generalizations towards which we hold a certain sort of attitude (a view advocated by Ramsey 1978, chapter 6B, for example).

I think we are thus entitled to conclude both that non-factualism has often in fact put its faith in psychology; and that at least in certain central cases it is bound to do so. Our final sceptical task is to show that this faith is misplaced. Beliefs cannot be adequately distinguished from other propositional attitudes; not at any rate unless we are prepared to rely, as the non-factualist here cannot, on a characterization of this psychological distinction in semantic or metaphysical terms.

The chapter is in three main parts. I first consider the Humean approach to the belief/desire distinction, based on Hume's account of the origins of action. This account continues to be widely taken for granted, even in these anti-Humean times. As a result, the task of explaining action in psychological terms may seem a promising source of the sort of distinction we are looking for. I shall argue that the promise is empty, however. In search of the source of the belief/desire distinction the modern philosopher ends up, as Hume himself did, appealing once more to semantics: claiming that beliefs, unlike desires, have truth conditions.

The second part of the chapter reaches a similar conclusion along a more general, more contemporary and perhaps initially more promising route. It considers the suggestion that we look to recent cognitive theory for an account of the *content* of a propositional attitude. A species of propositional attitude is a class of beliefs if and only if there is a class of contents such that the species concerned is the set of beliefs with contents of that class. (This is the psychological analogue of the condition for a species of utterance to be represented as a class of assertions.) Hence a basis for the belief/non-belief distinction might be expected to be a corollary of an adequate account of content. I shall argue, however, that contemporary accounts of the nature of content do nothing to settle this issue – or rather they settle it, if at all, in the semantic or metaphysical terms that the non-factualist has already renounced.

Finally, I shall turn to some arguments claimed to show that indicative conditionals are not expressions of genuine beliefs. In view of the logical role of conditionals, this is potentially the most significant non-factualist theory of all. My approach will again be deflationary: I shall try to show that these arguments parallel some striking but clearly inconclusive objections to a factualist treatment of ordinary probabilistic judgements.

### 5.1 DISTINGUISHING DESIRES: THE HUMEAN STORY

The Humean argues that reason is incapable of action – in modern form, that beliefs are motivationally inert – and hence that an action explanation is incomplete unless it appeals, at least implicitly, to some sort of desire or motivating attitude. The view is as orthodox as it was once controversial. Indeed, it is at the heart of so-called ‘folk psychology’ (a fact that should not be uncritically attributed to the folk influence of Hume, one suspects). I want to show that the Humean argument is inadequate, however, even in its most charitable contemporary reconstruction. It fails to show that belief cannot motivate action; and hence it fails to show that moral commitments (even if agreed to motivate actions) are not beliefs.

Our interest in the Humean argument stems from its application to these semantic questions. But it may be as well to point out that in challenging the principle that belief is non-motivational, we are not committed to rejecting totally the orthodox view of the explanation of action. For in one crucial respect Hume’s conclusion is stronger than the orthodox belief/desire model requires. It would be in keeping with the spirit of this view to say that the explanation of action involves two sorts of belief: beliefs about the desirability, or value, of certain states of affairs; and beliefs about the bearing of the action in question on the obtaining of those states of affairs. This modified view could agree, in any actual case, on what constitutes an explanation of an action. It differs from the Humean only as to how the attitudes mentioned by such an explanation should be described.

This difference is crucial to the use made of Hume’s principle by non-factualists in ethics. The non-factualist needs to show that moral statements lie on the right side of a psychological distinction – in other words, that they do not express beliefs. Given Hume’s principle, it only needs to be shown that moral commitments do motivate actions; for this puts them on the desire side of Hume’s distinction. Clearly, however, someone unconvinced by the Humean argument could quite well endorse the standard model of action explanation; accept that moral judgements express mental states on the active or motivating side; and yet insist that moral commitments are genuine beliefs about moral facts. This was the response of the radical factualist we met in chapter 4.

Such a view admits two possible accounts of the place of desire. One account would recognize a category of desires, distinct from beliefs, but claim that motivation is not the exclusive preserve of attitudes of this category (moral commitments being one sort of exception). The problem will then be to say what distinguishes beliefs from desires, given that it is not the power to motivate action. This problem might incline us to the second sort of account, recognizing no significant foundation for a distinction between beliefs and desires. This view would further divide into two possibilities: the substantial claim that all desires are beliefs; and the weaker sceptical claim that there is simply no adequate justification for not so representing an arbitrary propositional attitude. Our interests again lie with the sceptic.

Let us turn then to the case for Hume’s principle. Why should not beliefs be motivational? The Humean’s first line of attack might be something like the following argument. It is intended to show that the belief that a contemplated action A would have some property F (or would bring

about an outcome having property F) cannot in itself constitute a reason for an agent to perform A. If there were such a case (the argument claims) we could imagine a similar agent who would dislike F, and would hence be motivated not to do A by a desire that not F. This being so, the action A cannot in fact have been caused by or be explained in terms of the agent's belief that A would be F. It is also relevant that the agent desired that F, or at least did not desire that not F. In other words, the causal or explanatory story needs to mention the agent's desires.

I shall call this *the imagined dislike argument (IDA)*. Hume himself appeals to IDA, or something very like it, in this notorious passage:

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. (1978, p. 416)

Hume's immediate point is that reason does not control the passions. But given that the passions do control action, the argument becomes IDA – as indeed the latter example illustrates, in speaking of choice rather than preference. Hume's suggestion seems to be that the belief that an action will prevent his total ruin will not explain his performing that action, because he might have desired total ruin (in preference, at any rate, to some other consequence of preventing ruin, such as the unease of another party).

IDA has two major flaws, however. The first is to rely on the principle that if a causal explanation would be inadequate in the *presence* of some extra condition, then it is inadequate unless it mentions the *absence* of that further condition. This principle is no more plausible here than it is for causal explanations in general. In almost every case, causes are at best sufficient in *the circumstances* for their effects; and to insist on an explicit description of the relevant circumstances would be to cry for the moon. (We rely on *ceteris paribus* clauses because we cannot fill in these gaps.) In the present case, to admit such a principle would be to abandon any systematic approach to the psychological explanation of action. After all, there is no end to the things that even the most fastidious agent does not dislike.

IDA's second fault is just as damning: given that the issue is whether desires are anything but a special class of beliefs, the argument simply begs the question. For if some beliefs are desires, to believe that an outcome would be F might be the very same thing as desiring that outcome. In this case the possibility that the argument imagines – that of someone holding the belief, but desiring an outcome such that not-F – would be that of a person who desires what he takes to be undesirable. An advocate of the view that desires are a species of belief seems entitled to regard this as no possibility at all.

Neither of these objections is particularly profound. So it may seem doubtful whether IDA does justice to Hume's distinction, or explains its contemporary appeal. However, I think the argument does point us in the right direction. It seems to derive its initial plausibility from a model of the mind which we find intuitively appealing. Perhaps it thus assumes what it is supposed to establish. But at least it has the virtue of directing us to the crucial assumption.

We may think of the issue in terms of a simple model of the mental structure of an intentional agent. The functional role of desire seems to be to connect two other components: on the one hand, a unit (*Cog*) responsible for cognitive deliberation; on the other, a unit (*Will*) whose function is to initiate bodily action, in response to the outputs from *Cog*. There are several possible models of *Will* and its relationship to *Cog*. At one extreme we could regard it as a kind of store, or processing house, for intentions already formed in *Cog*. At the other extreme we

could take it to be the unit in which judgements arriving from Cog are matched with a store of desires to yield intentions. However, since we want a model in which the action that interests us takes place at the boundary between Cog and Will, let us avoid both extremes. Let us view Will as a structure on which intentions ‘condense’, as a consequence of suitable states of Cog.

In these terms, Hume’s principle is that no intention can condense in the absence of a desire – that desires are a necessary catalyst to the intention forming process. IDA rests on the claim that whatever beliefs might be present in Cog, we could imagine a desire whose presence would prevent a given intention condensing from those beliefs. The argument is a poor one, but catalytic view of desire has a lingering appeal.

However, we know that in the chemical case, altering either the composition of a vapour or the nature of the condensation chamber may make a catalyst redundant. So the analogy suggests that we consider the effect of changes in the corresponding aspects of the mental model: to Cog on the one side, and Will on the other. As yet we have no reason to regard these factors as cognitive constants. So there is nothing to stop us trying to imagine agents who differ in these respects, from each other, and more importantly from the way the Humean view takes us to be. The appeal of the orthodox view may rest on the fact that we have not yet noticed the alternatives.

There are two main possibilities to consider: the causal gap supposedly filled by desires might be bridged from either side. The first possibility would be to expand Cog, so as to admit beliefs with the capacity to produce intentions in a Will of the orthodox kind. The second would be to endow Will with the disposition to form intentions in the presence of certain orthodox beliefs. (I ignore the obvious hybrid possibilities.)

In the former case, we want a belief whose capacity to rationalize an action does not depend on an agent’s possession of certain desires. It seems to me that we can get a good idea of what such a belief would be like, by reflecting on familiar expressions of moral or rational necessity – ‘I am bound to tell the truth’, for example, or ‘I must go down to the sea again.’ Superficially, at any rate, utterances such as these can be regarded as expressions of beliefs that possible actions have certain properties. Moreover, such beliefs (if they really are beliefs) seem to provide reasons for actions which are insensitive to an agent’s desires about actions with the corresponding properties. The belief that one must go down to the sea seems capable not merely of making one do so, but of overriding, in the process, considerable distaste for messing about in boats.

These points are notoriously disputable. There are two lines of attack. One will urge that believing that I must do something, or am bound to do it, only provides me with a reason for doing it because, like everybody else, I have an implicit desire to do what I must. The other will take the opposite tack and argue that because utterances such as these express reasons for actions, they do not express beliefs. However, both lines of argument seem to depend on the very claim in question: namely, that beliefs alone cannot motivate actions. To settle the issue, we would need a criterion with which to distinguish beliefs from desires. Hume’s argument claims to offer such a criterion, saying that desires are distinguished in virtue of their active or motivating role in the production of action. But it now appears that the argument itself depends on the very criterion it is intended to supply. As McDowell puts it,

the idea of the world as motivationally inert is not an independent hard datum. It is simply the metaphysical counterpart of the thesis that states of will and cognitive states are distinct existences; which is exactly what is in question. (1978, p. 19)

Thus it seems that we have no non-circular reason to deny the possibility of an intrinsically rationalizing belief: a belief whose capacity to motivate and rationalize a contemplated action does not depend on co-present desires (even if its motivational effect could be overridden by certain such desires, were they to be present). And if we have reason not actually to affirm the possibility, it lies partly in the fact that the distinction between the belief and desire is itself so obscure.

I said that there are two possible ways to bridge the claimed gap between Cog and Will. The second is to imagine Will itself equipped to handle a somewhat broader range of inputs. Why shouldn't Will, or the process by which an agent forms intentions to act, simply be built to respond to certain beliefs about the properties of actions and states of affairs?

Consider, for example, a classical economic rationalist, whose actions are based entirely on the profit motive. Such an agent forms intentions to perform those actions he believes will maximize his net monetary return. One might object that the rationalist *wants* to make a profit. But why should he? Why should his mind not simply be organized so that believing that an action would maximize his profits normally leads him to form the intention to undertake it (in just the way, and to whatever extent, belief/desire combinations are ordinarily supposed to lead to intentions)?

One possible answer appeals to Davidsonian themes. We ourselves wouldn't be motivated to act as the rationalist does unless we wanted to make a profit, so we can only make sense of his behaviour – rationalize it – by ascribing to him such a desire. This objection might lead us too far astray. In making the ascription of attitudes to cognitive systems radically dependent on the ascriber's own system of attitudes, it threatens the entire project of speculative psychology. Fortunately, we can evade it. We want to know, *inter alia*, whether our own Wills are ever activated by beliefs alone. But in the face of a purported real case of this sort of thing, the Davidsonian move will simply be powerless. It cannot insist that *our* rationalization of the action in question would need a desire, for that will be the very point at issue.

I conclude that motivating beliefs are a conceptual possibility. If we begin with an orthodox Humean model of ourselves, we can imagine two kinds of psychological modification that would allow beliefs to motivate actions. The effect of this is not simply that the practical question of the nature of the motivation of action remains unsettled, but rather that the question itself is under threat. Hume's principle not only entails that action requires desires, as well as beliefs; it also offers a criterion for distinguishing desires from beliefs. Without it, the non-factualist lacks a foundation for that distinction.

Like Hume himself, the non-factualist may therefore feel inclined to try another tack – to say that desires are distinguished from beliefs in virtue of the fact that it is 'impossible ... they can be pronounced either true or false' (1978, p. 458). This would be to distinguish beliefs in virtue of their semantic properties; and it is a popular move. However, we shall see that it takes the non-factualist back to positions already abandoned – positions whose untenability forced the non-factualist to seek a psychological basis for the fact/non-fact distinction.

This may seem an immediate consequence of such an appeal to semantic properties. After all, the attempt to analyse assertion in terms of truth was the first target of our sceptical challenge. For the moment, however, a significant possibility remains open to the non-factualist. Psychological truth might succeed where linguistic truth failed. If semantic properties apply primarily to mental states, and only derivatively to language, then the arguments of chapter 2

may have missed the mark. Truth might indeed be insubstantial, but nevertheless play a crucial role in establishing a psychological basis for the fact-stating/non-fact-stating distinction.

To exclude this case, we shall need to consider some possible foundations for psychological semantics. This brings us to the concerns of contemporary cognitive theory. And it brings us to another possible response to the argument of this section. We have seen that the requirements of action explanation do not yield the psychological distinction the non-factualist is looking for. On reflection, however, this avenue was always rather limited. At best it offered a distinction between beliefs and the range of motivating attitudes we have loosely called ‘desires’. Even if this had met the needs of the ethical non-factualist, it would have been largely irrelevant to other non-factualists. If beliefs do not motivate actions, then neither, for example, do the partial beliefs invoked by the non-factualist account of probability.

From this point of view, our mistake may seem to have been to concentrate on too small an aspect of the functional organization of the mind – to concentrate on the initiation of action, to the exclusion of other functions of the mind. As it happens, this is exactly the mistake of which contemporary functionalist accounts of the mind accuse their behaviourist predecessors. In the next section we want to enquire whether these contemporary theories succeed where Hume failed, in distinguishing beliefs from non-beliefs.

## 5.2 FUNCTION AND COGNITIVE CONTENT

We may seem to be faced with a daunting diversity of contemporary views on the nature of the mind. One of Professor Fodor’s relatives ‘sighs for the days when well-brought-up philosophers of mind kept themselves occupied for hours on end analysing their behavioural dispositions’ (Fodor 1985, p. 76). But things are not as bad as they seem. For one thing, as Fodor puts it, ‘the chaotic appearances are actually misleading. A rather surprising amount of agreement has emerged, if not about who’s winning, at least about how the game has to be played’ (1985, p. 76). We can exploit this agreement – and Fodor’s guide to the options – for our present purposes.

More importantly, however, it turns out that there is very little to say: none of the contemporary options tries to analyse the distinction we are looking for, or indeed makes a point of claiming to have the resources for doing so. Most contemporary theories simply take for granted the Humean model of action explanation – in the guise, as they put it, of ‘folk psychology’. This provides the framework within which they tackle such questions as the nature of content. And those that do not assume the belief/desire model of action explanation are those that reject folk psychology, and with it the apparatus of propositional attitudes. So on neither side is there a developed claim to provide a basis for the distinction between beliefs and other sorts of attitudes; nor is there any prospect of such a basis.

This claim may seem implausible. After all, we know that we can think of the question as to whether desires (say) are a kind of belief, as that of the existence of an appropriate species of *content*. Roughly, desires can be represented as beliefs if and only if there is a content modifier *Des* such that the desire that P can be represented as the belief that *Des(P)*. And cognitive science has had a great deal to say about the notion of content. Surely there is something that will tell us whether there could be such a *Des*?

It seems not. The point is perhaps most clearly illustrated in what Fodor calls ‘the Representational Theory of Mind’ (*RTM*), according to which propositional attitudes are relations to representations in a ‘language of thought’. Fodor gives the following ‘canonical formulation of this view’:

For any organism O and for any proposition P, there is a relation R and a mental representation MP such that: MP means that (expresses the proposition that) P; and O believes that P iff O bears R to MP. (And similarly, R desires that P iff O bears some *different* relation R', to MP. And so forth ...) (1985, p. 88)

Clearly, the distinctions between the propositional attitudes are here built in from the start: the desire that P is interpreted in terms of the relation R' to MP, rather than in terms of the relation R to DES(MP) (DES being the mental representation of the content modifier *Des*).

From the Humean point of view this looks perfectly natural. If we question the Humean categories, however, it begins to seem rather arbitrary. For mental representations and the attitude relations R, R', etc., are both individuated in the same kind of way: in terms of their causal or functional roles. One of the great attractions of the RTM is its ability to exploit the computer metaphor, and conceive of mental processes as 'transformations of mental representations' (Fodor 1985, p. 95). Such transformations are causal processes whose outcomes depend on two factors: the syntax of the mental representations with which the process begins, and the mode in which those representations are initially stored – whether they are 'believed', 'desired' or whatever. But why not reduce these two factors to one, by regarding the mode of storage as an aspect of syntax? After all, in any actual case the available syntax will be capable of representing an enormous range of distinctions. It is surely capable of a handful more.

It is true that if we knew the syntax of the language of thought – in advance, as it were – then the question of representing mode of storage as a syntactical distinction would not arise. The syntactical guidebook would tell us that there was no way of representing certain distinctions within the language; leaving us no choice but to regard their role as that of a mode of storage of a sentence of the language. The non-factualist cannot hope for an answer of this kind, however. For what is the semantic role of the sentences of the language of thought? Simply that of the fact-stating sentences of natural language – or so the non-factualist must say, for it is these sentences he takes to express beliefs. In other words, the task of delineating the syntax of the language of thought is simply the original task of delineating fact-stating discourse: the very task for which the non-factualist now wants to distinguish beliefs from non-beliefs.

So the non-factualist is not entitled to appeal to the syntax of a language of thought, in order to distinguish beliefs. Indeed, no one is entitled to appeal to the syntax of a language of thought in order to determine which functional roles should be counted content-determining and which attitude-determining. For (as Mellor 1988 emphasizes) semantics must precede syntax, in this sense: our grounds for regarding a system of states and causal processes as a syntactic system of representations and transformations must be that we assign its elements a semantic role. The title 'linguistic' has to be earned.

There is little agreement on how it should be earned. Fodor again: 'Of the semanticity of mental representations we have, as things now stand, no adequate account' (1985, p. 99). It seems to me that proponents of the RTM have not always been sufficiently aware of the extent to which it depends on the assumption that this problem is solvable, at least in principle. For how else could an opposing view – that of Stich (1983) – present itself as a *purely* syntactic account? Stich says that the

core idea of the STM [Syntactic Theory of Mind] – the idea that makes it *syntactic* – is that generalizations detailing causal relations among the hypothesized neurological states are to be specified indirectly via the formal relations among the syntactic objects to which the neurological state types are mapped.' (1983, p. 151)

But how is this different from a formal description of (an aspect of) the causal structure – say an account of the operation of a telephone exchange that describes it in terms of a formal model of its switching operations? In the absence of the semantic notion of representation, the label ‘syntactic’ is quite inappropriate.

Stich evidently regards the significance of the ‘syntactic’ character of his theory of mind as more than that of a mere formal description. This shows up in the fact that he entertains (though he does not insist on) versions of the STM that follow folk psychology in postulating ‘two distinct classes of states underlying behaviour, belief-like states and desire-like states, with state types of *both* classes mapped to a single class of formal objects’ (1983, p. 151). Clearly, if the syntactical story was simply a formal model there would be no reason not to treat the belief/desire distinction as itself syntactical (in the manner, presumably, of the belief/disbelief distinction). Belief-like states and desire-like states would then get mapped to different but related classes of formal objects (and we would again have no reason not to regard desire-like states as a special kind of belief-like state). From a formal point of view, this is the obvious approach; and only the vestiges of the semantic point of view of folk psychology could recommend the alternative.

I emphasize that this is a criticism only of Stich’s ‘syntactic’ characterization of his theory, and not of the theory itself. On the contrary, I think Stich’s articulation of a semantics-free approach to the mind is of considerable dialectical value to the proposal I want to develop in part II (where the task, in a sense, will be to explain how the semantics gets in). The effort would be misguided, to say the least, if it could not begin from a foundation that does not already include what we want to put in. Stich has made it very much less implausible that there could be such a foundation.

Back to the RTM: it might be objected that in suggesting that this view leads us back to the claim that desires are distinguished from beliefs in virtue of their lack of truth conditions, I have ignored the RTM’s account of content. Given this account, it might be suggested, the RTM can characterize the sentences of the language of thought without recourse to the semantic distinctions of natural language (and hence, for our purposes, without leading us back to our starting point). In other words (the suggestion would go) the RTM can explain the fact that it does not represent the desire that P in terms of DES(P): namely by showing that its account of content does not leave room for the necessary content modifier, *Des*.

This objection might be more forceful if there were an agreed approach to the problem of the content of propositional attitudes. As it stands, the closest thing to an agreed position is that content is individuated, in part at least, by the causal or functional roles of the attitudes concerned. This suggestion trades on the appealing hypothesis that the inferential role of propositional contents – the sort of thing that is traditionally supposed to individuate *propositions* – is modelled in the causal role of the relevant mental states. This association is at the heart of the computer metaphor: ‘computers ... just *are* environments in which the causal role of a symbol token is made to parallel the inferential role of the proposition it expresses’ (Fodor 1985, pp. 93–4).

Appealing as this suggestion is, however, it does nothing to justify the belief/non-belief distinction. On the contrary, it again takes for granted a distinction between those aspects of the functional role of a state token that determine its content, and those that determine that it is a belief, a desire, or whatever – a distinction that again looks particularly odd, if we bear in mind the ratio of content differences to attitude type differences in the finished theory (and take off our folk-psychological blinkers). It is as though the Taxation Department kept separate databases for its male and female clients: we would want to know why gender could not be represented

explicitly, like everything else that the Tax Department knows about us. Why buy a filing cabinet, if you intend to keep everything in two or three colossal files? Or to put it another way, why start a separate file for desires, but not start one for disbeliefs, say, or beliefs about nutrition?

One last suggestion: perhaps we have concentrated on the wrong aspect of content. Many cognitive theorists think that there are two aspects to content: the internal aspect, that (more or less) matches internal functional role; and an external aspect, that depends on the actual reference of the attitude concerned – on what in the world it is in fact an attitude *about*. (See McGinn 1982, for example. Fodor himself now favours the view that content is entirely a matter of the external aspect; see Fodor 1987.) With the external aspect taken into account, the range of contents is circumscribed by the range of facts. But clearly this takes us back to our starting point. We wanted the belief/non-belief distinction in the first place because the direct metaphysical attacks looked so unpromising.

In summary, then, we find that if content is individuated in terms of functional role, there is no apparent reason to distinguish more than one type of propositional attitude; and that if it is individuated in terms of external semantic relations, the limits of belief depend on the limits of fact. Neither possibility provides what the non-factualist now needs in order to answer the sceptic or the radical factualist: a psychological criterion with which to distinguish beliefs from other attitudes, and thereby to distinguish fact-stating from non-fact-stating discourse.

### \*5.3 CONDITIONAL FACTS

There is one other important case in which what can be taken to be functionalist arguments have been widely thought to support a non-factualist conclusion. The case is that of indicative conditionals, and the arguments rest on Ernest Adams's (1965) hypothesis, now widely accepted, that the 'assertibility' of an indicative conditional goes by the conditional probability of the consequent, given the antecedent (rather than by the probability of the corresponding material conditional, say).

To see what is meant by Adams's claim, let us look at the ordinary case, with which Adams contrasts the behaviour of indicative conditionals. In what circumstances, in general, will a speaker be willing to assent to a statement *S*? As we have seen, the standard answer is something like this: When he or she believes that *S*, or at least has a partial belief (or degree of confidence) that *S* that in the context in question is sufficiently close to a full belief that *S*. This answer is of course very rough, but not too rough. On the one hand, it is clearly a subjective condition, not immediately concerned with objective justification for assent to *S*. On the other, it does not imply that speakers have to think *about* their beliefs, in order to assent to *S*. We do as a matter of fact tend to assent to those propositions in which we have high degrees of confidence, the claim goes, even if we are rarely aware that this is what is going on.

We saw in chapter 4 that a determined factualist about probabilities will equate degrees of partial belief, or degrees of confidence, with beliefs about probabilities. The normal rule of assertibility then becomes the principle that assertibility goes by increasing probability. 'Assertibility goes by increasing *subjective* probability' is the way it is often put, but this can be a source of confusion. If initially applied to a degree of confidence, the term 'subjective probability' encourages the (factualist) view that for any partial belief there is an associated full belief about a probability. What probability is this? Why, the *subjective* probability! The term 'subjective probability' thus comes to refer ambiguously to the original degree of confidence, now thought of as a belief about a probability, and to the probability that figures in the content of this belief. In fact, of course, if degrees of confidence are beliefs about probabilities, then these probabilities need be

no more subjective than the subject matter of any other belief. While the beliefs themselves are subjective only in the same sense: namely, that they are our beliefs, rather than somebody else's. For this reason, I shall avoid 'subjective probability', instead referring to the mental states concerned as *credences*. The term 'credence' leaves it open as to whether, with the radical factualist, we should think of credences as beliefs about probability.

Adams's hypothesis about indicative conditionals amounts to the claim that what in general determines whether we assent to such a conditional is not our credence in, say, the corresponding material conditional, but rather what we may call our *conditional credence* in the consequent, given the antecedent. For a factualist about probability a conditional credence is a belief about a conditional probability; and Adams's hypothesis comes to the claim that the assertibility of an indicative conditional goes by perceived conditional probability, rather than by the perceived probability of a truth-functional conditional.

It is harder to say what a non-factualist about probability should make of the notion of conditional credence, and hence of Adams's hypothesis. For one thing, if absolute (that is, non-conditional) credences are not beliefs about probabilities then, given Adams's hypothesis, the orthodox definition of conditional probability –

$$(5.1) P(C/A) = P(C\&A)/P(A), \text{ so long as } P(A) \text{ is non-zero –}$$

seems to provide no suitable account of conditional credences. The main problem is that we often assent to an indicative conditional – and therefore, on this story, have a conditional credence – when we clearly have no corresponding absolute credence in the antecedent, the consequent, or their conjunction. I am confident that if it is raining in Moscow then the domes of the Kremlin are wet; but I simply have no idea as to whether it is raining there. (For this and related arguments, and a discussion of what the non-factualist should make of them, see Price 1986.)

However, our present concern is to show that despite certain arguments to the contrary, a factualist about probability can be a factualist about conditionals. Given the factualist's view of absolute credences, it is easy to account for the above cases, and yet to reconcile Adams's hypothesis with the standard definition of conditional probability. For we can believe that the *ratio* of two probabilities is high, without holding an opinion as to the value of either. The factualist about probability thus regards a conditional credence of degree  $n$  in  $C$  given  $A$ , as a belief that  $P(C\&A)/P(A) = n$ ; and interprets Adams's hypothesis as the claim that the assertibility of the indicative conditional 'If  $A$  then  $C$ ' goes by the value of this ratio. Roughly, we assent to such a conditional when we believe that this ratio is sufficiently close to 1.

We can think of such principles of assertibility as fragments of a functionalist account of the mind. They belong to that part of such a theory that relates mental states to linguistic behaviour. As such, moreover, these principles suggest a ground for a non-factualist treatment of conditionals. For they suggest that the general functional account of assent behaviour, summed up in the principle that assertibility goes by degree of (absolute) credence, does not apply to conditionals. Conditionals seem to need a functional account of their own – in which case we have a theoretical basis for distinguishing them from utterances to which the general account does apply.

The argument rests on the assumption that the principle of assertibility for conditionals is not a special case of the general principle. In other words, it requires that conditional credence is not simply a species of absolute credence. If there were a conditional, ' $\rightarrow$ ' such that

$$(5.2) P(A \rightarrow C) = P(C/A)$$

the factalist could endorse Adams's hypothesis, and yet, simply by claiming that 'If A then C' is to be interpreted as ' $A \rightarrow C$ ', be able to say that for indicative conditionals, as for statements of fact in general, assertibility goes by absolute credence.

This suggestion was an early response to Adams's (1965) formulation of his hypothesis (for example by Stalnaker 1970 – 5.2 is sometimes referred to as 'Stalnaker's hypothesis'). But there are two arguments that have been claimed to show that there can be no such conditional ' $\rightarrow$ '. The first is due to Lewis (1976) and the second to Carlstrom and Hill (1978). I want to show, however, that in both cases the supposed problem for this reading of conditional credence is already a fairly obvious but non-fatal feature of a factalist view of absolute credence – and a feature that a factalist about conditionals can accommodate in much the same way as a factalist about probabilities.

Lewis's argument offers a *reductio* of the assumption that there is a conditional ' $\rightarrow$ ' whose probability is given by 5.2. The key is a standard principle of decomposition by cases – a straightforward consequence of the standard axioms of probability:

$$(5.3) P(X) = P(Y)P(X/Y) + P(\sim Y)P(X/\sim Y).$$

Substituting  $A \rightarrow C$  for X and C for Y, 5.3 gives us

$$(5.4) P(A \rightarrow C) = P(C)P(A \rightarrow C/C) + P(\sim C)P(A \rightarrow C/\sim C).$$

Assuming for the moment the equality

$$(5.5) P(X \rightarrow Y/Z) = P(Y/X \& Z),$$

from 5.4 we derive

$$(5.6) P(A \rightarrow C) = P(C)P(C/A \& C) + P(\sim C)P(C/A \& \sim C)$$

and hence, since  $P(C/A \& C) = 1$  and  $P(C/A \& \sim C) = 0$ ,

$$(5.7) P(A \rightarrow C) = P(C).$$

From 5.2 it now follows that

$$(5.8) P(C/A) = P(C).$$

This conclusion is perhaps absurdity enough. It amounts to the result that any pair of propositions X and Y are probabilistically independent, provided only that  $P(X \& Y)$  and  $P(X \& \sim Y)$  are non-zero. But Lewis derives two further results: that there cannot be three incompatible propositions X, Y and Z that all have non-zero probability; and hence that the probability function has at most four values.

Lewis's argument uses 5.5, which he derives quite straightforwardly from the standard definition of conditional probability (5.1) and the assumption, in effect, that rational agents change their credences by *conditionalizing*: i.e., that on coming to believe that X, a rational agent assigns to any proposition Y a probability equal to the conditional probability she previously ascribed to Y

given X. But the details need not concern us, I think, because 5.5 is more general than the argument requires, and we can get what we need for 5.6 a little more directly.

Given the conditionalization assumption ('COND', for short), it follows that  $P(A \rightarrow C/C)$  equals  $P_C(A \rightarrow C)$  – that is, the probability that (as rational agents) we would ascribe to  $A \rightarrow C$ , if we came to believe that C. But by 5.2,

$$(5.9) P_C(A \rightarrow C) = P_C(C/A)$$

and since  $P_C(C) = 1$ ,

$$(5.10) P_C(C/A) = 1.$$

This shows that  $P(A \rightarrow C/C) = 1$ , and similarly we can use the fact that  $P_{\sim C}(C) = 0$  to show that  $P(A \rightarrow C/\sim C) = 0$ , giving 5.7.

In order to show how the factualist might respond to Lewis's argument, I want to exploit certain similarities it bears to a classic argument for fatalism. This argument is generally agreed to be fallacious. But the key to the fallacy is the key to the factualist's response to Lewis (which is not to say that Lewis's argument is fallacious – more on this in a moment).

The traditional function of the argument in question is to encourage an audience to risk its life on the proponent's behalf. In the following fine example (for which I am indebted to Hugh Mellor), the proponent is Henry V, admonishing Westmoreland for wishing for reinforcements for their impending battle with the French:

... No, my fair cousin:  
 If we are mark'd to die, we are enow  
 To do our country loss; and if to live,  
 The fewer men the greater share of honour. (Henry V, Act IV, *Sc.* III)

Philosophers are not insensitive to the pragmatic virtues of this style of argument. Leibniz (1951, p. 153) remarks that 'the sophism which ends in a decision to trouble oneself over nothing will haply be useful sometimes to induce certain people to face danger fearlessly.' He comments that 'it has been applied in particular to 'Turkish soldiers', but has doubts about its efficacy: 'It seems that hashish is a more important factor than this sophism, not to mention the fact that this resolute spirit in the Turks has greatly belied itself in our days.' Pragmatics aside, however, Leibniz, like most philosophers, is convinced that the argument is fallacious.

Its similarity to Lewis's argument stems from two factors: the first that Lewis's principle of decomposition (5.3) can be thought of as a generalization of the logical principle of argument by Dilemma, used by the fatalist; and the second that both arguments involve hypothetical judgements about conditionals. In the case of the fatalist argument, the latter characteristic may not be immediately apparent. It is essential, however, for the fatalist must convince his victim not only that bravery is better in either eventuality, but also that the victim's choice will not otherwise determine the course of events. In the above case, for example, Henry must convince Westmoreland that whether (A) they fight without reinforcements is irrelevant to whether (C) they are victorious. For otherwise Westmoreland will rightly point out that their choice is not between A and not-A (A being better whether C or not-C). Rather it is between (A and not-C) and (not-A and C) – the latter presumably being preferable.

The fatalist thus argues as follows:

- 1  $C \vee \sim C$ .
- 2  $C \rightarrow (A \rightarrow C) \& (\sim A \rightarrow C)$  – i.e., if C, then if A then C and if not-A then C, so that A is irrelevant to whether C.
- 3  $\sim C \rightarrow (A \rightarrow \sim C) \& (\sim A \rightarrow \sim C)$  – i.e., if not-C, then if A then not-C and if not-A then not-C, so A is again irrelevant to whether C.
- 4 Hence, by Dilemma, A is irrelevant to whether C.

It is premisses 2 and 3 that involve hypothetical assessments of conditionals.

As well as these two similarities, of course, there are some very significant differences between this argument and Lewis's. The two arguments are quite different in intent. Lewis's is a *reductio*. Indeed, the absurdity that Lewis derives from 5.2 parallels the fatalist's final conclusion. In arguing for 5.5, moreover, Lewis justifies premisses whose analogues (premisses 2 and 3) the fatalist takes for granted. Nevertheless, I think that the arguments have enough in common to allow us to use our grounds for rejecting the fatalist's as a guide to a possible answer to Lewis's.

Let us focus on premisses 2 and 3, which compare with Lewis's use of the equalities  $P(A \rightarrow C/C) = 1$  and  $P(A \rightarrow C/\sim C) = 0$ . At this point both arguments rely on the procedure of assessing a conditional under an assumption concerning the truth of its consequent. I think we can get a good idea of what a factualist might object to about this procedure by constructing a parallel involving, in place of conditionals, a claim about probabilities themselves.

Let PROP be the proposition that  $P(C) = 1$ . The following argument claims to show that  $P(\text{PROP}) = P(C)$ . By 5.3,

$$(5.11) P(\text{PROP}) = P(C)P(\text{PROP}/C) + P(\sim C)P(\text{PROP}/\sim C).$$

Now by COND,  $P(\text{PROP}/C)$  is the probability we would ascribe to PROP, if we came to believe that C. In these circumstances, we would ascribe probability 1 to C itself (this follows from COND and the fact that  $P(C/C) = 1$ ). In other words (now assuming a factualist's view of credence), we would come to believe that  $P(C) = 1$  – that is, that PROP. This doesn't yet tell us what *probability* we would ascribe to PROP; but applying COND once more, this time using the fact that  $P(\text{PROP}/\text{PROP}) = 1$ , gives us what we need: in these circumstances,  $P(\text{PROP}) = 1$ . Similarly, we can show that  $P(\text{PROP}/\sim C) = 0$ . By 5.11, this gives us  $P(\text{PROP}) = P(C)$ .

Somewhat less subtly, this time following the fatalist rather than Lewis, we can 'show' that PROP is materially equivalent to C itself:

- 1  $C \vee \sim C$ .
- 2 Given that C, then there is no possibility that not-C, and so  $P(C) = 1$  – i.e., PROP is true, as is C itself.
- 3 Given that not-C, on the other hand, it is certainly not certain that C, so PROP is false – again like C itself.
- 4 Hence, by Dilemma, PROP always has the same truth value as C.

It is clear that on any ordinary view of probability, both conclusions are absurd. The first rules out cases (easily found in physics, for example) in which the probability of an outcome necessarily lies in some range that excludes 1; while the second rules out all probabilities other than 1 and 0. But it is easy to see on what grounds a factualist will find these arguments

fallacious. They assume, in effect, that PROP has a fixed meaning, such as to allow us to assess its truth value or probability in hypothetical contexts. Whereas in fact the sense of a probabilistic judgement depends very much on its context. The fallacy becomes apparent (the factualist will say) if we make explicit the normally implicit indexicality of such a judgement. If we think of PROP as

(5.12) Given the evidence to hand,  $P(C) = 1$ ,

for example, then it becomes clear that in assessing PROP under the assumptions that C and that not C – in adding each assumption to the available evidence, in effect – we assess a different proposition in each case. We are interested in whether, given the evidence *actually* to hand,  $P(C) = 1$ ; it is no help to be told that if *further* evidence were to hand (namely the fact that C), it would then be the case that  $P(C) = 1$ .

In the long run, we should clearly expect from the factualist a more precise account of the indexicality of probabilistic judgements. We shall return to this topic in chapter 7. For the moment the important thing is that from the factualist's point of view an appeal to implicit indexicality seems to work as well for conditionals as for probability. If we think of  $A \rightarrow C$  on the lines of

(5.13) Given the evidence to hand, C follows from A,

then both Lewis's argument and the argument for fatalism commit the same indexical fallacy as the probabilistic arguments. They show that on the assumption of certain *further evidence* – the hypotheses that C and that not-C – A is irrelevant to C; but in general these facts are themselves irrelevant to the question at issue, which is whether *on the actual evidence* C follows from (or is probabilistically dependent on) A.

Against this indexical view of indicatives, it might be objected that it violates the presumption, as Lewis puts it, that 'our indicative conditional has a fixed interpretation, the same for speakers with different beliefs, and for one speaker before and after a change in his beliefs.' How else, Lewis asks, 'are disagreements about a conditional possible, or changes of mind?' (1976, p. 301). The thought is thus that if conditionals involve an implicit indexical reference to evidence, then as evidence varies between speakers, and for a given speaker over time, so will the content of conditionals. This seems contrary to the intuition that different speakers, or the same speaker at different times, can disagree about a conditional, without, as it were, arguing at cross purposes. For the moment, however, it is sufficient answer to this objection that speakers apparently cope with the same difficulties in the case of probabilistic judgements (and certainly do so with many other sorts of indexical expression). The important thing is that the two cases seem to be in the same boat.

Let us turn then to the second argument claimed to show that Adams's hypothesis cannot be treated as a special case of the general principle of assertibility for statements of fact: that of Carlstrom and Hill. Here we cannot do better than to reproduce Carlstrom and Hill's own presentation of the argument (with minor notational changes):

We will derive a contradiction from the assumption that truth conditions can be assigned to the indicative conditional  $A \rightarrow C$ , for arbitrarily chosen A and C, in some possibly complex way such that the probability that  $A \rightarrow C$  is true is identical with the conditional probability  $P(A \& C)/P(A)$ . Let A and C be contingent and logically independent sentences, and let A, A&C, and  $A \rightarrow C$  be true at some possible world W1. The truth value of  $A \rightarrow C$  at a world is

not fixed by the truth values of A and C, for none of the sixteen equivalence classes of Boolean sentences constructible from A and C has a member with the property that its probability of truth is always identical with  $P(A\&C)/P(A)$ . Thus there are worlds  $W_2$  and  $W_3$  with the following properties:  $A \rightarrow C$  is false at  $W_2$  but true at  $W_3$ , and each of A and  $A\&C$  has the same truth value at both worlds. We can imagine two possible ways in which probabilities are assigned to the class of worlds. Under the first assignment, one of the two worlds  $W_1$  and  $W_2$  is probably the case, but each is about as likely to be the case as the other. In such a situation a sentence true at both worlds has a probability of nearly one of being true, a sentence true at just one of the worlds has a probability of about one half, and a sentence true at neither world has a probability of about zero. The second assignment is just like the first except that the worlds  $W_1$  and  $W_3$  are considered in place of  $W_1$  and  $W_2$ . Now since  $A \rightarrow C$  is true at  $W_1$  and false at  $W_2$ , its probability of truth under the first assignment is about one half. Under the second assignment its probability is nearly one. But the probabilities of both A and  $A\&C$  are about the same for both assignments, and in both cases, the probability of A is considerably larger than zero. Hence, the conditional probabilities are about the same. Although the probabilities that the conditional is true are quite different, the conditional probabilities are very close. Conditional probability is not probability of truth. (1978, pp. 156–7)

I want to show that as with Lewis's argument, the problem that this argument identifies is already a feature of a factualist interpretation of non-conditional credences. In this case the relevant point is so obvious that it may seem trivial. For example, let B be the proposition that the probability of a given proposition A is 0.9. Since  $P(A) = 0.9$  does not imply that A, it is possible that B is true and A false. Let W be a world in which this is the case. Now imagine an assignment of probabilities to worlds that gives W a probability very close to one. In terms of this assignment it is true that  $P(A)$  is close to zero, and yet also very probable that  $P(A) = 0.9$ . In other words, this assignment justifies the assertions both that B is false, and that B is very probably true.

The fallacy is fairly obvious: if we want to assign probabilities to possible worlds, and hence to think of the probability of a proposition as the sum of the probabilities of the worlds in which that proposition is true, then we must take it that judgements about probabilities are not true or false at individual worlds, but rather in virtue of facts about the space of worlds as a whole. Probabilistic judgements must thus be viewed as modal judgements.

Again, our concern is not with the adequacy of this form of factualism about probabilities, but rather with the point that the same kind of answer defeats Carlstrom and Hill's analogous objection to a factualist treatment of indicative conditionals based on Adams's hypothesis. I think the most striking way to illustrate this is simply to *define* ' $A \rightarrow C$ ' as ' $P(C/A) = 1$ '. It is true that in one sense this violates Adams's hypothesis – more on this in a moment – but it preserves something similar. We have already noted that it is a consequence of the factualist view of credence that for any proposition X, to believe (or be disposed to assent to) X is to believe (or be disposed to assent to)  $P(X) = 1$ . Let us express this consequence as

$$(5.14) \text{Bel}(X) \leftrightarrow \text{Bel}(P(X) = 1).$$

Given Adams's hypothesis, this means that if we accept 5.2, we are in any case committed to

$$(5.15) \text{Bel}(A \rightarrow C) \leftrightarrow \text{Bel}(P(C/A) = 1).$$

The present suggestion is in effect that we should regard 5.15 as a consequence not of Adams's hypothesis and 5.14, but simply of the fact that 'A→C' and 'P(C/A) = 1' are synonymous. The consequence of 5.14 will rather be that

$$(5.16) \text{Bel}(A \rightarrow C) \leftrightarrow \text{Bel}(P(P(C/A) = 1) = 1).$$

To the extent that it tells us why someone who believes that the probability of C given A is (effectively) 1 also believes that if A then C, this suggestion agrees with – indeed, *explains* – Adams's hypothesis. But it will see it as a mistake to regard Adams's hypothesis as an assertibility condition, in the above sense: the same kind of mistake as we would make in claiming that 'X is a bachelor' is governed by the rule that one should be prepared to assent to it just in case one believes that X is an unmarried male.

In the present case, the difference shows up in cases in which P(C/A) is assigned a value other than 1. Suppose, for example, that we believe that P(C/A) = 0.75. On Adams's hypothesis, this implies that we take 'If A then C' to be assertible to degree 0.75; whereas on the present suggestion, since P(C/A) = 0.75 implies that P(C/A) is not one, 'A→C' will simply be believed to be false. As it happens, this suggestion thus makes better sense than Adams's hypothesis of at least some uses of indicative conditionals. Consider this exchange, for example:

S: If his heart has stopped, then he is dead.

R: That's not true – some people have survived cardiac arrests, so even if his heart has stopped, he may still be alive.

In such cases conditionals seem to be rejected outright, even if the corresponding conditional probability is quite high.

However, our present concern is not to defend this view of indicative conditionals, but simply to make the point that for a factualist about absolute credences, the Carlstrom and Hill argument, like Lewis's, raises no new problem for a factualist view of conditionals. Carlstrom and Hill's *reductio* depends on imposing on conditional truths a constraint the factualist is already bound to reject for probabilistic truths. This is really all we need here, though the above suggestion goes further. It proposes a factualist interpretation of indicative conditionals, automatically available to factualists about non-conditional credences, that makes sense of the grounds, if not the letter, of Adams's hypothesis: the intuition that assent to an indicative conditional requires a high conditional credence.

I conclude that the arguments of Lewis and of Carlstrom and Hill do not prevent a factualist from endorsing Adams's hypothesis (or something like it), and yet treating the resulting assertibility rule for indicative conditionals as a special case of the rule for statements in general. Earlier we construed these assertibility rules as fragments of a functionalist account of mind and linguistic behaviour. It seemed that indicative conditionals might be a case in which functionalism could provide the belief/non-belief distinction required by a psychologically grounded non-factualism. But it now appears that for conditionals, as in general, the functionalist is unable to justify a division between those functional differences that reflect differences of content, and those that reflect a difference in the type of attitude concerned.

It should be mentioned here that some factualists about conditionals – notably Lewis himself and Frank Jackson – have responded to Adams's hypothesis in a very different way. Broadly, they accept that indicative conditionals employ a special rule of assertibility, but consider this to be an addition to the appropriate case of the rule for truth-conditional utterances in general. On this

account indicative conditionals have truth conditions – those of the material conditional, in fact – and their rule of assertibility requires, *inter alia*, a belief in their (probable) truth. But for reasons to do with the conversational role of such utterances, it also requires that the corresponding conditional credence be sufficiently high. (Lewis 1976, Jackson 1987.)

The major difficulty for such an approach is to explain the significance of the claim that conditionals have truth conditions. It cannot be to account for ordinary practices of acceptance and rejection of such conditionals; for to recognize a need for Adams's hypothesis is to recognize that the truth conditions of the material conditional do not explain ordinary usage. Moreover, we concluded in chapter 3 that if we are prepared to explain away usage, any meaningful utterance can be construed as truth-conditional. In a parallel fashion, any utterance can be construed as expressing a belief. But the construction does not guarantee that such a belief will play any non-eliminable role in a functional account of the use of such utterances. In a similar way, it seems to be the functional story associated with Adams's hypothesis – the special rule of assertibility – that does the work in a Lewis or Jackson account of the use of indicative conditionals. Lewis and Jackson accept that in virtue of Lewis's *reductio* argument (and perhaps also Carlstrom and Hill's) this part of the story cannot be interpreted in terms of the general account of statement of fact. The above argument suggests that this is a mistake. Provided that we are factualists about probability, and in particular interpret non-conditional credences as genuine beliefs, we can also be factualists about indicative conditionals.

What, incidentally, if we are not factualists about non-conditional credences? Are we then prevented by Adams's hypothesis and the Lewis and Carlstrom and Hill arguments from being factualists about conditional credences? It seems to me not. I mentioned earlier that in this case there are strong grounds for rejecting the usual definition of conditional probability in terms of non-conditional probability. A conditional credence cannot be thought of as the possession of a pair of absolute credences of appropriate content and degree, for in most cases in which we need a conditional credence (according to Adams's hypothesis), we simply do not have the relevant absolute credences.

In the (1986) paper in which I offered this argument, I suggested that we should therefore think of conditional credences as what I called 'inferential dispositions': dispositions to adopt a consequent credence of a certain degree, on acquiring an antecedent belief. This is not the only possibility. For example, Ramsey (1978, chapter 3) thought of both conditional and absolute credences as theoretical psychological entities, to be characterized in the first instance in terms of their effects on behaviour. Both can be measured in terms of choice of betting odds. The difference is that conditional credences call for conditional bets.

The various possibilities agree on one thing, however. The equality 5.1, normally used to define conditional probability, is no longer sacrosanct. At best, it is likely to be what it is for Ramsey: a derived result about the relationship between the absolute and conditional credences of a rational agent. The present significance of this new attitude to 5.1 lies in the fact that it suggests another answer to the arguments of Lewis and Carlstrom and Hill. In so far as these arguments rely on the usual definition of conditional probability, it is open to someone who rejects this definition – instead regarding conditional credences as an independent class of psychological entities – to object that Lewis and Carlstrom and Hill misuse what is at best a contingent connection between conditional and absolute credences.

The objection is actually more appropriate for the Carlstrom and Hill argument than for Lewis's. We saw that Lewis's argument can be expressed without reference to 5.1. In the Carlstrom and Hill case, however, 5.1 is crucial. In this case the claimed absurdity was just that it should be

possible for different systems of belief to assign the same probability to  $A$  and  $A\&C$ , and yet different values to  $P(A\rightarrow C)$  and hence to  $P(C/A)$ . But if 5.1 is at best dependent on an assumption of rationality, then there is nothing contradictory in the possible existence of such systems of belief. (Should irrationality be *impossible*?) There might be a contradiction if the Carlstrom and Hill argument showed us that rational systems of belief could differ in this way, but the argument offers nothing of this kind.

One might think – I speak here from experience, as my (1986) paper shows – that in taking a non-factualist view of absolute credences, and hence being led to view conditional credence as a primitive psychological category, one thereby adopts a non-factualist view of conditionals themselves. But on the contrary, it now appears that a primitive view of conditional credence in fact makes things easier for a factualist account of conditionals. Such a view can already help itself to our earlier responses to the Lewis and Carlstrom and Hill arguments; and now, in virtue of its new attitude to 5.1, has another very promising objection to the latter argument.

If we characterize conditional credences as inferential dispositions, or something similar, then at first sight it seems that we assign them a different role from belief in a functionalist theory of the mind. Conditional credences become dispositions to add to our beliefs – to move from one belief to another. But at second sight we should appreciate that there is nothing in the functionalist theory to show that this functional role is not itself performed by a class of beliefs. It is the familiar problem: as psychologically inclined non-factualists, we must do more than show that a class of utterances are the expressions of mental states with a special functional role; we must exclude the possibility that the functional distinction in question marks a difference in attitude content, rather than a difference in attitude type.

**PART II**

**Truth and Its Limits: An Explanatory View**

## 6

**The Function of Truth**

For the benefit of readers who may have just joined us, let me summarize events so far. We have looked at a range of attempts to analyse the notion of statementhood – to say what it is about an utterance that qualifies it to be ranked in the fact-stating part of discourse. However, none of these attempts seemed able to exclude the possibility that all discourse is fact-stating discourse. None therefore meets the needs of either of the programmes we identified in chapter 1 – of non-factualists or marker theorists, who have a common interest in a non-trivial fact-stating/non-fact-stating distinction, even if conflicting views as to where it should be drawn.

One option at this point would be to walk away from the problem, concluding that non-factualists and marker theorists are simply misguided in expecting language to divide in this way. I think this would be premature, however. Before we blame the problem for the failure of the analytic solution, we should consider the possibility of other solutions. In the remainder of the book I want to argue for another approach. I want to show that so long as we concentrate on *explanation* rather than *analysis*, we can find considerable illumination in the concerns of non-factualism.

As I indicated in chapter 1, this explanatory approach will be founded on a novel theory of truth. The attractions of this theory are not limited to its ability to make sense of non-factualism, in my view, significant as I think this is. So, as a guide particularly to the reader whose interest is in the more general issues, let me indicate the structure of the remaining four chapters.

In this chapter I first characterize the explanatory approach, abstractly and by example; and then try to head off some objections to its application to a linguistic distinction of the kind we are interested in. I shall talk at this stage of the fact-stating/non-fact-stating distinction, but parallel remarks would apply to a distinction between the truth-bearing and non-truth-bearing parts of language. In any case, I then concentrate on truth, proposing an account of its function in discourse. The chapter concludes with some remarks on the relation between this account and other recent ‘explanatory’ approaches to truth.

Chapter 7 begins with a defence of the proposed account against a number of objections. From section 7.6 I turn to the account’s application to non-factualism. Chapter 8 continues this topic, mentioning also the non-indicative moods. Chapter 9, finally, returns to general issues. I discuss the connection between the proposed account of truth and Fregean general semantics, and relate it to other recent positions on truth, realism and objectivity.

## 6.1 EXPLANATION WITHOUT ANALYSIS

As theorists we often find ourselves with the task of explaining distinctions marked in the ordinary lives of ourselves and others. We notice that in cases of a certain kind, our subjects treat some instances in one way, others in another. Sometimes the behavioural difference will amount to application or non-application of a descriptive term. Sometimes it might consist in some other aspect of our linguistic behaviour. Or sometimes it might be entirely non-linguistic. Whichever it is, however, these behavioural distinctions call for explanation. Such patterns do not seem arbitrary – we want to know how they arise.

One sort of explanation tries to identify some feature of the two sorts of case, to which our subjects' behaviour can be seen as responsive. This is the analytic or reductive approach. On this view, to put it crudely, their choice between behaviour X and behaviour Y is governed by their ability to recognize the presence of certain properties PX and PY, and their conformity to the rule that they do X when they detect PX and Y when they detect PY. In the particular case in which the behavioural difference amounts to the application or non-application of a descriptive term, this amounts to the view that they use the term to refer to one of the properties in question; its use again being dependent on their ability to detect such a property in their environment.

An analytic account should thus be explanatory; the appeal to underlying properties should be capable of accounting for the behavioural difference with which we began. An adequate explanatory account need not be analytic, however. There are ways of accounting for behavioural distinctions that do not depend on an appeal to recognition (even implicit recognition) of underlying properties. This is perhaps most obvious in the case of non-linguistic behaviour – that of other species, as well as some of our own. Why do we tend to fall asleep every evening, for example? Recognition of the time of day obviously has less to do with it than has the physiological basis of our circadian rhythm.

In linguistic cases it is perhaps less obvious that the appropriate explanation may be non-analytic. All the same, some such accounts are quite familiar. They arise most naturally from a detached perspective: that of the anthropologist, say, or the historian of science. For example, we would not expect a fully recognitional account of the origins of a system of religious or defunct scientific distinctions; for we want to say that there may be nothing in the world that matches the distinctions concerned.

Some terminology: from now on I shall use 'explanatory account' as an abbreviation for 'non-analytic explanatory account'. This is not only shorter, but gets the emphasis right. Explanation comes second in an analytic account. Explanatory approaches put explanation first.

I want to suggest an explanatory approach to two linguistic distinctions that are very far from defunct; two distinctions marked in our ordinary use of the notions of truth and falsity. One of these is the distinction whose analytic treatment part I was concerned to discredit – that between statements and non-statements, or truth-bearing and non-truth-bearing uses of language. The other is the distinction we mark within the truth-bearing part of language, in applying the descriptions 'true' and 'false' selectively. I want to explain the former distinction in terms of an account of the latter. In other words, I want to show that we can explain the limits of truth in language in terms of an account of its function in central cases.

In both cases the explanation is intended to be non-analytic. I want to show that our use of truth and its limits can be explained without being taken to reflect or be responsive to any such real properties of linguistic items. Because these distinctions are so much a part of our own practice, however, the detached explanatory perspective does not come easily. To soften the target a little, I want to begin with an example in which there seems to have been an analogous shift from an analytic to an explanatory approach to a perceived distinction in human life. It is the case of social class.

Social stratification is a striking feature of most societies. It is not simply a theoretical notion, important only to anthropologists and sociologists, but part of the lives of ordinary people. Many aspects of a person's life depend on his or her caste, or class, as it is perceived by the rest of the

community in which he or she lives. Ordinary folk apply such notions to themselves, and the theorist begins with the notions of folk sociology.

Folk sociology takes folk usage at face value. After all, what is folk sociology except folk usage about society? Folk usage applies class properties to individuals. Hence it gives the impression that the way to a theory of class is to *analyse* these properties – to discover, for example, the distinguishing characteristics of the aristocrat and the commoner, the prince and the untouchable.

The simplest such theory perhaps takes the analysis no further than the folk class properties themselves. Class is thus itself regarded as part of the natural order of the world. An individual is seen to belong to a class in much the way that an animal or plant belongs to a species. We now find such a view absurd and repugnant. We should not underestimate its plausibility, however, particularly in societies without significant class mobility, and in combination with a suitable account of the origins of such an order.

A more sophisticated approach, less clearly repudiated in contemporary thought, is to analyse class distinctions in terms of more basic properties of individuals. So long as these properties are taken to be changeable, this approach allows for class mobility. Unlike the first, it can also distinguish between real and apparent class, and hence allow for the class exiles of fiction and fact – lumberjack lords, Downing Street grocers, and the like. Moreover, it promises an explanation of some of the social manifestations of class. Indeed, it may seem to offer a justification of social stratification – the idea that because the rich are better at something, they not only gets but deserves the gravy.

Attractive as they have often seemed, such accounts of class founder on one of three main fallacies. For a start they may appeal to underlying properties that although genuinely individual, are not unevenly distributed between the social classes. Appeals to biological properties of individuals are likely to fail for this reason, for example. Unlike carthorses, the working classes are not innately adapted to the labouring life.

The second kind of fallacy is to call on properties that although both individual and appropriately distributed, bear the wrong explanatory relationship to the relevant social distinctions. This can happen in several ways. The properties concerned may be effects rather than causes of the class distinctions they are supposed to account for – think of the correlation between access to education and social class, for example. Alternatively, their correlation might stem from a common cause, or simply be accidental.

The third fallacy is perhaps the most pervasive, however. It is to appeal to properties that are themselves constituted in social terms. It would be involved, for example, in an attempt to explain class distinctions in terms of differential ownership of capital or possession of political power. Property and power are themselves in some sense social constructs. We may well clarify the problem of class by describing it in these terms – come to understand that some social distinctions are of more significance than others. But to clarify a distinction in this way is not to explain how it comes to exist in the first place.

In rejecting such accounts of class, we recognize, I think, that an adequate theory of class would not be simply an analysis of the class properties of folk sociology. We distinguish the task of describing the class structure of a given society from that of explaining the existence of such distinctions. As noted, the descriptive task may have an analytic component. But we would not

expect the explanation of social stratification to appeal to non-accidental differences between individuals.

What the explanation of class should appeal to is a matter for some debate. By way of illustration we might mention two approaches, both in some sense evolutionary. On the first, class would be viewed as a product of notions of kinship – of their tendency over the generations to accentuate, rigidify and defend what were originally minor and accidental variations in social role. Classes would condense like galaxies from clouds of dust; or perhaps more accurately, like organisms from the primordial soup. On the second view, in contrast, a social class would be compared more to an adaptive variation than to an organism or species in its own right. Class distinctions would thus be explained in terms of their functional roles, and hence their contributions to the reproductive success of a community as a whole.

Whatever the appropriate story, however, the class case makes two important points. It shows that the use of a distinction that seems initially to reflect intrinsic differences between individuals may turn out to require a non-analytic style of explanation. And it illustrates that in the grip of the analytic perspective, we are liable to take as explanations what are at best redescriptions of the distinctions we want to account for. It seems to me that the sorts of views we considered in part I make the same kind of mistake about the fact-stating/non-fact-stating distinction. Naïve theories of class attempt to explain class distinctions in terms of intrinsic differences between individuals. Analogously, these linguistic views try to account for the distinctive features of fact-stating discourse in terms of some intrinsic property, supposedly distinctive of fact-stating uses of language. As in the case of class, the approach fails for two main reasons. Either the claimed underlying property is not appropriately distributed (or cannot be shown to be so distributed, independently of the distinction it is supposed to explain). Or the supposed intrinsic property is itself a manifestation of the difference of usage that calls for explanation.

However, I think it must be conceded that although the class analogy might suggest this possibility, it does little to support it. It is clear, I take it, that the proper approach to class is explanatory; but far from obvious that the linguistic distinction might even admit the same kind of treatment. This difference seems to stem from an important disanalogy between the two cases. We incline to the lottery view of class selection: whatever the origins or function of social stratification, an individual's class is decided, in the main, by the accident of birth. The lottery view is much less plausible in the linguistic case, however. The 'individuals' differentiated by the linguistic distinction are speech acts, linguistic tasks, or something of the kind. They are not biological individuals, whose interests we know to be often subordinate, from a biological point of view, to those of a species or community. They compare instead to enduring characteristics of a species. We want to explain some striking differentiation of the tasks for which we employ language – the distinctions marked in English by the use of the indicative mood, or the application of the notions of truth and falsity, for example. As a result, the linguistic distinction seems more tied to an explanatory base than is true of the class case.

It is rather as if we had observed that a community uses two quite different methods of performing the task of moving themselves from place to place. Sometimes they 'go on foot', at other times they 'take the Tube'. The community may have little sense that these activities have something in common. They may not realize that in principle anywhere that can be reached by taking the Tube could also be reached on foot. As theorists, however, we would expect an explanation of this behaviour not simply to account for the fact that the movement task is differentiated in this way, but to explain why particular movement tasks tend to be performed in one way rather than the other. We note for example that from Highgate to High Barnet, almost all our subjects 'take the Tube'. We would be surprised if this were simply an accident – if the

habit were not explicable in terms of certain features of that particular movement task (such as the distance from Highgate to High Barnet, and the availability of the Northern Line).

In the linguistic case, similarly, we expect that as an enduring feature of human behaviour the use of the statement form will turn out to be well adapted to the needs of particular linguistic tasks – to the sociolinguistic function it has presumably developed to serve. I think this explains why the analytic or reductive approach is less clearly inappropriate here than in the case of class. If an account of fact-stating discourse needs to appeal to some pre-existing difference between linguistic tasks, in order to explain the selective application of the features of usage characteristic of statements, then this difference seems itself to form the basis of a reductive account. For can we not say that statements are those utterances that are used to perform tasks having the underlying property that explains the assertoric form of usage?

I think there are three things that may seem to recommend this suggestion. The first is the lingering appeal of the analytic approaches discussed in part I. In the face of the evidence, our intuitions doggedly maintain that we can explain the assertoric form in terms of its use in expressing beliefs, carrying truth values or representing states of affairs. Putting this aside, two factors remain: first that it is difficult to see how we could explain the development of the assertoric form, except as a manifestation of the kind of base property that would enable the suggestion to go through; and secondly that an appealing account of the use of language seems to require that speakers be aware of some such base property, in order to explain their choice of particular forms of words.

In the next section I shall try to cure any lingering attachment to intuitions of this kind. Readers not so afflicted may move directly to section 6.3. But as an indication of the kind of resolution these factors allow, consider again our anthropological subjects who ‘go on foot’ or ‘take the Tube’. I said that although these people might have little conception that these were different ways of doing the same thing, we would expect to find explanatory differences between the ‘movement tasks’ for which they typically choose one mode rather than the other. But we would recognize, I think, that the correlation between these underlying properties and our subject’s choice of transport might be far from perfect. Being creatures of convention and habit, they might sometimes take the Tube when (as we can see) it would be better for them to walk, and *vice versa*. These correlation failures might be either type-type, as when it would generally be better to make the journey in question by the other means; or type-token, as when the usual advantage of the chosen mode is outweighed by special circumstances (frozen points at Camden Town, say). From the explanatory point of view these failures are of little significance. An evolutionary account relies on long run advantages, and tolerates an element of historical accident. But clearly they defeat any simple reduction of the acts of taking the Tube and going on foot to the underlying properties that explain their differentiation in nature. And the same may be true in the linguistic case. Whatever the usual function of fact-stating or truth-bearing discourse, we should not be surprised to find cases (tokens and types) in which this function goes unserved. This possibility prevents any simple analysis of assertion in terms of the underlying properties being held to account for its form and existence in language.

## \*6.2 ADAPTION AND LINGUISTIC CHOICE

We want to explore the idea that the statement form has developed to serve some identifiable sociolinguistic function. We want to resist, however, the resulting pressure to *characterize* statements as speech acts that serve this function. I mentioned two intuitions that might seem to favour this move: first, that it would allow us to explain the development of the assertoric form as a manifestation of the relevant functional property; and secondly that an adequate account of

the use of language might seem to require that speakers be aware of some such property, in order to explain their choice of particular forms of words.

Concerning the differentiation of language into assertoric and non-assertoric forms, there seem to be two possible accounts. One would allow that the development of the assertoric form has been at least partly guided by speakers' own intuitions about the appropriate form of their language, in the light of its functions. The other would insist that such development is essentially 'blind', not significantly influenced by speakers' understanding.

The latter account seems bound to resort to an evolutionary story, explaining the distinctive characteristics of assertoric and non-assertoric discourse as the result of a gradual process of adaptation and selection. The same will be true of the former account, to whatever extent it relies on 'blind' development. In neither case, of course, need 'evolutionary' imply a genetic basis for the variation and retention of linguistic characteristics. Nor is it clear that linguistic selection need depend on reproductive success. In humans as in certain other animals, new patterns of behaviour can spread through a population in very much less than a generation; and hence very much faster than could be explained by their contribution to fertility.

All the same, the likely evolutionary character of the development of language does a lot to undermine the intuitions that seemed to favour the analytic approach. For one thing, an evolutionary theory is capable of explaining the development of a behavioural distinction, without treating it as the manifestation of some pre-existing distinction. This is one of the great achievements of Darwinian theory: the fact that it can account for functional differentiation in nature, without taking the development of species and other biological kinds to consist, as it were, in the teleological colonization of previously vacant roles. On the contrary, the process of development creates the roles as it goes along. Hence the vacuity of the doctrine, sometimes advanced, that evolution 'fills every available niche'. Within certain physical constraints, evolution creates the niches as it creates their occupants.

Take the case closest to hand. Darwinism frees us of the idea that the human hand is a directed solution to a pre-existing functional problem – an idea embodied not only in 'designer' theories of creation, but also in the Lamarckian view that adaptation occurs as organisms strive to reach certain goals. After Darwin we can say in all seriousness that the functional role of the hand – its biological niche – is the performance of manual tasks. This truth is as trivial as it seems. Only a teleological view requires that the functions of the hand have anything more in common than the fact that they are the functions of the hand. For a Darwinian the interesting truths lie elsewhere: in the story concerning the development of the hand from its precursors, and the contributions of its developing functions to survival and success. And so much, I think, for the feeling that an analytic account is required to explain the development of the assertoric form of usage. It rests on a failure to appreciate the potential of non-teleological accounts of the development of biological characteristics.

The second intuition also rests on a teleological viewpoint, but at a different level. Whereas the first intuition concerned the development of forms of usage themselves, the second has to do with their application on individual occasions of use. It stems from the feeling that any significant feature of an utterance should in principle be explicable in terms, in part, of the speaker's *understanding* of the relevant part of the language. We think of the choice of a particular form of words as an intentional action, and therefore as something that ought to be explicable, in part, in terms of the speaker's beliefs about the significance of those words. To accept otherwise would seem to be to consign the relevant feature of linguistic competence to the ranks of the instinctive, the habitual and the 'wooden' (Blackburn 1984, p. 96).

I think we are thus inclined to think of the choice of the assertoric form as something like the choice between synonyms – what goes on when a competent speaker chooses to say ‘Tube’ in preference to ‘Underground’, ‘Metro’, or ‘Subway’, for example. Such a choice is presumably based on a prior understanding of the nature of the speech act in which the chosen word will play a part. It is difficult to shake off the idea that the ‘choice’ of the indicative mood and other features of the assertoric form is similarly based on an implicit grasp of an underlying property of an intended communicative act. Since such an underlying property would seem to provide the basis for an analysis of the notion of statement of fact, this ‘choice’ model of utterance lends credence to the analytic approach.

Appealing as the choice model may seem, it cannot be applied without restriction. Unlike Humpty Dumpty, we find our choice of words constrained by our intended meaning. Of course, these constraints usually allow us a good deal of freedom. Thanks to the existence of synonyms, there are usually many ways of ‘saying the same thing’. But the fact that there are such constraints means that we may in fact have no choice about the assertoric form, and hence no need to be aware of an underlying property in virtue of which that form is appropriate.

It might be objected that even where language gives us no choice, we have in some sense to know our intended meaning, in order to be in a position to determine the correct words for the job. If so, then linguistic intentions are once again prior to the selection of the assertoric form. But there are big problems for such an account. In particular, it seems to depend on the existence of judgements that are not linguistic in form. For what does a speaker conclude, when he judges that the sentence ‘P’ is the appropriate way to express what he means to say? That ‘P’ expresses the proposition that P? This judgement is not tautologous, as of course we discover if we translate it into a language other than English. But it is not the sort of thing that a competent English speaker could be in any doubt about. So to the extent that reasoning is conducted in English, this judgement cannot be a significant prerequisite for the decision to say ‘P’.

What if we move to another language – perhaps the language of thought? Then we do have a non-trivial judgement, of the kind of thing that a would-be English speaker would have to know. Let us suppose that in the language of thought, this judgement is expressed by ‘Q’. In other words, we are supposing

R: In the language of thought, ‘Q’ expresses the proposition that in English, ‘P’ expresses the proposition that P.

The suggestion is that in order to account for our speaker’s utterance of ‘P’, we have to admit that he judged that ‘P’ is an appropriate way of saying that P. Since this judgement is trivial in English, we are supposing that he expresses it in another language – the language of thought, for example. (Interposing other languages at this point would only delay the inevitable.) Presumably he makes this judgement by tokening ‘Q’ in an appropriate ‘judgement register’. In effect, he says ‘Q’ to himself. But why ‘Q’? Is *this* choice guided by the judgement that R? If so, then the problem recurs again. If not, then why is there a problem in the first place?

Thus at some point this view seems committed to the existence of non-linguistic judgements. This takes the view perilously close to the instinctive, habitual and wooden. The mark of ‘non-woodenness’ seems to be our ability to reflect on our own beliefs and practices. But how can we reflect on what we cannot express? In other words, it looks as if the preference for the choice model of the use of the assertoric form (and hence for an analytic account of statement of fact) rests on an unattainable ideal. In the use of language, as elsewhere, our freedom of choice cannot be boundless. Our choices are always constrained by the terms in which the options present

themselves. We cannot choose what we cannot conceptualize. The above argument suggests that in the case of ‘choosing one’s words’, the options present themselves in linguistic form. Choosing one’s words is not a matter of first settling on the particular (fully specified) speech act one wants to perform, and secondly selecting the appropriate words for the job. There is only one kind of choice; though it may go in several steps, as one refines one’s intentions, in the process of settling on a particular form of words.

There is a nice parallel here with accounts of the origin of life. Like teleological misinterpretations of Darwin, creation myths tend to embody an unwarranted conceptual distinction between a biological kind and its ‘place’ or niche in nature. For example, the biblical account may leave us with the impression that when God created a species, He made two decisions: to include a particular niche in the Great Plan, and to create the creature that fills that niche. These decisions might be taken in either order, and we find suggestions of both orderings in the biblical account. The more obviously absurd is that of creation before assignation to a biological role. This is suggested, for example, by God’s famous injunction – to creatures already equipped, presumably, with their reproductive instincts and apparatus – to ‘Go forth and multiply.’ The reverse ordering is suggested, for example, by the account of the creation of Eve. We are given the impression that God first decides to provide Adam with a mate, and then decides to create Woman for that purpose. But there is room for two decisions here only to the extent that the first decision is incomplete. God might have initially decided that Adam should have a mate without making up His mind on some of the details. (Should Eve eat Adam after mating, for example?) This leaves room for a series of decisions. However, to the extent that God has filled in such details on the Great Plan, the actual creation of Woman is not a separate decision but merely the fulfilment of a prior intention. A place in the Great Plan is delineated by the properties of its occupant. To decide to fill the place is thus to decide to create the occupant.

Analogously, the above view of linguistic competence involves what amounts to a two choice account of the ‘creation’ of an utterance: the idea that in deciding to speak, we first decide what kind of linguistic act we wish to produce – what niche we want to fill – and secondly choose the appropriate sentence for the job. It seems to me that this model involves the same sort of misconception as the two-choice view of the creation of life. Just as the common feature of the hand’s activities is simply that they are *manual*, so what unifies the speech acts for which we use the statement form is just that they are utterances in the statement form. We cannot choose the statement form on the basis that we recognize that an intended speech act belongs to a certain kind, because to get that far is already to have made a decision about the use of the statement form.

Given two sentences of whatever form, we can of course choose to utter one, both, or neither. In doing so we rely, presumably, on some conception of the effect of an utterance of either kind. What we must resist, I think, is the idea that this understanding of the role of a potential utterance is somehow independent of our awareness of the linguistic expressions whose use is at issue. In the end, we must live with the fact that linguistic acts present themselves to us as *linguistic* acts, and not in terms of some pre-linguistic characterization.

All the same, in this and the next two chapters it will sometimes be convenient to speak as if we took it that language had been developed to serve a range of pre-defined functions; in particular, to express a range of different sorts of mental states – beliefs and suchlike. As is often the case in biology, a teleological perspective will thus be a useful expository device. This psychological scaffolding will be thoroughly dispensable, however. The work will be done by the notion of behavioural dispositions, and we won’t need to assume that such dispositions serve to delineate

the familiar psychological categories – in advance, as it were, of the linguistic distinctions we shall use them to explain.

### 6.3 TRUTH: THE EXPLANATORY PERSPECTIVE

In chapter 2 I mentioned Strawson's (1949) 'amen' account of truth, which makes a theory out of the observation that one function of the term 'true' is to provide a uniform means of endorsing a statement made by another speaker. Strawson argues that the description 'true' is primarily a device to confirm or corroborate a previous utterance. He doesn't say what 'false' is, but there are at least two possibilities: that it 'confirms' the negation of the original utterance; or, without assuming an account of negation, that it indicates some sort of rejection or 'disconfirmation'.

Strawson's theory is not an analysis of truth. It is rather an account of the function of the term 'true' in discourse – an account that does not presuppose that the function of such a term must be primarily referential. Strawson does not tell us what truth is, but offers a theory as to why we should talk as if there were such a property. In our terms it is thus an explanatory theory of truth.

It is not a very satisfactory theory, however. For we noted that if corroboration were the main function of 'true' it would be unclear why the term should not be applied more widely – to questions and commands, for example. It would often be useful to have a simple device for putting one's weight behind a question asked by someone else: for indicating, in effect, that one is asking the same question oneself. Similarly, there are cases in which it would be useful to be able to say in a simple way that a question did not have one's endorsement. The same is true of commands, requests, indeed of practically any sort of utterance. This amen theory of truth does not explain why we should be allowed the conversational convenience in some cases but not in others. So if a functional theory of truth is to account for the non-indicative moods, the function of 'That's true' had better be more than that of 'Amen'.

A more promising approach lies in the observation that to call an utterance 'true' is to say that it is *correct*, or *right*; to call it 'false' is to say that it is *incorrect*, or *wrong*. As Dummett puts it, 'the roots of the notions of truth and falsity lie in the distinction between a speaker's being, objectively, *right* or *wrong* in what he says when he makes an assertion' (1978, p. *xvii*). We might be wary of Dummett's use of the notions of assertion and objectivity – these are the sorts of notions we want to explain – but the emphasis on the *normativity* of judgements of truth and falsity is a step in the right direction.

Whatever their origins, normative terms seem to function as verbal carrots and sticks, with which we can reward and punish certain sorts of behaviour. As with more tangible incentives, such rewards may be actual (for past behaviour) or merely conditional (on future behaviour). But unlike that of carrots and sticks, the reward or punishment bestowed in a normative judgement is in a sense self-inflicted. This not because we ourselves wield these verbal implements – that is true of carrots and sticks – but because their effect is conditioned and conventional. A child who has not learnt the meaning of 'good' and 'bad' cannot be encouraged to modify his behaviour by being told that it is good to eat his breakfast and bad to jump on the cat's tail (which is why parents are often tempted to resort to less convention-governed incentives).

Non-verbal incentives can also depend on convention, of course. A child who has not learnt the relevant conventions will not be moved by the offer of a fiver if he eats his breakfast. And just as

we might be taught to be ‘bad’, or to regard ‘Everything you say is false’ as a pat on the back, so we might learn to attach monetary value to almost any kind of physical object.

In general, at any rate, the use of such conventional rewards and punishments presumably serves some point. Why else should it have developed? In the case that interests us, we might hope and expect that the point of use of the normative terms ‘true’ and ‘false’ is such as to explain the ordinary limits of their application. We might hope to show, in other words, that the application of these terms to assertions serves a purpose that would not be served by applying them to, for example, questions and commands.

Again, we would thus be explaining ‘true’ and ‘false’ not analytically, but simply in terms of their functional role. We would be asking not what these terms refer to, but rather what they do for us (and hence why we might have developed them). This approach would be uncontroversial for a notion such as monetary value (or the valuations inherent in class distinctions, to return to our earlier example). However, it is apt to seem misguided in the case of normative notions such as truth, moral value and rationality. In the monetary case, we recognize that monetary value is conventional not simply as regards the objects we choose to attach it to, but also in the sense that the very notion of monetary value is clearly a social product (and hence conventional). But we are reluctant to allow that normative terms involve both these kinds of conventionality. The first kind is obvious enough: it is simply the point that we could have meant by ‘bad’ what we presently mean by ‘good’. But we are inclined to assimilate this with the point that, for example, we could have meant by ‘kipper’ what we actually mean by ‘sardine’. That is, we resist the suggestion that what we would be talking about in using these different words is also a social construct, rather than a natural feature of the world.

Accordingly, I think, we consign ourselves to some of the great treadmills of Western philosophy – the attempts to analyse the concepts of Truth, Reason, and the rest, and hence to understand their normativity and the role they play in our lives. These well-worn steps continue to seem more attractive than the conventionalist alternatives. A striking indication of the depth of this preference lies in the tendency on the part of philosophers who abandon the attempt to analyse truth in terms of correspondence to reality, to seek to analyse it in some other terms. We thus have the pragmatist tradition of coherence accounts of truth, of attempts to identify ‘true’ with ‘useful’, ‘accepted at the end of inquiry’, and the like. True, in the pragmatist tradition we also find the view that philosophy should not attempt to analyse truth. But even here there seems to be reluctance to take on the task of *explaining* our use of the notion of truth. Instead such pragmatists ‘would simply like to change the subject’ – to ‘suggest that we do not ask questions about the nature of Truth’ (Rorty 1982, p. *xiv*). This is the quietist tendency I referred to earlier, to blame the problem for the failure of the analytic solution. We shall return to it in chapter 9; but it is further evidence of the near invisibility of the explanatory alternative in contemporary thought.

Perhaps we suspect that in taking the explanatory approach, the conventionalist buys the easy life at the cost of what gives point to life, namely the very values in question. In other words, we fear that if we allow ourselves the explanatory perspective, allow ourselves to ask why we talk in these terms, we lose our entitlement to use them, to continue to behave as if they matter. We can cheerfully regard money as a social convention, because we could countenance getting by without it. But how could we be philosophers, or indeed persons, without a sense of the reality of Truth, Reason and the Good?

At best, this could amount to an argument for not attempting to develop explanatory account of normative terms – a version of the argument that there are some questions it is safer not to ask.

Even so, I think it exaggerates the risk. In other areas of life, we seem to be able to combine an awareness of the functional role of appetites and values with an undiminished ability to feel their force. Abstinence does not seem to be assisted by the knowledge that one's craving has a biological basis. It might be objected that there is a difference between these cases and those involving the use of normative terms such as 'true' and 'good': namely, that since the latter depend on linguistic convention, it is easier for us to give them up, or to cease to feel their force. But if the use of the word 'convention' is thought to carry this implication, we can do without it. The explanatory approach need not claim that with the benefit of an explanation of our use of these normative terms, we could (let alone should) do anything to change the practice. Some linguistic practices may perhaps be accessible to revision. But many, perhaps most, may simply be wired into us, in circuitry over which we have no control.

#### 6.4 THE CASE FOR SEMANTIC INCENTIVES

We have digressed, however. The suggestion was that we regard the normative notions of truth and falsity as the core of a conventional system of punishment and reward, and hence try to explain the limits of their application. In other words, we should try to show that the application of these terms to assertions serves a purpose that would not be served by applying them for example to questions and commands.

What purpose could this be? I suggest that the normative character of the notions of truth and falsity provides speakers with an incentive to justify their utterances to others, and hence to participate in an important form of argument or dispute. At its crudest level, it ensures that speakers are rewarded when their utterances are endorsed by others, and punished when their utterances are rejected. Thus it gives everyone an interest in participating in and attempting to 'win' the relevant disputes. It gives us all an interest in persuading others to endorse our own utterances (and hence, of course, in making sure that our own utterances are such as to be endorsed by others).

The normative nature of the terms in which such disputes are conducted gives them short term value to individuals, but cannot constitute their long term value to the linguistic community as a whole. On the contrary, the development of such a system of linguistic incentives suggests an underlying behavioural advantage – something that would explain the linguistic development in evolutionary terms. This advantage presumably turns on the behavioural consequences of the mental states exposed to criticism in such linguistic disputes. In particular it might lie in the survival advantages of basing one's mental attitudes on as wide a body of experience as possible.

Rough as it is, the approach already suggests an explanation for the fact which embarrassed Strawson's theory – namely that 'true' and 'false' can be used to endorse and reject only certain sorts of utterances. Perhaps not all utterances serve a function such that there is reason to seek or expect agreement between speakers. Consider requests, for example. Imagine that the two of us are talking to a third person, Fred. I ask Fred to buy us each a cup of coffee. Fred looks at you, but you say nothing. Our utterances on the subject conflict, in the sense that no one person could have both asked and not asked Fred to buy a cup of coffee. But clearly this 'conflict' may be entirely benign. There may be no reason to expect us to make the same requests, of Fred, or in general – and hence no advantage (quite the contrary) in a linguistic device that would encourage us to do so.

The proposal is thus that some utterances normally express states of mind with respect to which there is survival value in encouraging speakers to dispute, in the above sense. For these states of mind, two heads are better than one, and hence there is a general advantage in exhibiting

differences, and in providing incentives to encourage the spread of well-justified beliefs. The use of the normative terms ‘true’ and ‘false’ is an important part of the apparatus that language has developed to exploit this advantage. However, the apparatus is not deployed where it is generally not appropriate. The roles of utterances such as questions, commands and requests are not such as to provide an advantage in seeking unanimity. On the contrary, these utterances are such that different speakers may reasonably conflict (in the above sense), even if fully acquainted with each other’s point of view. So ordinary usage does not provide the incentives for disputes in such cases – perhaps marking their immunity with the non-indicative moods.

We thus have the beginnings of a functional account of truth, grounded on the claim that its primary role is to encourage and facilitate a useful form of linguistic dispute. The present sketch leaves a lot to be desired, of course. How, for example, is the claimed difference between indicatives and non-indicatives to be reconciled with the results of chapter 3 – with the apparent possibility of treating non-indicatives as disguised assertions? I shall return to the non-indicatives, and to this query in particular, in chapter 8.

It may seem also that the proposed account leaves no room for non-factualism. An interesting non-factualist thesis is bound to concern a case in which utterances are ordinarily treated as true and false. The non-factualist claims that although such utterances appear to be statements of fact, they are really something else. But the non-analytic approach leaves no room for a distinction between appearance and reality in this case – no room for categories underlying use in terms of which the non-factualist’s claim could be expressed. We shall see, however, that the account leaves room for a distinction of a different kind. In particular, the complexities of the relevant form of linguistic dispute give rise to a range of cases in which the optimum application of the normative notions of truth and falsity is far from straightforward. The ordinary linguistic apparatus is stretched to its limits, and the strain shows up as doubts as to whether the utterances in question ‘have truth values’ – whether, in other words, these notions are applicable at all. Not surprisingly, these are just the cases in which non-factualism has seemed attractive. Far from excluding non-factualism, this approach thus does a fair job of explaining its continued appeal.

It is true that the approach entails that traditional non-factualism is misconceived, in taking the analytic stance – in claiming to discern a categorical distinction underlying use. But in place of this misconception, the account offers two distinct sorts of question: the first as to whether the use of a given class of utterances is fully assertoric, or rather involves the identified departures from the canonical use of truth and falsity; and the second as to whether the actual use is the most appropriate, in view of the account’s understanding of the function of the notions of truth and falsity. Traditional non-factualist arguments can be expected to bear on both questions. Indeed, in the next two chapters, as we examine the structure of linguistic disputes, we’ll repeatedly appeal to the sorts of considerations that have been taken to support non-factualist treatments of particular topics.

In the next chapter our first job will be to show that the relevant notions of dispute and disagreement do not themselves depend on the notions of truth and falsity. If we want to explain the notion of truth in terms of its role in fostering certain activities, our characterization of these activities cannot depend on truth. Before that, however, I want to distinguish this approach to truth from another approach, also described as ‘explanatory’, which has received attention in recent years.

## \*6.5 TRUTH: ANOTHER ‘EXPLANATORY’ PROGRAMME

We have come to the task of explaining truth through that of accounting for the distinction between fact-stating and non-fact-stating discourse. We are interested in explaining rather than analysing truth because we have concluded that unless an analysis of truth draws on a prior understanding of the nature of statement, then, like Strawson’s amen theory and the disquotational accounts of truth, it will be unable to explain the fact that not all utterances are treated as truth-valued. In explaining the existence of a more than disquotational (and hence non-universal) notion of truth, we hope to explain, in effect, the existence of the non-universal linguistic category of statement of fact.

In recent years several writers have noted that although the ordinary notion of truth (or at least the philosopher’s ordinary notion of truth) seems stronger than a disquotational notion, it is hard to see why we need a stronger notion. The disquotational notion (or something of similar ‘strength’, such as a Tarskian truth theory) seems to be all that we need for a theory of linguistic understanding. And in psychological theory, it seems to be the ‘internal’ cognitive role of an attitude, rather than any ‘external’ relation of correspondence with the world, that does the explanatory work. The recent appreciation of these points, and of their significance for an account of truth, is particularly due to Hartry Field (1972, 1978).

Field asks ‘why we need to be able to apply semantic notions like truth and reference to the sentences that people believe and desire’ (1978, p. 48). This question seems to me to involve very much the same explanatory perspective as the approach suggested above. It sees the task as that of explaining the ordinary use of the notion of truth. (I think Field’s answer is inadequate, however – more on this in a moment.) But in the discussion to which Field’s work has given rise, the phrase ‘explanatory theory of truth’ has come to mean something different. The emphasis has shifted from explanation *of* (the use of) the notion of truth, to explanation *by means of* the notion of truth.

This shift is associated with a slightly different starting point: a philosophical commitment to a ‘realist’ world view, and hence an interest in justifying, rather than merely explaining, the use of the kind of notion of truth that such a view seems to require. This approach to truth is particularly due to Putnam (1978), who suggests that a realist notion of truth may be required to account for the fact that linguistic behaviour contributes to our overall behavioural success: ‘The *success* of the “language using program” may well depend on the existence of a suitable correspondence between the words of language and things, and between the sentences of language and states of affairs’ (1978, p. 100).

Putnam’s approach makes truth a theoretical notion, an element in the kind of story that a theorist would have to tell about language and its place and nature. Hence there is at least a danger that this theoretical notion will have nothing to do with the ordinary notion of truth. In what sense is it the same notion? And even if it is the same notion, why do ordinary speakers need it? These questions emphasize the difference between an explanatory theory of truth in Putnam’s sense, and a theory that seeks to explain the ordinary use of ‘true’ and ‘false’. In principle a theory could be explanatory in both senses. Truth might be a necessary part of an essential folk theory of language and the world. (Devitt 1984, chapter 6, seems to take this view, though not on Putnam’s lines.) But the two approaches are independent. Putnam’s does not guarantee ours, and ours does not depend on Putnam’s – folk notions may be explicable without being explanatory.

All the same, we can distill several proposals for an explanatory theory of truth in our sense from the work of Putnam, Field, and others. If any of these succeeds in explaining the fact that the ordinary notion of truth is not merely disquotational, then presumably it will explain (where the disquotational theory could not) the fact that truth is restricted to a particular sort of discourse. Our own suggestion would thus be unnecessary, or perhaps repetitious.

Putnam's own suggestion has been criticized (McGinn 1982, p. 222; Devitt 1984, p. 90) on the grounds that the explanation of the success of linguistic behaviour does not require a realist or correspondence notion of truth; a notion of warranted assertibility will do the same job. As Devitt puts it:

The role claimed for truth in [Putnam's] explanation must be ... as follows: the expectations that a belief gives rise to tend not to be disappointed if the belief is true but tend to be disappointed if it is false. True beliefs tend not to face recalcitrant experience. The problem is that warranted or justified beliefs tend not to either. (1984, p. 90)

It seems to me that there is a more serious objection, however. Not only does the explanation of success not need a correspondence notion of truth; all it needs is a disquotational notion. For the central claim, in effect, is that beliefs tend to lead to success if and only if they are true. And this is just the sort of generalization that has long been recognized to provide gainful employment for a disquotational notion of truth. The classic example is Ramsey's: 'He is always right', or 'Everything he says is true.' Ramsey points out (1978, p. 45) that if we allow ourselves to quantify over propositions (or sentences, or some such), we can read this as 'For all P, if he says that P then P.' (This and more complicated cases are discussed at length by Grover, Camp and Belnap 1975.)

In the present case the claim at the heart of the appeal to success is something like

(6.1) True beliefs tend to lead to successful behaviour.

Applying Ramsey's technique gives us

(6.2) For all P, if P then the belief that P tends to lead to successful behaviour.

If 6.2 does indeed mean the same as 6.1, then the notion of truth cannot really be doing any explanatory work. This is not to deny that we can explain our success by means of the hypothesis that most of our beliefs are true. It is simply to point out that in the formulation of this hypothesis, 'true' may be nothing more than a useful syntactical device, as the disquotational theories claim. This does not make the hypothesis empty. There is plenty of room for doubt as to whether

(6.3) For most P, if we believe that P then P –

or, as it might be better put, as to whether

(6.4) For all P, if we believe that P then it is probable that P.

If it is any good, the argument from success provides reasons to allay this doubt. But what we thus have reason to accept does not concern a more than disquotational notion of truth.

Moreover, it seems to me that this objection applies equally to the accounts that McGinn and Devitt suggest in place of Putnam's success-based explanation of a realist notion of truth – as it does to the following suggestion from Field, from which McGinn and Devitt develop their accounts:

I think that the reason why we need to be able to apply semantic notions like truth and reference to the sentences that people believe and desire is that we hold the theory that people's beliefs are, in many circumstances, reliable indicators about the world; and the only way to state this theory is to use the notion of truth .... *Moreover, this theory is not a piece of gratuitous metaphysics that could easily be dispensed with:* it is central to our getting information about the world, for we are constantly using our opinions about other people's beliefs in forming opinions about the world .... These inferences evidently proceed by means of certain *reliability principles*, principles that say under what conditions a person's beliefs about certain things are likely to be true. (Field 1978, p. 48)

A typical reliability principle is presumably something of the form

(6.5) Everything mother believes is true.

However, we have just noted that there is no need to regard such a use of 'true' as anything more than disquotational. What matters is that

(6.6) For all P, if mother believes that P, then P.

It is true that eliminating apparent references to semantic properties is more difficult when we believe only that mother is reliable on a limited range of topics. As Field says, 'we want to be able to say in the theory that some people have very reliable beliefs *about physics* but very unreliable beliefs *about the state of the economy*, and so forth' (1978, pp. 48–9). If mother knows her onions, we may want to say that

(6.7) Everything that mother believes *about onions* is true.

This becomes

(6.8) For all P, if mother believes that P and P is about onions then P.

But we are still left with the notion of *aboutness*, that Field takes to require a theory of reference. However, there is scope for treating 'about', like 'true', as an eliminable syntactical convenience. If we allow ourselves to quantify over predicates, for example, we can get close to 6.7 with

(6.9) For all *F*, if mother believes that onions are *F* then onions are *F*.

So it looks as though Field's reliability principles require neither truth nor reference, in the realist's sense.

McGinn and Devitt both modify Field's proposal by placing more stress on a broader notion of communication. Devitt says that we need to explain the fact that speakers both learn from and teach each other, and that 'it is these explanations that require talk of truth' (1984, p. 92). In so far as his account concerns learning, Devitt follows Field, emphasizing our need for judgements about the truth of the beliefs of others. As I have argued, these judgements need no more than a

disquotational notion of truth (which is to say that they don't need truth at all, so long as they are prepared to put up with some syntactical inconvenience).

Concerning teaching, Devitt notes also that a speaker S must believe that an audience A will take an utterance or other sign as an indication that S has a certain belief (that it is raining, in Devitt's example). What matters is that A

takes the behaviour to 'mean that' it is raining; or, better, he takes it to be true if and only if it is raining. He takes S to be sincere, and so infers ... that S has a belief that is true if and only if it is raining. He takes S to be reliable and so comes to believe that it is raining. (1984, p. 93)

Putting aside reliability, the question is whether A's belief about the significance of S's behaviour (or S's belief about that belief) needs to mention non-disquotational truth. If we are allowing non-linguistic behaviour, it is surely misleading to say that A must believe that S 'means that' it is raining. If S means that it is raining then he is saying so, and his behaviour amounts to a speech act. In this case what A must believe is first that S has said that it is raining; and secondly (since A believes S sincere) that S believes it is raining. The non-linguistic case requires even less: simply the belief that the behaviour in question indicates that S believes that it is raining. Thus, there doesn't seem to be a role even for a disquotational use of truth.

It may be that the point Devitt is aiming for here is the one suggested by McGinn. McGinn proposes

that we locate reference in the point of communication – in the intentions with which assertions (and other kinds of speech act) are made. A hearer understands a speech act as an assertion just if he interprets it as performed with a certain point or intention – viz. to convey information about the world. ... To fulfil the intention to communicate knowledge (or belief) about the world in language the speaker must exploit signs standing in representational relations to things in the world. (1982, p. 226)

However, this suggestion seems to involve the view we rejected in chapter 2: namely, that assertions are distinguished from other kinds of speech act in virtue of being 'aimed at truth'. McGinn seems to be proposing that in order to understand the point of an assertion, we must see it as made with such an aim; and hence that we require a notion of truth, thought of in terms of 'representational relations to things in the world'. The lesson of chapter 2 is that as an explication of the notion of assertion, this gets the relationship between truth and assertion the wrong way round. If we try to characterize assertion in terms of the notion of being aimed at truth, we cannot say why not all utterances should be regarded as assertions.

The conclusion of chapter 2 was thus that in order to understand truth, we must first understand the nature of assertion. That is, we need to be in a position to understand why it is assertions, particularly, that admit the standards of correctness and incorrectness embodied in the notions of truth and falsity. At the time, the possibility remained open that this called for an independent analysis of the notion of assertion, on the basis of which we could hope to explain the function of 'true' and 'false'. We have now rejected the analytic approach, however. As a result, we have to think not so much of the role of truth in the individual case, as of the function of the assertoric form of discourse as a whole. We have to think of truth not as a property uniquely possessed by linguistic entities of a certain kind, but as the most striking component in a pattern of linguistic behaviour, whose significance we want to explain. Instead of explaining truth in

terms of assertion, we want to explain the assertoric pattern itself, of which truth seems simply the most striking facet.

Another part of the same pattern is perhaps the notion of representation itself: the very idea of ‘signs standing in representational relations to things in the world’, as McGinn puts it. Traditionally the notion of representation has seemed to be prior to truth. Truth has been viewed as *correct* representation. In order of development I think that these notions might be better taken the other way. Truth might arise first, in virtue of its role in argument; and representation follow, in virtue of our attempts to understand the nature of the ‘correctness’ of true beliefs and utterances. However, the important point is that explaining truth and explaining our idea that signs represent the world are part of the same enterprise. So although it is true of truth, as McGinn says of reference, ‘that finding a place for [it] forces us to take a world-directed view of communication’ (1982, p. 226), this is not the basis of an explanation of truth, but simply a comment on the nature of the explanatory task.

Thus it seems to me that the recent search for an ‘explanatory theory of truth’, stemming from the work of Field, does nothing to provide an explanation of truth in our sense. Nor, incidentally, does it seem to make any progress towards a theory that would be explanatory in Putnam’s sense – that is, in which truth is itself an explanatory notion. As we saw, the purported applications of such a notion seem to need nothing more than the disquotational property of truth.

## 7

**The Power of Negative Thinking**

I have suggested that we try to explain the existence and application of notions of truth and falsity in terms of their functional role. In particular, I suggested that their main function might be to encourage a useful form of linguistic behaviour, namely reasoned argument. It seems that in virtue of their normativity, these notions enable and encourage us to favour opinions with which we agree and to disfavour opinions with which we disagree. Disagreements thus become socially unstable, and we are led to defend our opinions, to argue and to seek agreement. In the long run, I suggested, such disputes can be expected to have a beneficial effect on the behavioural dispositions of individual speakers. They help to ensure that as individuals we hold and act on attitudes that reflect, to some extent, the combined wisdom of our linguistic community. Our behavioural dispositions can thus be tested against those of other speakers, before they are put to use in the world. The guiding principle is that it is better to be criticized for claiming that tigers are harmless than to discover one's mistake in the flesh.

Obviously, the notions of disagreement and dispute do a lot of work in an account of this kind. A natural suspicion is that these notions themselves need to be explained in terms of truth and falsity. This would put paid to the proposal to base a functional account of truth on its role in promoting such disputes. In this chapter our first task will be to explore this objection. We shall see that there are several forms it might take, but that none is inescapable. In the process we shall be led to a clearer account of the origins and conduct of such disputes and a better understanding of their behavioural advantages. That in turn will prepare the ground for the chapter's second main concern: to show how this functional approach to truth leaves room for non-factualism.

## 7.1 TRUTH AND THE TRIGGERS OF DISSENT

In its simplest form, the circularity objection would be the claim that a disagreement (of the relevant sort, at any rate) about an assertion P is a disagreement about the *truth* of P. However, in this form the objection is easily dismissed: for a dispute about the truth of P would have to be a dispute about the truth of 'P is true', and so on, *ad infinitum*.

A more interesting form of the circularity objection: How is it that we recognize the existence of a disagreement (and hence the grounds for a dispute) with another speaker, if not in virtue of the perception that he or she has said something *false*? The proposed account claims to explain our use of 'true' and 'false', in terms, primarily, of their role in these disputes. What it claims to explain, in other words, is our habit of describing the grounds for argument as being that our fellow disputee has said something *false*. The claim is that our use of this normative term serves the function of providing an incentive to the other speaker to enter into a dispute.

The account thus aims to explain the ordinary speaker's perception of the basis of a disagreement: the perception that someone else has spoken *falsely*. In order to explain it, however, the account needs to appeal to some more basic feature of the cases in which it takes this perception to be appropriate – to say that (perhaps unconsciously) we recognize the grounds for a dispute in terms more basic than the perception that someone else has spoken falsely. Similarly, the cue that a dispute is resolved cannot simply be the perception that the other disputee now speaks the truth, for the account wants to explain our inclination to describe it in these terms. So what could be the cues of argument, if not our judgements concerning the truth and falsity of the utterances of others?

A note here on terminology: I take dissent to be analogous to judgement rather than to assertion – that is, it happens in the head. A dispute begins when someone expresses dissent. Presumably some of the utility of truth and falsity lies in their encouraging dissenters to speak up. But even silent dissenters may profit from the experience, by reconsidering their own beliefs.

A speaker makes an assertion, another expresses dissent, and a dispute is under way. We want to know how the latter speaker came to believe that dissent was appropriate. And we want to avoid the suggestion that it is a matter of coming to believe that the former speaker spoke *falsely*.

The mistake thus avoided would be that of pitching the cues of dissent too high – of making them depend on perceptions of truth and falsity. We should also be careful not to pitch them too low. We want to say that the biological utility of linguistic disputes stems from the behavioural consequences of the states of mind that come under challenge (albeit indirectly, in the form of their linguistic proxies) in such a dispute. So it is important that dissent be cued to something of behavioural significance. However, these behavioural consequences are long run affairs, and no doubt hideously complicated. It would be absurd to require that in individual disputes, speakers are cued by an understanding of the behavioural significance of the opposing points of view. Moreover, this would actually discourage disputes in cases in which the behavioural significance of differing beliefs is balanced by differing desires: I want to see Fred, and believe he is in the kitchen; you want to avoid him, and believe he is not in the kitchen; we both head for the kitchen, but one of us will be disappointed. So it is important that dissent be cued not by an awareness of the behavioural consequences themselves, but by some manifestation of the mental states that, in the long run, have those consequences. We need to separate the practice of dissent from the underlying facts that explain its development.

We want to be able to say that the triggers of dissent lie somewhere between these two extremes – between perception of truth values themselves, and perceptions of behavioural consequences. The answer might seem obvious: shouldn't a speaker simply look for cases in which he asserts 'P' and another speaker asserts 'Not-P' (or *vice versa*)? This is too quick, however. For one thing, speakers may express conflicting points of view in completely different words. There need be no easy syntactical clues. More importantly, your 'Not-P' may not conflict with anything I have *said*; but if it conflicts with something I *believe* (i.e., that P), then I ought to dissent. So it appears that we need to credit speakers with the competence to discern that an utterance expresses a mental state incompatible with their own beliefs. There seem to be two components to such a competence. The first is the ability to *interpret* an utterance – to see what a speaker means. The second is the ability to detect incompatibilities between one's own beliefs on the one side, and a claim embodied in a received utterance on the other.

On some views of meaning and interpretation, the former ability depends on the notion of truth – for to interpret an utterance is to know under what conditions it would be true. However, we have seen that such views seem to require only a thin, disquotational notion of truth. This was the basis of our objection in chapter 2 to the proposal to derive a substantial theory of truth from a Tarskian truth theory, constrained by the needs of radical interpretation. So the ability to interpret an utterance does not seem a source of circularity in the proposal to base a thick notion of truth on its role in encouraging argument. (I do not mean that it is clear what the notion of meaning and interpretation will come to under such a proposal; I discuss this in chapter 9.)

This leaves the ability to detect incompatibilities between beliefs and hypotheses. The present view is not alone in crediting speakers with an ability of this kind. The question is not whether we have such an ability, but whether it depends on a grasp of the kind of notion of truth we want dispute behaviour to explain. As far as I can see, such a view could only arise from the idea

that in judging that beliefs X and Y are incompatible, we are really judging that these beliefs could not both be true. This use of ‘true’ is less obviously thin and disquotational than some we have encountered, but only because it is applied to beliefs rather than to sentences. But why suppose that incompatibility judgements need to be *about* beliefs and hypotheses? What speakers need is not the ability to reason about their own mental states, but rather the ability to employ these mental states in reasoning about the world in general. Prompted by the suggestion that Q, they need not to decide that the hypothesis that Q is incompatible with their belief that P, but simply to employ their belief that P in deciding that not-Q. It is a nice question just how we notice such incompatibilities. But it is hard to see why anyone should think that perceiving that P and Q could not both be true is any more easy, or primitive, than perceiving that P implies not-Q, or than utilizing one’s belief that P in ‘perceiving’ that not-Q.

Thus for the moment I take it that what triggers dissent from the assertion that Q is simply the belief that not-Q. (Here ‘belief’ need mean nothing more than ‘disposition to assert’.) This invites the objection that a grasp of negation itself depends on a grasp of the notions of truth and falsity, along the standard truth-functional lines. I have some sympathy with this objection, to the extent that I doubt whether we are entitled to help ourselves to the notion of negation in this way. I think it would be more accurate to say that negation has its origins in the expression of dissent. However, I doubt that this makes dissent dependent on falsity. Instead the view might be something like this: in acquiring a language, we learn to assent to sentences in some circumstances and dissent from them in other circumstances; dissent, like assent, is triggered by the activation of the sorts of habits thus acquired. Such a view of linguistic competence has independent advantages. (In Price 1983b I argue that it enables us to block Dummett’s move from manifestation constraints on linguistic knowledge to a non-classical logic.) But I cannot defend it here.

## 7.2 TRUTH AND THE WORTH OF BELIEF

Let us turn to another form of the circularity objection. According to the proposed account, the utility of argument (and hence our use of ‘true’ and ‘false’) lies in its tendency to ‘improve’ the commitments with which we meet the world. Thus it rests on the principle that some commitments are ‘better’ or more valuable than others. But what is the measure of the worth of a belief, except its truth or falsity? Isn’t a belief poor, or incorrect, just in case it is *false*? The objection is thus that in explaining the function of ‘true’ and ‘false’, the proposed account will have to appeal to the idea that commitments can be true or false in some prior sense.

The way to meet this objection is to note that on any plausible account, truth cannot be the only measure of the worth of belief. For imagine for a moment that there were a substantial property of truth, possessed by certain of our mental states. Why should we aim for beliefs that possess this property? Not for their own sake, presumably, but because such beliefs are valuable in some other sense – because they dispose to behavioural success, for example. Vague as this notion is, there is clearly no simple correlation, on anyone’s view, between truth and behavioural appropriateness. False beliefs can lead to good outcomes, true beliefs can lead to bad ones, and many beliefs in fact have no behavioural consequences. Unlike truth, behavioural ‘rightness’ is bound to be a long run matter.

So from the point of view of behavioural utility, true belief is not an end in itself. Truth is at best an end in virtue of its usual consequences – something we should aim for in virtue of the long run behavioural advantages of doing so. If one already takes for granted a view of truth as a substantial property of mental states, this will be the natural way to construe its role in discourse; though it will need to be explained why this property should be a good guide to behavioural

success. This is not the only possible view, however. Our proposal seeks to *explain* the ordinary notion of truth in terms of its role in discourse. Instead of beginning with the conviction that truth is a substantial property, we begin with the observation that ordinary speakers display such a conviction (and then seek to explain this fact). We thus regard truth as a ‘mythical’ goal of enquiry, whose popularity is to be accounted for in terms of the benefits of subscribing to such a myth. Of course, it remains to be shown that the main characteristics of the notion of truth can be explained on such a basis. However, the one characteristic that is immediately explained is the fact that truth is treated as the ordinary goal of enquiry. Here both sides agree on the appearances; and only a prior commitment to a more substantial view of truth could support the objection that our view fails to make sense of the fact that truth is the real end that we aim for in revising our beliefs.

The above objection calls also for the following reply. In the end, far from construing the benefits of argument as the production of true beliefs, our account should not depend on the notion of *belief* at all. As I said in chapter 6, our use of belief at this stage is no more than an expository convenience. It is dispensable, in principle, in favour of talk of behavioural dispositions. If this were not so then the present objection would be rather beside the point. A well-founded notion of belief, or indeed of representational attitudes in general, would bring with it not only an account of truth but also a basis for the fact-stating/non-fact-stating distinction. So relying on a substantial notion of belief would be damning in itself. The worst that could be said of relying on a substantial notion of *true* belief would be that it opens the stable door after the horse has bolted.

Thus far in this chapter we have been mainly concerned to show that the notions of disagreement and dissent do not themselves depend on those of truth and falsity. It is time to pay more attention to disputes themselves. We are interested in the nature and origins of their advantages, and hence in showing that in order to exploit these advantages, the goal of argument must have more or less the characteristics of the ordinary notion of truth. And we want to be on the lookout for the origins of non-factualism.

### 7.3 THE VALUE OF ARGUMENT

Arguments arise from disagreement and are resolved by agreement. In some sense, they thus serve to diminish disagreement among the members of a linguistic community. What could be the point of such a function? Why should disagreement be collectively or individually disadvantageous?

Let us for a start approach these questions in terms of an orthodox notion of belief. On such a view, beliefs have truth values, and a disagreement entails that at least one of the speakers involved has a false belief. This suggests that the disadvantageousness of disagreement rests on that of false belief. Resolving disagreements helps to correct false beliefs – not infallibly, of course, for a dispute may lead both parties to false beliefs – but perhaps more often than not.

But what is wrong with false beliefs? Here, as we saw, the answer will presumably be that false beliefs tend to lead to inappropriate behaviour. ‘Tend to’ is as strong as the claim can go – as we have seen, it is possible for mistakes at this level to cancel out, in virtue of the holistic nature of decisions to act. So on the standard view the justification for argument is something like this: verbal disagreements – for simplicity, pairs of utterances of the forms P and not-P – tend to indicate that one participant holds a false belief. (This ‘tend to’ stems from several sources, but perhaps most significantly from the fact that even the most well-intentioned conversationalists

may not ‘mean the same thing’ – express the same belief – by saying P.) False beliefs tend to be a behavioural handicap. And argument tends to replace false beliefs with true beliefs.

I want to suggest that the orthodox notion of belief – of a truth-valued mental state – plays no significant role in this account. What matters is a certain correlation between utterance and action in general: roughly, that within a speech community, utterances of a given sentence tend to be correlated to mental states with similar causal or functional roles in the determination of behaviour. An utterance type is thus correlated with a class of mental states that share what I shall call the *Same Boat Property* (SBP). A class of mental states have the SBP if their typical behavioural consequences are such that their behavioural appropriateness, or utility, is predominately similar across a speech community. If a mental state has the SBP, then if it is appropriate for any one of us, it is appropriate for all – we are all in the same boat.

The notion of the SBP is far from precise; and not, I suspect, capable of being made significantly more precise. But provided that we do not expect it to bear too much weight, I think it will be useful at several points. Later it will help us to explain some of the variations from the statement form that are associated with non-factualism. More immediately, it clarifies the role of truth in the above account of the benefits of argument. For given the SBP, we can offer a justification for argument very much like the one outlined above: verbal disagreements tend to indicate that one party is behaviourally disadvantaged (or potentially so); arguments tend to improve matters. This does not yet establish that in the earlier version, the standard conception of belief is redundant. After all, it might be the standard conception that entitles us to the SBP (for the class of mental states typically correlated with a given utterance). Indeed, it may seem that the required case of the SBP is a trivial consequence of the truth-conditional view of belief; that beliefs with the same truth condition are clearly appropriate in the same circumstances – namely, when and only when their truth conditions obtain. However, this ignores the fact that there is no simple correlation between truth and appropriate behaviour. So the orthodox account above of the value of agreement does not *derive* the SBP (from sameness of truth conditions); it assumes it, in the form of the assumption that throughout a community, false beliefs tend to be a behavioural liability.

The principle that some mental states have the SBP can be regarded as a theoretical hypothesis. The account requires that it hold to a sufficient extent, in some form, to explain the utility of a procedure whose function is to encourage resolution of certain disagreements between the commitments of different speakers. In the end, if need be, it can derive its plausibility partly from that of the theory that requires it; and partly from the demonstration that particular violations of the principle turn out to be correlated with departures in usage from the procedure that it explains.

The step from the hypothesis that certain of our mental states have the SBP to the claim that argument is advantageous depends on two main assumptions. One is that this is a case in which individual interests recommend co-operation, which of course is far from true in general, even within a single species. But this is hardly likely to be controversial, given that language itself is a co-operative and social activity. The co-operative basis of dispute behaviour seems simply a special case. (After all, we are free to avoid disputes – to decline to co-operate – if it seems to our advantage.) More importantly, the step depends on the assumption that in general, at any rate, pushing for agreement does more good than harm. Clearly disputes can be counter-productive, leaving both participants with inappropriate commitments. However, it seems reasonable to assume that on average argument does improve matters, both by encouraging reconsideration of individual grounds for judgement, and by the effect of pooling of evidence.

This account of the function of disagreement and argument thus rests on the notion that in virtue of their behavioural consequences, direct and indirect, some commitments are more appropriate than others; and moreover, that the appropriateness of a given commitment is more or less constant throughout a linguistic community. These factors enable individual speakers to benefit from a social activity, made possible by language: the drive to seek agreement. This drive is fuelled by the conventional notion of an external standard for belief and judgement, in virtue of which disagreements are socially unstable. The concept of truth is the catalyst that drives our systems of commitment through reactive argument, in search of the equilibrium that comes with agreement.

So why does disagreement matter? The account gives two answers. Internally, from the participant's perspective, disagreement matters because it is the cue for a mode of criticism – we criticize judgements with which we disagree and are criticized ourselves in turn. Externally, from the explanatory perspective, disagreement matters because in the long run it is an indication of behavioural liability. How are these perspectives related? The account suggests that the external explains the existence of the internal – or more specifically, of the special notion of *correctness* on which the internal perspective rests.

Note that agreement is not an end in itself, on this view, on either the internal or external levels. Externally, agreement need have no value of its own, but only the contingent virtue that stems from being the normal result of an otherwise useful process. And internally, it is simply the state in which disputants accept that they have achieved their common goal – that is, as they see it, the truth of the matter.

Indeed, it is crucial that disputants should not treat agreement as an end in itself. In the absence of any felt allegiance to an external goal, agreement is easily engineered – for example, as in many animal societies, by a system of deference to authority. I don't mean to suggest that argument is free of deference to authority, and such like. Clearly it isn't, in practice. Nor do I deny that such ways of reaching agreement can be valuable, even in human societies. The point is rather that there is a quite different benefit, that stems (in practice imperfectly) from the conception of an external goal for assertion and judgement. This benefit comes not from the social benefits of agreement itself, but from the improvements in individual behavioural performance that stem from the reactive process of which agreement is the stable state.

Imagine, for example, a community of creatures who use sounds as signals for a variety of purposes. Certain sounds might be uttered in response to significant features of their environment – food, dangers of various kinds, and so on; others, perhaps, in advance of significant kinds of behaviour, not otherwise predictable by other members of the community. In both cases it is easy to see how agreement – uttering the same sounds at any given time – might be valuable in itself. To the extent that the community's success depends on co-ordinated behaviour, it will be important that the members of the community are disposed to the same behaviour at the same time.

This aside, however, there might be a quite separate point to noting and resolving disagreements. It might enable individuals to improve their own behavioural commitments, even when these commitments have no role to play in the co-ordinated behaviour of the community as a whole. Such resolutions could be achieved in a number of ways. One possibility, obviously appropriate in some cases, would be to defer to experts. On visual matters, bats would do well to defer to hawks; on auditory ones, the reverse. (This doesn't mean that hawks and bats need co-ordinate any resulting behaviour, of course.) On some matters, however, the community might possess no experts. It would then seem advantageous for everybody to be encouraged to have their say;

in other words, for disagreement to be resolved by argument among equals. I am suggesting that it is in this context that the normative notions of truth and falsity are important. They are at least part of what it takes to encourage the members of a community to argue when disagreements are noticed. The method is not infallible, of course. Once in place, for one thing, truth itself becomes a matter on which authoritarians may exercise authority. But an imperfect system of norms may be very much more useful than no system at all.

#### 7.4 TWO COMPARISONS: ELLIS AND HABERMAS

The above point marks one of the main differences between the account of truth I am proposing here and one of the very few explicitly explanatory approaches to truth in the recent philosophical literature: the account suggested by Brian Ellis, in a paper called ‘Truth as a mode of evaluation’ (1980). Ellis distinguishes ‘objective and subjective modes of speech’:

To say that *p* is true, or more conventionally, just *p*, is for me to express my belief that *p* objectively. But for me to say that I believe that *p* is to express my belief that *p* subjectively .... In speaking in the objective mode, the belief or attitude we express is disembodied and put forward neutrally for discussion as though it were a thing that had some independent existence .... The participants in the discussion see themselves as arguing about this neutral detached object, and deciding whether to accept or reject it. (1980, pp. 88–9)

Ellis associates truth with the objective mode of speech. He leaves it unclear, I think, why this mode should employ both the forms ‘*p*’ and ‘“*p*” is true.’ That point aside, he suggests that objectification is a

manifestation of a process, ... necessary for human communication, which makes it possible for human beings to develop what is roughly a *communal* system of beliefs and attitudes, which in turn is necessary for co-ordinated and co-operative action .... We are social beings, and our survival depends on cooperative activity. How we act normally depends on our desires, values, and belief systems .... And we may suppose that intersubjective agreement in our belief and value systems at least *facilitates*, even if it is not a necessary condition for, such activity. (1980, pp. 89–90)

Thus it seems that Ellis sees the advantage of the objective mode almost entirely in terms of benefits of agreement *per se*. It is true that for social creatures agreement or commonality of purpose is often useful, indeed indispensable. As I have said, however, I don’t think that this can explain why agreement should be sought in argument, rather than say imposed by authority. In so far as what matters is simply that we all do the same thing, it would be more economical for evolution to arrange that we all defer to a leader. The fact that we take the trouble to argue suggests a further benefit, not guaranteed by easier paths to agreement. I have suggested that this benefit lies in the tendency of argument to improve a participant’s individual adaptation to his or her environment. I think there is something of this idea at only one point in Ellis’s paper:

Basically, we are people with belief systems who are concerned to interact with each other, *and so to widen the bases of experience upon which our belief systems depend*, and to reach intersubjective agreement about things. (1980, p. 98, my italics)

In virtue of its emphasis on *agreement*, Ellis’s approach to truth seems vulnerable to the usual objection to coherence theories. It is often felt that such theories fail to make sense of the intuition that truth has something to do with answerability to some *external* standard – the

intuition, to put it crudely, that we can agree until we are blue in the face, yet still be *wrong*. Correspondence theories try to make sense of this world-imposed ‘rightness’, as the basis of an *analysis* of truth.

Like Ellis’s, the present approach rejects this attempt at analysis, in favour of an explanation of our possession of the notion of truth. Unlike Ellis’s, however, my account retains the intuition that world-imposed rightness has something to do with truth. On my account it is crucial to the explanation of truth that some commitments are (or tend to be) more appropriate than others. This ‘appropriateness’ is world-imposed. Behaviour that is successful in one world would be disastrous in another. This notion of appropriateness is not an analytic base for truth, however, but part of the explanatory base. In the terms I distinguished above, appropriateness belongs to the external perspective, whereas truth belongs to the internal perspective.

I have emphasized the argumentative role of truth and its contrast to authority as a means of achieving agreement. These two points may call to mind another similarity the present account bears to contemporary work on truth. Both are prominent in the theory of truth developed in recent years by Jürgen Habermas. Here I can do little more than to note this connection. Two cautions seem to be in order, however. For one thing, it is unclear to what extent Habermas’s theory is intended or can be regarded as an explanatory account of truth. Certainly it has been interpreted as a variant of a Peircean consensus *analysis* of truth (see Pettit 1982, for example). Secondly, it is far from clear that Habermas would countenance the naturalistic and evolutionary flavour of the account here proposed; Hesse (1978, p. 390) suggests that ‘any such interpretation of human evolution seems to fall foul of Habermas’ own strictures against Chomsky’s naturalism [cf. Habermas 1970] – it is itself a social theory and hence presupposes its own normative standpoint.’ (I return to this kind of objection in chapter 9.)

## 7.5 ARGUMENT AND THE SOURCE OF BELIEF

We have been discussing the value of disagreement and argument, and hence trying to account for the nature of the ordinary notion of truth. So far, we have characterized the benefits of argument in terms of its effect on the very commitments whose linguistic expressions are the overt matters of dispute. It is time to recognize another benefit. An inappropriate commitment may be not only a behavioural liability in itself, but also a manifestation of a potentially more serious liability – a sign of a flaw in the system of background beliefs, rules or habits of inference, from which the commitment in question arose. It seems that disputes have the function not merely of correcting individual mistakes, but also of encouraging improvements in the general beliefs and dispositions that lead to those mistakes.

It will be helpful at this point to appeal to a simple model of the origins of belief. Let us say that a person typically forms a belief in virtue of two things: a prior *evidential* belief, and a general *inferential* disposition to adopt a belief of one kind, given a belief of another kind. Since our use of this model will be largely heuristic, accuracy need not be one of its virtues. To our folk-psychological intuitions, however, many cases of belief formation do seem to be of much this kind.

The model allows disagreements to be traced to two possible sources. If we disagree about the truth of some proposition P, we may have had different evidential beliefs, or we may have employed different rules of inference. Our disagreement may turn out to be *evidential* or *inferential* (or a combination of both, of course). And our dispute, if we engage in one, has the potential not only to rid one of us of a behaviourally disadvantageous attitude to P, but also to correct the source of that mistake.

I think it would be an interesting and profitable task to examine the structure of such arguments in more detail. The model seems to throw light on the connection between truth, inference and validity. It seems that in the process of argument, speakers attempt to trace the claimed ‘incorrectness’ of their opponent’s assertions to its roots in any proffered defence. At the same time they try to import ‘correctness’ from the claimed roots of their own position. Central to this is the notion of an inference that claims to preserve ‘correctness’, and can itself be challenged on the grounds that it fails to do so. However, our immediate concern is with another application of the model. As yet there is nothing in the notions of disagreement and argument to indicate that there is a place for non-factualism. This model will help us to explain the limits of truth, and hence the origins of non-factualism.

In section 7.3 I emphasized the importance of the SBP (‘Same Boat Property’), only in virtue of which can it be true that if two people hold conflicting commitments, then one of them is likely to be behaviourally disadvantaged. Without the SBP there would be no reason to encourage agreement, and hence none to treat disagreement in evaluative terms – as an indication that someone has made a *mistake*.

Given that we do think in these terms, argument seems to be an attempt to trace the origins of the disagreement to some evidential or inferential basis. In effect, argument consists in a search for what both parties will agree to be the original mistake. In general, in other words, we argue as we should if the justification for evaluative treatment extends to the precursors of the belief at the centre of an initial disagreement; as we should if the SBP applies as much to evidential beliefs and rules of inference as to the judgements that lead us into conflict.

However, it turns out that this extension of the SBP is not always in order. There are cases in which it seems proper that speakers be considered to be *initially* in different boats – and therefore initially *justified* in disagreeing – even though not justified in preserving that disagreement, when each becomes aware of the other’s point of view. We shall see that such cases populate the margins of factual discourse. It is in these ill-defined fields of usage that non-factualism finds its place.

## 7.6 PROBABILITY AND THE END OF ARGUMENT

The most striking cases, in my view, are those involving probabilistic judgements. Such judgements reflect evidence that is less than complete, in the following sense. If I am interested in whether P, then I make a judgement such as ‘It is probable that P’, or ‘There is an  $n\%$  chance that P’, only because my evidence does not permit me actually to determine whether P. I do the best I can with the limited evidence available to me.

However, my evidence may differ from yours. It need not be that either of us is mistaken about our evidence, but simply that we have access to different facts. So my best judgement about the probability that P may differ from yours. But when I declare that it is probable that P, you may well reply that on the contrary, it is rather unlikely that P. There appears to be a disagreement, the basis for an argument. When the facts emerge, however, we needn’t insist that in the first place one of us must have been mistaken. On the contrary, I may well recognize that in your place I would have made your assessment of the probability that P, and *vice versa*.

It is important to appreciate that the SBP does still apply in a limited form. Once we become aware of the difference between our judgements of the probability that P, then we are in the same boat. On the one hand, the behavioural consequences of probabilistic judgements are the same for both of us (subject to the usual qualifications). And on the other, since in principle we

now have access to each other's evidence, there is no longer a systematic difference between us, such as to justify different judgements about the same probability. So there is a reason for us to seek agreement, and hence for convention to apply the normative incentives that encourage us to do so.

The fact that the SBP does not generalize to the *initial* disagreement suggests that in a properly organized linguistic community, speakers would be prepared to withhold criticism of each other's initial utterances. Indeed, they may wish to commend, rather than criticize, if it becomes apparent that the initial difference of opinion rests on their each having access to a different body of evidence. So if 'true' and 'false' are evaluative terms whose function is to encourage useful revision of judgements and patterns of inference, then we should expect that speakers who discover that a disagreement is of this kind would be reluctant to say that the other party's initial claim was *false*. And this does seem to be what happens. When we discover that our disagreement has this sort of basis, we are content to say that we were both correct, given the evidence we then had; and we go on to consider the bearing of the combined evidence.

Some terminology: I shall say that in such a case the initial disagreement *evaporates* or becomes *insubstantial*. A disagreement is *substantial* so long as the parties concerned regard each other as mistaken, and evaporates when such evaluations are no longer felt to be appropriate. We shall see that this can happen in several ways.

In the probabilistic case, disputes seem to run as follows. Any disagreement is initially treated as substantial. If it then transpires that the disagreement rests simply on the fact that the disputees originally had access to different bodies of evidence – in other words, that each recognizes that the other has made best use of his or her own evidence – then the original disagreement evaporates. Once both parties have access to their combined evidence, however, any remaining disagreement is treated as substantial.

What determines whether a disagreement about a particular sort of judgement should always be treated substantially? The following answer might suggest itself: a disagreement is or at least should be treated as substantial if it is a dispute about a matter of fact – if it answers to an objective reality, in virtue of which one party must be mistaken, as long as they disagree. As will by now be apparent, I think that this gets things back to front. I suggest that judgements of a certain kind are factual just to the extent that we do treat all disputes involving such judgements as substantial. Evaporative disagreement marks the limits of what we might characterize as the factual pattern of usage. In other words, the crucial feature of the factual pattern is that it requires us to say that of any pair of conflicting judgements, at least one must be false.

In chapter 6 I distinguished the descriptive and explanatory components of a non-analytic approach to the notion of statement of fact. This characterization belongs on the descriptive side. The job of the descriptive part of such a theory is to characterize the manifestations of the distinctively fact-stating form of language; and to do so in such terms as to enable us to explain its existence. I think that by focusing on dispute behaviour – especially on the extent of substantial disagreement – we can do much to explain the relevant distinctions we find in usage in terms of the value and limits of argument. At the same time, I think we can do considerable justice to non-factualism, at least in its more important applications. For we can show that the intuitions that seem to support non-factualism are very much those that incline us to leave room for evaporative disagreements. Non-factualism seems largely a product of these intuitions, viewed through the lens of philosophical analysis. The present approach rejects the analytic perspective, but it enables us to explain the intuitions that that perspective distorts.

In the remainder of this chapter I want to defend the suggestion that factual discourse be characterized in these terms against a variety of objections. In particular, it needs to be shown that the distinction between substantial and insubstantial disagreements puts the limits of factuality in roughly the right place. By this I don't mean that we can tell in advance where the right place is; but simply that it would be a *reductio* of the proposal to characterize statements in this way, to show that evaporative disputes commonly occur about utterances whose factual nature is not in doubt. So to begin with we need to show that what gives rise to evaporative disputes in the probabilistic case is not a feature of assertoric utterance in general. We shall also want to show – anticipating a likely objection – that evaporation cannot adequately be explained by the supposition that probabilistic judgements are indexical.

## 7.7 COMMITMENT AND THE CASE FOR TOLERANCE

We have observed that probabilistic disagreements are liable to evaporate when they turn out to stem from the fact that speakers have relied on different evidence. The same might seem true of almost any kind of judgement. It often happens that although we disagree with another speaker, we recognize that in her position, with her evidence, we would have come to the same conclusion ourselves. Accordingly, we are inclined to excuse her mistake. Doesn't this show that almost any assertoric utterance can give rise to an evaporative dispute?

I think that these cases are actually quite different from those that can stem from probabilistic judgements. The difference shows up in the use of the labels 'true' and 'false' in the two sorts of case. In the probability case, we observed that we are reluctant to say that a probabilistic judgement is false when it turns out to have been properly based on different evidence from our own. On the contrary, we are likely to say that it was 'true, given the speaker's evidence', or something of the sort. But now consider a case of the other kind. In fading light, a guest mistakes my rabbit for a rat. 'There's a huge rat in your garden!', he cries. I reassure him, and politely excuse his mistake: 'That rabbit *does* look ratty at this time of night.' Excusing the mistake is not the same as acknowledging that no mistake has been made, however. If it came to the point, then, clearly, I would continue to maintain that the guest's assertion was false. From his point of view, it was understandable, perhaps, but certainly not true. There is no room for points of view on the question as to whether a creature is a rabbit or a rat.

The special character of probabilistic judgements shows up particularly in the following fact. It can be reckoned correct to make a probabilistic judgement even if one knows of the existence of evidence which, if one had it to hand, might make one revise that judgement. For example, a doctor might say, 'You are probably not infectious, but if we had your test results we would have a better idea'; or even, 'You are probably not infectious, but the test results will tell us one way or the other.'

Contrast these examples with the following: 'You are not infectious, but if we had the test results, we would have a better idea'; or 'You are not infectious, but the test results will tell us one way or the other.' (Or: 'That's a rat, but in better light it might turn out to be a rabbit.') Such assertions take away with one hand what they have given us with the other. Utterances such as 'You are not infectious', or 'There's a rat in the garden', seem to be expressions of 'final' commitment, in the sense that to accept such an assertion is to accept that further evidence is unnecessary. If there is a rat in the garden, then any evidence that seems to suggest the contrary is misleading. As long as we are committed to claim that there is a rat, we are committed to the view that the case is closed – additional evidence is mistaken or irrelevant, if not confirmatory. The above examples violate this 'closure principle', explicitly rejecting the commitment to ignore further evidence.

Gilbert Harman has discussed this feature of commitment. In a recent book on reasoning he observes ‘that people find it difficult to accept things only tentatively .... There is a strong tendency quickly to convert such tentative acceptance into full acceptance’ (1986, p. 49). Harman characterizes ‘full acceptance’ as follows:

Belief in or full acceptance of P involves two things. First, one allows oneself to use P as part of one’s starting point in further theoretical and practical thinking. Second, one takes the issue to be closed in the sense that, when one fully accepts P, one is no longer investigating whether P is true. ... Full acceptance of P may therefore involve an implicit commitment to the claim that further investigation of P would not be worthwhile, for example, in leading one to discover further evidence that would lead to a reassessment of one’s conclusion. (1986, pp. 47–8)

It is a nice question why we make such final commitments. Clearly, they are often unjustified. The very fact that we can appreciate this, that we can change our mind, shows that in a wider sense our actual commitments are not final. On the whole, we do not close off the possibility of conflicting evidence. Why then don’t we express ourselves in a form that allows for this possibility? Why doesn’t my startled guest cry, ‘There is probably a large rat in the garden!’? Or more interestingly, why isn’t ‘There is a large rat in the garden’ simply understood in this way, so as to allow for insubstantial disagreements? (The interpretation would not have to be probabilistic. ‘It seems that P’ has much the same effect.)

A possible answer is that in the development of language full acceptance and assertion simply came first. Perhaps we could do without it, at least in theory. But in practice it developed before the various forms of qualified assertion, and so we have both. Harman suggests a less contingent constraint: that in practice there may be no alternative to full acceptance, for creatures of our limited mental abilities.

If one had unlimited powers of record keeping and an unlimited ability to survey ever more complex structures of argument, replies, rebuttals, and so on, it would be rational always to accept things only tentatively as working hypotheses, never ending inquiry. But since one does not have such unlimited powers ..., one is forced to limit the amount of inquiry in which one is engaged and one must fully accept most of the conclusions one accepts, thereby ending enquiry. Tentative acceptance ... cannot be the general rule. (1986, p. 50)

These factors suggest that most commitment is full commitment, and hence that most expressions of commitment are normally expressions of full commitment. But they do not seem to entail that ordinary expressions of commitment must always be taken as expressions of full commitment. That is, they seem compatible with the possibility that all or most assertions can in exceptional circumstances be the basis of an evaporative disagreement. I do not want to deny this possibility, which in my terms is the suggestion that the (descriptive) distinction between statements and non-statements is always a matter of degree. But the issue is somewhat premature, until we have a clearer idea of the range of possible ways in which utterances can give rise to evaporative disagreements. For the moment I want simply to make the point that probabilistic judgements are subject to evaporative disagreements, for a reason not characteristic of assertoric judgements in general.

This reason can now be described as follows. Probabilistic judgement is a species of intrinsically ‘non-final’ judgement. Probability has its origins in situations in which we feel that final commitment is unwarranted. In this context not only is judgement sensitive to evidence, but

there is no warrant for ruling out further evidence. Suppose for example that we are interested in whether P, and that the context is such that we make judgements about the probability that P. These judgements are based on our evidence about P – and we would not make such probabilistic judgments if we thought that the evidence was sufficient to warrant a final judgement that P (or that not-P). In other words, we hold open the possibility that further evidence would make a difference to our views as to whether P. But this means that whatever judgement we do make must be non-final – we must hold open the possibility of revising it, if new evidence comes to hand. So the non-finality of judgement about probability stems from its origins in the expression of non-final judgements about other topics.

It is illuminating to contrast this explanation to a characteristic argument for non-factualism about probability, that we encountered in chapter 4. That argument focused on the same feature of probabilistic judgement – its origins in the expression of qualified judgement about other matters – and hence observed that probabilistic assertions typically express ‘partial beliefs’, or degrees of confidence. The non-factualist wanted to infer from this that such assertions do not express full beliefs, and hence that probabilistic utterances are not statements of fact. However, we saw that a factualist need not accept this inference; for the non-factualist has not shown that the same mental state could not be both a partial belief that P and a full belief that it is probable that P. On the contrary, the factualist might suggest, the notion of a degree of confidence, functionally construed, provides a characterization of the content of a belief about probability.

The present argument does not depend on this questionable step. It is not intended to show that probabilistic statements are not factual, but rather to explain the fact that their usage differs in a significant respect from that of typical non-probabilistic statements. The argument does not take this difference to prove that probabilistic utterances are not ‘genuinely’ assertoric, descriptive, or what-ever; for it belongs with a point of view that rejects such underlying distinctions. It simply attempts to explain, in terms of the most evident function of probability in discourse, a subtle but important aspect of the application of the notions of truth and falsity to probabilistic judgements.

It may seem surprising that a disagreement about the very same sort of judgement can be sometimes substantial and sometimes insubstantial. There is a simple explanation, however. In the probability case, for example, I said that disagreements evaporate in cases in which it turns out that agents are initially relying on different pieces of evidence. Cases like this are more common in some areas than in others. In scientific discussions, for example, it can usually be taken for granted that people share the same evidence, and moreover that new evidence will not come to hand, at least in the context of discussion. Probabilistic judgements can then be treated as final, and evaporation will be rare.

Even in more ordinary contexts, there seem to be good reasons for treating a probabilistic dispute as substantial, until it proves not to be. For one thing, it helps to ensure that the important disputes – the cases in which somebody has made a mistake – don’t slip through the net. It is better to risk prosecuting a few innocent differences than to risk excusing a guilty one. Secondly, we noted that if the grounds for an insubstantial disagreement are that speakers have initially relied on different evidence, then those grounds are effectively cancelled by the communication that calls attention to the disagreement. Communication puts the SBP into effect – if two people can disagree, then they can compare and combine their evidence. In order that they have an incentive to do so, it seems important that any current disagreement be treated as substantial. Only if such a disagreement is treated as if there is a single correct (or ‘true’) answer, will participants be encouraged to reassess their positions, to re-examine the latest evidence, and hence to strive for agreement.

Thus in the probability case it seems that only non-current disagreements should be allowed to evaporate. (Not all non-current disagreements, of course, but only those that turn out to have stemmed from access to different bodies of evidence.) The basis for such disagreements is automatically cancelled by communication. It is natural to wonder whether this is a peculiarity of the probability case; or whether, on the other hand, all current disagreements should be regarded as substantial. To settle the question, we shall need to examine some non-probabilistic sources of evaporative disagreement. That will be the main task of chapter 8. Before turning to non-probabilistic cases, however, I want to deal with the objection that a factualist can explain the existence of evaporative disagreements about probabilities by appealing to indexicality – by allowing that probabilistic judgements make implicit indexical reference to the evidence on which they are based.

## 7.8 THE INDEXICAL OBJECTION

An evident feature of indexical judgement is that apparent disagreements may turn out to stem from a misunderstanding as to the intended reference of an indexical term. For example, the judgements ‘It is cold here’ and ‘It is warm here’ are incompatible, in the sense that no one could correctly assent to both at the same time; but obviously both may be correct if said by speakers at different places or times. In chapter 5 we saw that a probabilistic factualist has other reasons to think of ‘It is probable that P’ along the lines of ‘*Given the present evidence*, it is probable that P.’ Does not such a reading, coupled with the above feature of indexical judgement in general, explain what happens when two speakers base judgements about the probability of the same state of affairs on different bodies of evidence?

There is an important difference between the two cases, however. Imagine for example that you phone me and say ‘It is snowing here today.’ I will disagree with you if and only if I think it is not snowing where you are; and agree if and only if I think that it is snowing there. But in a probabilistic case I may disagree with you even if I think your evidence does support your assessment of the probability at issue; and agree with you even if I think that it does not. In other words, a disagreement may exist in the probability case, even if both parties acknowledge that neither has incorrectly assessed his or her own evidence. There is no analogous possibility in the usual indexical case.

The objector will reply that in the probability case this possibility depends on a shift in the reference of the indexical term, in the course of a conversation. It is simply the possibility that, for example, I might agree that given *your evidence* it is probable that P; and yet disagree that given *the presently available evidence* it is probable that P. Clearly such changes of reference are likely to occur in conversation. At least in principle, conversation allows us access to each other’s information on the topic at hand. We may begin with different bodies of ‘available evidence’, but conversation ensures that what is now available to each of us is our combined evidence.

But can this account for the facts? The impression that it does so seems to me to trade on the vagueness of the notion of ‘available evidence’. There are various readings of this phrase, ranging from the completely subjective – ‘The evidence of which I am actually aware’ – to the fairly objective – ‘The evidence available in principle.’ At the subjective end of the scale, we certainly explain how apparent disagreements about probability can evaporate; but at the cost of making it mysterious why they should be treated as disagreements in the first place. Take a forensic example: you say that your evidence points to organized criminals with contacts in high places; I think that mine suggests a clever twelve-year-old with a personal computer. My initial reaction is not to think that we disagree but to assume that we must be relying on different evidence. Moreover, since there is no evident disagreement between us there is no apparent motive for us

to take matters any further. What needs to be explained is the fact that in a case like this we should and do go on to consider our combined evidence. On my account the motive is provided by our treatment of such a case as a substantial disagreement, answerable to a normative notion of correctness. (This is the *internal* motive, of course; the external justification comes separately. More on the latter below.) But on this subjective version of the indexical interpretation, the sanitized exchange provides no motive for anything more than polite acknowledgement of different but thoroughly compatible judgements.

So much for a subjective interpretation of the supposed reference to ‘available evidence’. What if this phrase is read as something less unambiguously subjective – for example, as something like ‘the evidence reasonably accessible’? The problem is now a dilemma. On the one hand, if this reading is taken sufficiently strongly to exclude the possibility that two parties to a conversation have different ‘available evidence’, then it excludes the very cases the indexical treatment is supposed to explain – those that seem to allow evaporative disagreements about probability. If on the other hand the reading allows for differences between speakers in their ‘available evidence’, then the above problem recurs. The supposed disagreement doesn’t get started, because speakers who understand probabilistic assertions in these terms will be inclined simply to attribute what would otherwise be a conflict to differences in the evidence that each refers to.

Thus it seems to me that an indexical reading cannot explain this distinctive feature of probabilistic judgement. The impression that it does so rests on a systematic equivocation concerning the supposed indexical reference – in effect, on a vacillation between the horns of the above dilemma.

## 7.9 EVAPORATION AND THE PROPERTY OF TRUTH

At this point a determined opponent, though conceding that the indexical interpretation does not provide an orthodox explanation of evaporation, might hold out the possibility of some other such explanation – some other way of accounting for evaporative disagreements within an orthodox truth-conditional interpretation of probabilistic judgement. I want to finish with an argument that seems to count against this possibility.

The analytic view thinks of truth and falsity as properties; the goal of analysis is an understanding of the nature of such properties. Suppose that P is a sentence (or statement, proposition, or whatever) of the kind that bears these properties. ‘P is true’ is then a respectable assertoric sentence in its own right – it can be used to say that a certain linguistic entity has a certain property. Moreover, since it is not probabilistic it cannot be the subject of the kind of evaporative dispute described above.

Imagine now that P itself is probabilistic. The analytic view faces a dilemma. On the one hand we have just observed that if two speakers disagree about a simple predication such as ‘P is true’ then one of them is mistaken, whatever their evidence – one of them has spoken falsely. Yet on the other hand to disagree about ‘P is true’ is to disagree about P. If the former disagreement entails that one party is mistaken, then so does the latter. If we allow evaporative disagreements about P, we must allow them about ‘P is true.’

If truth and falsity are properties, in other words, their literal ascription must be confined to sentences not subject to the ‘no fault’ evaporative disagreements characteristic of probabilistic judgements. It follows that our ordinary treatment of probabilistic judgements is not entirely compatible with the truth-conditional model. From the orthodox point of view, either we misuse

such judgements, or they do not possess genuine truth conditions; for their use is incompatible with their having the property for possession of which truth conditions are conditions.

This point has been obscured, I think, by the assumption that the features of use on which it rests can be explained by indexicality – by some sensitivity of probabilistic truth to evidential context. However, we have seen that the appeal to indexicality itself rests on an equivocation in the relevant notion of ‘available evidence’. In any case, the result is an embarrassment for the analytic approach. In one way or another, it detaches the analyst’s desired ascription of truth conditions from its evidential basis in usage. It implies that we either use truth-bearing sentences as if they were not truth-bearing, or use non-truth-bearing sentences as if they were truth-bearing.

Note that the present approach encounters no such problem. An explanatory view is not committed to the idea that truth is a property, seeking merely to explain the fact that ordinary speakers are inclined to treat it as such. If it turns out that ordinary usage is sometimes incompatible with a treatment of truth as a property, then so much to the credit of the explanatory view – at least so long as it can account for this fact.

I have argued that in the probabilistic case the appropriate account turns on the role of probabilistic judgement in contexts of incomplete evidence. In chapter 8 I want to show that by appealing to different factors in the same general spirit, it is possible to explain the peculiarities of certain other areas of discourse – areas which, like probabilistic usage, have been popular targets for the non-factualist approach.

## 8

**The Limits of Intolerance**

We are trying to account for the ordinary use of the notions of truth and falsity in terms of their function in encouraging argument. In view of our original concern with non-factualism it is important that the account should work ‘at the edges’ – in the borderline cases in which talk of truth and falsity seems to some degree inappropriate, and in which non-factualism has hence seemed particularly attractive. We looked at one such case in chapter 7. I argued that our occasional reluctance to apply the notions of truth and falsity to probabilistic judgements can be correlated with a systematic failure of the normal functional justification for the use of truth and falsity.

In this chapter I want to examine some further cases, and show that again the functional theory of truth can explain the kinds of peculiarities that have been taken to favour non-factualist treatments of the topics concerned. I’ll pay most attention to moral and conditional judgements (the topics whose non-factualist treatment we considered in part I). But I shall also comment on the account’s potential in other areas in which non-factualism has been appealing. And I’ll finish, as promised, with some remarks on the non-indicatives.

Perhaps I should say again that in attempting to explain the appeal of non-factualism, I won’t be trying to make a case for non-factualism as normally construed. Non-factualism normally takes for granted the analytic approach. It sees itself as concerned to determine the true extent of the property that distinguishes fact-stating discourse. Part I discredited that endeavour. As we’ll see, however, there is considerable scope for reconstruction of non-factualist arguments along explanatory lines. A subsidiary concern of this chapter will be to distinguish several stances that remain open in this connection, when the analytic project has been closed down.

## 8.1 RELATIVISM AND EVAPORATIVE ARGUMENT

In chapter 7 we drew a rough distinction between two sorts of linguistic dispute, according to the root of an initial disagreement: in an *evidential* dispute the root lies in speakers’ evidential beliefs; in an *inferential* dispute it lies in the rules or habits of inference that lead from evidential beliefs to the conflicting judgements that are the immediate grounds for disagreement. We noted, however, that disagreements can have multiple roots; and the distinction between evidential beliefs and rules of inference is in any case tendentious.

At that stage we had not raised the possibility of evaporative disagreements. However, the one sort of evaporative disagreement we have so far considered is clearly evidential rather than inferential. It turns on the fact that in making probabilistic judgements, speakers may have access to different bodies of evidence. So long as the speakers concerned do not disagree with each other’s evidential beliefs or probabilistic inference rules, we have an evaporative evidential disagreement.

Can there be evaporative inferential disagreements? I think that there can, and indeed that some of the best-known tenets of the non-factualist movement amount to the endorsement of this possibility. For consider moral or aesthetic relativism – the view that different individuals or groups may hold to incompatible but equally ‘valid’ aesthetic or moral principles. Such a possibility would reveal itself in a case in which the two groups or individuals make conflicting ethical or aesthetic judgements, on the basis of agreed non-moral or non-aesthetic facts. Since

the non-evaluative evidence is agreed, the disagreement must stem from the general evaluative principles, in virtue of which each group or individual infers moral or aesthetic conclusions from such a basis. In other words, in our present terminology, the disagreement has an inferential basis. The relativist claims that in such a case, neither side need have grounds for regarding the other as mistaken. But in our terms, such no fault disagreement is the mark of insubstantiality; and hence the relativist seems to be urging the possibility of evaporative inferential disagreements.

This case bears out some points I made earlier. It illustrates both the element of reconstruction of non-factualism that follows the rejection of the analytic viewpoint, and the fact that the reconstruction leaves open several distinct kinds of non-factualist conclusion. In appealing to the possibility of relativism, non-factualists in ethics and aesthetics have on the whole been concerned to argue that moral and aesthetic truth *is* relative – that it is relative, as a matter of fact, and therefore should be treated as relative. It follows from this view that if a community treats moral truths as absolute, its members are mistaken about the nature of moral judgement. In other words, the non-factualist traditionally begins with claims about the ‘real’ status of moral judgements, and proceeds to prescriptions concerning the use of moral utterances. In rejecting the analytic point of view, we seem bound to reject this form of non-factualism. We have given up the idea that there is an independent standpoint, with respect to which there is anything more to being a statement of fact than being treated as such. We may now describe usage and hence explain it, but there is no longer a place for trying to analyse usage and hence prescribe it.

I think that a good deal of contemporary non-factualism can be reconstrued in this descriptive sense, admittedly at the cost of systematic disrespect for the intentions of its authors. The trick is simply to reinterpret what the analytic non-factualist normally appeals to by way of evidence – such things as our ordinary intuitions about the truth and falsity of the kinds of judgement in question. From the non-analytic point of view this appears not as evidence for the presence or absence of some underlying property, but rather as the basis of a descriptive taxonomy of use. Much of orthodox non-factualism is thus directly accessible to our first kind of reconstructed non-factualist, whose role is to catalogue the subtle variations of usage on the boundaries of the statement form.

Perhaps this kind of non-factualism need not be merely descriptive. It cannot prescribe ordinary usage as such, but it might seek to correct any corruption of usage that could be attributed to the influence of a mistaken philosophical theory. All the same, to a traditional non-factualist this looks like a dull substitute for prescription grounded on analysis. Is there room for anything more colourful – for a non-analytic non-factualism that would seek to revise ordinary usage itself? I think there are two possibilities, stemming from two kinds of reflection on the use of the class of judgements in question.

## 8.2 SECOND ORDER EXPLANATORY NON-FACTUALISM

The first possibility would involve the recognition that actual practices of disagreement simply fail to resolve certain sorts of dispute. In the moral case, for example, it would flow from the perception that moral disagreements may simply reach a stalemate, in which neither party can do more than to continue to affirm some disputed moral principle. This perception needn't itself recommend relativism. But the inherently unstable nature of such a stalemate, in which each party retains the conviction that the other is at fault, might seem to recommend a change in usage – a degree of relativistic tolerance.

The second possibility is perhaps only a more theoretical version of the first. It would concentrate not so much on the actual procedures we employ in argument as on the role of a type of discourse in our lives, and that of argument in general. From such a standpoint it might be argued that particular types of utterance would be better treated in one way rather than the other – ‘better’ being cashed in terms of the overall value to the individual, community or species of such forms of behaviour. Given the present functional account of truth, for example, it might be argued that we would do better to allow for evaporative moral and aesthetic disagreements.

Thus there seems to be room for a non-descriptive non-factualism – a second order theory whose prescriptions concerning usage would flow from a view of the significance of factual usage in general, and an understanding of the role of the particular area of discourse whose status is in doubt. I shall not directly attempt to decide to what extent existing arguments for or against relativism about morality and other topics can be interpreted in this light. But I do want to illustrate the range of considerations that might be relevant to such a second order non-factualism; and some of these, as we shall see, call to mind standard moves in the discussions of relativism about morality and other matters.

Let us begin at one extreme, with a challenge to the conceptual possibility of the second order stance. In the moral case, for example, the stance seems to depend on the claim that it is possible to discuss the merits of moral principles – of different rules of inference from non-moral beliefs to moral conclusions. This seems to depend in turn on an understanding of the meaning of moral judgements which would be prior to the adoption of a particular set of acceptance rules for such judgements. After all, these acceptance rules are what the discussion is about. Now the objection: such an understanding is impossible, because what we mean by moral terms is a function, *inter alia*, of what we take to be their assertion conditions. In a slogan, meaning *is* use, and therefore cannot prescribe use.

In effect, this is the objection that relativism collapses into the indisputable but uninteresting possibility that different speakers may simply use the same word with different meanings. The proper reply is obviously that some uses are more important than others. In the moral and aesthetic cases, at any rate, the feature of use that does most to fix meaning seems to be not the adoption rules in virtue of which speakers acquire commitments, but the ‘application rules’ – the conventions in virtue of which commitments have behavioural consequences. It is relatively easy to imagine someone who adopts moral and aesthetic commitments in different circumstances from ourselves. But as we have already had cause to observe, it is difficult to make sense of someone who means what we mean by ‘good’, ‘beautiful’, and so on, and yet who is not motivated much as we ourselves are by commitments involving these terms. There are certain exceptional cases, of course – for example cases in which the effect of such a judgement is overridden by some other motivation – but we would be inclined to regard the fact that someone lacked that pull *in general* as evidence that he meant something different from us by the terms concerned.

Conversely, it seems that these definite behavioural consequences prevent arguments over matters of moral and aesthetic principle from degenerating into terminological border disputes. If a person’s moral and aesthetic commitments – those of her commitments that are expressed in what are for us moral and aesthetic terms – have for her the same kind of behavioural consequences as such commitments would have for us, we take this as evidence that she means the same by these terms as we do. We are then in a position to compare her moral and aesthetic sensibilities to our own; and free to conclude, with the relativist, that although such a person’s sensibilities differ from our own, she need not be *mistaken*.

To show that relativism is a conceptual possibility is not to show that we are or should be relativists, however. The probabilistic case illustrates this point. As we observed in chapter 4, simple probabilistic judgements also have fairly definite consequences for behaviour. Roughly, judgements of increasing probability are reflected in an increasing willingness to risk a given loss for a given gain, in an action that depends for success on the state of affairs whose probability is in question. If someone consistently failed to act in this way, having sincerely professed to what appeared to be probabilistic commitments, we would be inclined to say that he did not mean what we mean by ‘probable’ and associated terms. Conversely, if someone does apply what appear to be probabilistic judgements in this way, then we are inclined to assume that she does mean what we mean. And we continue to assume it, even if we disagree with the statistical principles on the basis of which she makes her judgements. Probabilistic relativism is thus a conceptual possibility. However, it seems to have little of the attraction of its moral and aesthetic counterparts. Once evidential differences have been sorted out, we don’t say that *chance* lies in the eye of the beholder.

If we were to try to justify this difference between the probabilistic and ethical cases, I think it would have to be in terms of the claim that barring evidential differences, probabilistic judgements involve no systematic failure of the SBP. Presumably our ability to make probabilistic judgements has some survival value, and presumably its value would vary if we varied our probabilistic inference rules. (It is easy to think of cases in which the wrong rules would regularly lead to disaster.) Access to evidence aside, there seems to be no difference between speakers that could explain how the optimal probabilistic inference rules could be different in different cases. And in the absence of such a difference, there may seem to be no case for probabilistic relativism.

However, this is to overlook the possibility that there might simply be no optimal set of probabilistic inference rules. Any such set of rules is a strategy for making judgements on the basis of less than conclusive information. In principle there might be a range of such strategies, none clearly better than any other. Probabilistic relativism then begins to look appealing, I think, on inferential as well as evidential grounds. This suggests that the apparent lack of evaporative inferential disagreements in ordinary probabilistic usage may simply reflect the fact that ordinary probabilistic reasoning is not complex enough to have encountered the difficult cases. (If so then we might do better to look at the reasoning of professional statisticians. I am not sure to what extent differences between statistical schools of thought can be viewed in these terms.)

### 8.3 RELATIVISTIC PRACTICE: FOR AND AGAINST

The point remains that in the moral and aesthetic cases, as for probability, relativism is neither entailed nor recommended simply by the fact that it is a conceptual possibility. One further factor is whether moral and aesthetic judgements have the SBP. The relativist’s traditional appeal to subjectivity can be seen as an argument that the property does not hold in these cases (and hence, in our terms, that moral and aesthetic disagreements are properly treated as potentially evaporative). The subjectivist claims that moral and aesthetic utterances are expressions of desires, motivations, or something of the kind. Ideally such attitudes should be correlated with dispositions to feel pleasure or displeasure in certain states of affairs. (Masochism is not a particularly good survival strategy.) Hence if different speakers can have different such dispositions, the SBP need not apply: if pleasure lies in the eye of the beholder, then so (we should allow) does beauty.

This sort of argument can be challenged at a number of points, however. In particular, it seems to depend on the view that appetites or motivations are simply givens, and not themselves

contingent products of cognitive processes. If motivations are allowed to be changeable, then there are at least two lines of argument in favour of non-relativistic moral and aesthetic practices.

The first argument would rest on the claim that some practices are in some sense ‘better’ than others, and hence that there is a place for the kind of improvement that is encouraged by the notion that expressions of motivation are answerable to an external standard. ‘Better’ might be cashed in several ways. Perhaps most plausibly, it might be argued that some tastes and appetites are potentially more rewarding than others. Hence there would seem to be an individual advantage in a social process of moral and aesthetic discussion, argument and refinement. The more educated one’s palate, the greater one’s enjoyment, on this view – and hence we all gain, on the whole, by the process of comparison, criticism and improvement that is encouraged by the practice of treating moral and aesthetic judgements as true and false.

This argument seems more plausible for the aesthetic than the moral case. For the moral case, however, the recognition that motivations are changeable states suggests another line of argument. For it suggests that we regard moral judgements not as expressions but as sources of motivational states. Roughly, we tend towards what we take to be good not because ‘taking to be good’ is a manifestation of a pre-existing tendency but because ‘good’ functions as a conventional reward for action. Clearly, this is a view of moral discourse that parallels the present account of truth and falsity. It is a likely development of the theme, familiar in ethical theory, that moral concepts play a regulative role in society. The view would thus be that morality provides a system of incentives that functions to encourage socially ‘desirable’ forms of behaviour (‘desirable’ being cashed in non-moral terms, of course).

However, such a system of incentives would be useless if everyone was held to be entitled to a reward, in virtue of satisfying his or her own individual standards. Relativism thus threatens the social function of morality – which perhaps explains why it is such a disturbing thesis, at least in extreme forms. The practice requires that the system of rewards be seen as objective, as answerable to an external standard; or in other words, that speakers regard differences in their application of moral descriptions as indications of mistakes.

Thus we have two possible arguments for disallowing evaporative inferential disagreements on moral and aesthetic matters. The first is based on the claimed advantages of educated aesthetic and (less plausibly) moral tastes. The second turns on the regulative role of moral and (less plausibly) aesthetic judgement. Both, I think, leave scope for limited relativism. The advantages of conformity may sometimes be outweighed by the disadvantages of intolerance. ‘Live and let live’ may sometimes be the optimal survival strategy. Perhaps we should behave as if everyone in our own community is in the same boat, but allow that foreigners do things differently.

I do not propose to examine the above justifications for moral and aesthetic objectivity in any detail. Our present purpose is simply to illustrate the variety of factors that, from the non-analytic perspective, may be held to explain or recommend a particular pattern of usage. However, the arguments do call for a number of comments. The first of these concerns the parallel between the latter approach to morality and our suggested account of truth.

#### 8.4 TRUTH AS INCENTIVE – A CIRCULARITY?

We described ‘true’ and ‘false’ as incentives, whose purpose is to encourage speakers to seek agreement, and hence to participate in linguistic disputes. The second argument above suggests that in the moral case, such a system of incentives requires that speakers treat moral truth as objective – in other words, that they treat moral disagreements as substantial. This seems to

imply that in the case of truth and falsity themselves, the claimed system of incentives would itself require that speakers treat disagreements about the truth of utterances as substantial. If so, isn't the account in trouble? Doesn't it commit us to an endless hierarchy of disagreements, corresponding to iterations of the predicates '... is true' and '... is false'?

This is a nice point, to which I want to return in the next chapter. The brief answer is that although it is true that if 'true' and 'false' are to function as incentives then speakers must treat disputes about the truth of utterances as substantial, this is guaranteed, as long as it needs to be, by the general presumption that disagreements are substantial unless proved otherwise. Evaporative disagreements are the exceptions. There is a kind of circularity in the use of 'true' and 'false' as incentives, but it is not vicious. In order that 'true' and 'false' should function as incentives, their ascription takes the objective or declarative mode. Truth is presented as an objective property, that utterances either have or lack. We add truth to our idea of the world, to help us get on in it. But without truth, we would not have an objective or declarative mode of speech. What is distinctive about this mode of speech is that it is regarded as answerable to an external standard; that is, to truth.

### 8.5 PROJECTIVISM NOT VINDICATED

The second point arising from section 8.3 concerns an apparent contrast between the above case for a substantial treatment of moral disagreements and the kind of justification previously suggested – a difference that seems to echo the old idea that beliefs and desires have different 'directions of fit' to the world. In the previous non-evaluative cases (including the probabilistic case) we used the SBP to explain and justify the use of the normative notions of truth and falsity. We argued that because the behavioural consequences of such commitments are broadly similar across a linguistic community, there tends to be an individual advantage in a conventional system of incentives and disincentives that encourages attempts to reach agreement. The world puts the members of a community in the same boat, as it were, and hence we tend to do better if we co-operate. In the moral case, however, the above argument may seem to reverse things – to suggest that we should behave as if an external moral reality had put us all in the same boat, in order that morality might serve its community purpose.

This is an important objection. If sustained, it would suggest that the present account is a form of the projectivist theory we discussed in chapter 4. For it would seem to admit a distinction between those areas of discourse for which truth can be cashed in terms of a world-imposed notion of correctness, and those in which there is a useful function for a pretence of world-imposed correctness.

However, I think the objection is mistaken. It is correct in noting a difference, on the proposed account, between the grounds for the application of the notions of truth and falsity in different areas of discourse (at least if their application to normative discourse is explained by the latter of the two arguments above). Indeed, I think the explanatory approach should explicitly affirm this possibility. Once we detach the explanation of the use of truth from the search for an analytic base, we should see it as possible, even probable, that across the range of functions served by the assertoric form of language, the advantages of truth-encouraged argument will not be the same in all cases. Certainly, we should expect that these cases have something in common, to explain the fact that they all employ the assertoric form and the notion of truth. But this degree of uniformity is compatible with considerable diversity in its explanatory base. The advantages of argument may vary from case to case.

The objection is mistaken, however, in so far as it assumes that in the non-evaluative cases (or some sub-class of these cases), the account provides a ‘genuine’ notion of truth, cashed in terms of ‘world-imposed correctness’. I think this mistake stems from a tendency to construe the SBP in terms of the ‘success’ account of truth. From such a point of view it seems obvious what property is shared by the same belief in different heads: that of leading to behavioural success in the same circumstances – that is, on this view, the property of being true in the same circumstances. However, in introducing the SBP I emphasized that it should not be conceived in these terms. In order to account for an evaluative notion whose function seems to be to encourage (reasoned) agreement, we needed to claim that disagreement is in general an indication of some sort of behavioural disability. That in turn seemed to require the assumption of the SBP. Without something of the kind, we could well have disagreements between speakers in cases in which each holds what from his or her own point of view is the behaviourally optimal belief. I noted that this can indeed happen in individual cases, such as when the effects of a mistaken belief are balanced by a mistake elsewhere in an agent’s system of belief. But to account for truth in terms of its propensity to encourage agreement, we need to assume that it does not happen globally.

The SBP was thus a theoretical hypothesis, needed at this point in our explanatory account of truth. As such, it derives support from the demonstration that departures from the standard use of truth can plausibly be correlated with specific reasons to favour (or at least not to penalize) disagreement, in special cases – with systematic factors in virtue of which the SBP might be expected to fail.

This route to the SBP is non-committal as to the nature and origins of notion of ‘appropriateness’ on which it depends. As just noted, there is therefore an understandable temptation to think of it in terms of truth, ‘correct representation of reality’, or something of the kind. This temptation must be resisted, however, for these are the very concepts whose use we are trying to explain. Behavioural appropriateness belongs to an explanatory but not an analytic base for the suggested account of truth. The key to the explanatory approach is the realization that these things can be kept apart.

Provided we do keep them apart, the above objection can be seen to be mistaken. Certainly there may be a difference between the senses in which the ‘appropriateness’ of evaluative and non-evaluative judgements depends on more than individual preference or point of view. But this does not entail that in the latter case ‘appropriate’ can be equated with ‘true’. The difference lies within the scope of the explanatory approach as a whole. It does not, as for the projectivist, mark a point at which explanation takes over from analysis.

## 8.6 PRESCRIBING USAGE

The final point arising from section 8.3 concerns the sense in which the arguments for and against aesthetic and moral relativism can be considered to be ‘prescriptive’. It needs to be emphasized that their conclusions are ‘external’ rather than ‘internal’ prescriptions for moral and aesthetic usage. That is, they arise not from *within* a particular moral or aesthetic practice, but from an external standpoint – a point of reflection on the practice as a whole.

This distinction is not always easy to draw, but I think it has a place in many areas of discourse. Many practices yield prescriptive principles – principles that will seem *to the practitioner* to follow from the ‘ground rules’ of the practice in question. But these are internal prescriptions. They flow from the practice, and we cannot feel their force unless we already subscribe to the practice from which they stem.

External prescriptions, on the other hand, depend on an ability to detach ourselves from the practice in question, and hence to make recommendations about its conduct from some wider standpoint. This exercise is problematic in a number of ways. Even if we can achieve the required external perspective – survey usage from outside, as it were – there are difficulties with the notion of a recommendation from such a perspective. Here, as in other areas of the human and social sciences, the boundary between justification and explanation is far from precise. Naturalistic explanation tends to be presented as conservative defence.

Whether conservative or reformist, however, similar questions arise. In what sense is the recommended practice better than others? And how is it possible for us to adopt such a recommendation – to decide on a particular pattern of usage? These are difficult issues, and I have no definite answers to offer. But it does seem that we are not entirely constrained by current convention. In philosophy and in ordinary life we do have some ability to reflect on and to modify these kinds of patterns of linguistic practice; even if the process of change seems to involve not so much an explicit decision on current and possible usage as a communal shuffle in a direction we cannot quite describe.

I have suggested that some of the concerns of the non-factualist movement can be viewed in this light, as attempts to prescribe the appropriate pattern of usage for particular classes of utterances – to show, in particular, that the application of the notions of truth and falsity is sometimes inappropriate. I noted that to interpret non-factualism in this way is to separate it from its usual analytic presuppositions. However, we saw that prescriptive arguments of the above kind have a natural place in the analytic view. So it may be that we should regard the prescriptive side of non-factualism as merely a legacy of its usual analytic origins, and not try to reconstruct it on an explanatory basis.

My impression is that this is unduly pessimistic, and that there has been an aspect of non-factualism, not particularly tied to the analytic view, involving a potentially reformist process of reflection on usage. The point is incidental, however. Our main concern is with the explanatory view itself. In the remainder of this chapter I want to consider some further applications. The first of these involves the third of the main examples we discussed in part I – the case of conditionals.

## 8.7 CONDITIONALS

As we saw in chapter 5, the difficulties for a straightforward factualist interpretation of indicative conditionals seem to stem from the behaviour of conditionals with false antecedents. The suggestion that the indicative conditional is simply the truth-functional material conditional comes to grief on the fact that English speakers do not take grounds for denying an antecedent as grounds for asserting a conditional in which it appears. This fact underpins Adams's hypothesis: the claim that the assertibility of an indicative conditional follows the subjective conditional probability of the consequent, given the antecedent. And Adams's hypothesis, in turn, underpins the anti-factualist arguments we discussed in chapter 5. Even if, as I claimed, these arguments are not conclusive, it is clear that the behaviour of conditionals with false antecedents makes things very much more difficult for a factualist interpretation.

From the explanatory viewpoint, however, there is a ready account of this behaviour. Suppose that the main function of the commitments we express with indicative conditionals lies in hypothetical reasoning. Crudely, the judgement that if P then Q is a kind of insurance, or contingency plan, against the possibility that P. (This seems plausible enough, but the following explanation could be modified to cope with other possibilities.) Presumably some contingency

plans are better than others. Add the communal element – the SBP – and we have what it takes to make the argumentative quest for agreement a useful social practice. Individuals then stand to benefit, improving their contingency plans by comparing them to those of others. And hence there is a role for truth, the dialectical catalyst whose function is to provide the (internal) point of argument – to make disagreement seem to matter.

What then of false antecedents? How should we assess an insurance policy against an eventuality that we know has not arisen, and will not arise? That is, what bearing should it have on our assessment of a plan to deal with the contingency that P, that we learn that not-P? From the explanatory point of view, the suggestion that a false antecedent makes a conditional true (or rather provides a reason for calling it ‘true’) amounts, in effect, to the claim that any plan against an unrealized possibility is worthy of approbation.

I think that this convention would have several disadvantages. The best that can be said for it is that any such plan is bound to be harmless, in the immediate sense, because it will never be put into effect. But this is not to say that we should encourage each other to carry such policies. For one thing, even a redundant policy bears a cost. To encourage redundant plans would be to encourage unnecessary use of limited cognitive resources. For another thing, to encourage such redundant conditional commitments would be to devalue the incentives and disincentives on offer in the important cases – those in which such a commitment is not known to be redundant.

Most importantly, however, to reward conditional judgements on the grounds of their redundancy would be to undermine the function of such rewards in influencing *habits* of judgement. We noted that argument tends to improve not merely the commitments that form the immediate grounds for a disagreement, but also the general principles of inference that lead to such commitments. In the conditional case, the latter form of improvement seems to require that speakers not escape criticism simply in virtue of the fact that an inappropriately adopted conditional commitment turns out to be redundant (due to a false antecedent). Otherwise it would be as if we were to excuse (or worse still, commend) a habitual drunken driver, on the grounds that on the occasion in question he or she had been lucky enough to avoid an accident.

A similar range of considerations seems to explain why we should not regard conditional judgements with false antecedents as uniformly false. The first factor is different. The problem would be not that we would waste cognitive space by encouraging redundant commitments, but that we would underutilize the available resources by discouraging potentially useful commitments. (Why take on a life insurance policy that in addition to the ordinary premium imposes a heavy penalty for failing to die?) But the second and third factors are the same: this practice would devalue the incentives and disincentives in the more significant cases; and most importantly, would interfere with the optimal use of truth and falsity in encouraging the best habits of conditional judgement.

This seems to explain why (*pace* the material conditional interpretation) we do not argue about the truth of indicative conditionals whose antecedents we take to be false. Conversely, we can also explain why we do disagree about pairs of conditionals that on the material conditional view might both be true. We are ordinarily inclined to treat ‘If P then Q’ and ‘If P then not-Q’ as incompatible, rather than as compatible claims that together imply that not-P (as we should if we were dealing with material conditionals). The contingency plan model of conditional judgement explains why this should be so. To take on both commitments would be to subscribe to a pair of plans that could not both be put into effect. Faced with the competing claims we therefore cannot accept both, unless in the sense that we affirm that neither will be put into effect. The disadvantage of the material conditional is that it makes this compromise the natural choice. Like

the indexical treatment of probabilistic judgement, it defuses potentially valuable disputes. The appropriate course is therefore to treat the co-occurrence in a conversation of ‘If P then Q’ and ‘If P then not-Q’ as the basis of a substantial disagreement, and hence as *prima facie* grounds for a dispute. In this way we are provided with an incentive to call in our evidence, and to attempt to justify our conditional commitments.

In the process we may find that the initial disagreement rested on the fact that the two sides relied on different evidence, and that the combined evidence shows that not-P. We will thus be led to the conclusion that the material conditional interpretation would have made immediate. But we are led to it at the proper time. Properly conducted contingency planning may lead to the conclusion that a particular contingency will not arise, on the grounds that it would entail inconsistent consequences. That is, it may lead to the realization that such planning is inappropriate in the case in question. But this is a far cry from discounting a possibility whenever we find ourselves faced with incompatible plans for dealing with it.

Thus, it seems possible to explain in functional terms the most striking differences between the ordinary use of truth and falsity with respect to indicative conditionals and the pattern that would be appropriate for material conditionals. Of course, there are other factors that a more detailed treatment of indicative conditionals would need to take into account. One such concerns the connection between indicative and subjunctive conditionals. The focus of a disagreement about an indicative conditional will often shift to a connected subjunctive, particularly if doubt is raised about the truth of the antecedent. Suppose for example that we are arguing about whether if Rudolph is here, he is wearing a wig. If we learn that Rudolph is not here, we are inclined to go on to argue about whether if he had been here, he would have been wearing a wig. So a proper treatment along functional lines will need to consider subjunctives, as well as indicatives.

However, if we allow that in such a case the function of the original argument is partly to improve our habits of conditional judgement, there is an appealing explanation of the move to the subjunctive. For even if we find out that the immediate judgement is redundant (because Rudolph is not here), the more general potential benefits provide a reason for a continued dispute. As long as we take it that there is still a substantial disagreement between us, we are liable to reassess the principles on which our original judgments were based; and hence, on the whole, to do better in the future. This suggests that subjunctive conditionals might themselves be explained as tokens in a process of reasoning and argument whose significance lies at another level: in the general principles on the basis of which we form the contingency plans with which we meet the future.

We saw above that indicative conditionals give rise to a species of evaporative disagreement, in cases in which the effect of combined evidence is to convince the parties to a dispute that the conditional in question has a false antecedent. Subjunctive conditionals seem to be subject to a different species. As has often been noted, our assessments of counterfactual conditionals depend on background assumptions. In order to reason about what we take not to be the case, we must rely on some but not all of what we do take to be the case. To appeal to everything we currently accept would trivialize the exercise, since in particular we could appeal to the fact that not-P in reasoning about what would be the case if P.

How then do we select the background information to which we appeal? However we do it, it is clearly sensitive to context. As a result, a disagreement about a subjunctive conditional may turn out to stem from the fact that the participants have ‘held fixed’ different parts of their common background evidence. Thus:

- (8.1) If the author of *Waverley* had been alive today, he would have been over two hundred years old (because Scott, the author of *Waverley*, was born in 1771);

and yet

- (8.2) If the author of *Waverley* were alive today, he or she would certainly be less than one hundred and thirty years old (because nobody lives longer than that).

Such cases suggest that a dispute about a subjunctive conditional can evaporate, if it turns out that the participants are taking for granted different aspects of an agreed body of background information. This behaviour seems consistent with the suggestion that the significance of subjunctive conditionals lies in the general beliefs on which they are based. Evaporation here seems to follow recognition that there is no disagreement about these underlying generalizations.

Subjunctives are a difficult case, however. No doubt there is much more to be said about the connections between on the one hand their functions in speech and thought, and on the other their usage in argument and as bearers of truth and falsity. I want to conclude these remarks by mentioning an advantage of the present explanatory approach that seems particularly prominent in the case of conditionals.

In recent discussions of indicative conditionals it has been noted that their interpretation is constrained by their evident affinities, in ordinary usage, with disjunctions on the one side and subjunctive conditionals on the other. (See particularly Jackson 1984, pp. 68–9; 1987, pp. 67–70.) These affinities are problematic for any approach that wants to draw a major semantic distinction between disjunctions and indicatives or between indicatives and subjunctives – for example, the view that disjunctions have ‘genuine’ truth conditions (given truth-functionally), whereas indicative and subjunctive conditionals are ‘projections’ of non-truth-conditional mental dispositions. Why should classes of utterances whose meanings are determined in such different ways behave so similarly? I want to emphasize that here, as in other areas of discourse, the explanatory approach offers a unitary approach. Hence it avoids these problems. At the same time, however, it has (or at any rate claims to have) the flexibility needed to account for the relevant range of variation we find in usage – variations in patterns of argument and in the application of the notions of truth and falsity.

## 8.8 PRESUPPOSITION

Conditionals have often been linked in philosophical discussion with assertions dependent on presuppositions – themselves a class of utterances whose truth-conditional has been thought to be problematic. Much of the discussion of the latter class concentrates on the question as to whether a sentence such as

- (8.3) The king of France is the author of *Waverley*

can be considered to have a truth value, given that it depends on a false presupposition. I want to comment briefly on this topic, partly as a further illustration of the explanatory approach to such issues, and partly to exhibit a contrast to the conditional case.

From the explanatory point of view the question as to whether 8.3 ‘really’ has a truth value is misconceived, of course. We can only ask whether and why such an utterance is or should be treated as having a truth value. This aside, it may seem that this case ought to be similar to that

of an indicative conditional. Isn't discovering that a presupposition is false similar in effect to discovering that an antecedent is false?

I think there is actually an important difference between these cases. The difference between 8.3 and

(8.4) If there is a king of France, then he is the author of *Waverley*

is that a commitment to 8.4, unlike to 8.3, involves no commitment to the claim that there is a king of France. Other things being equal, acceptance of 8.3 can thus be expected to be accompanied by behavioural effects not shared by acceptance of 8.4: namely, those effects that would be expected to flow from the belief that there is a king of France. These effects give a point to criticism on the grounds of the failure of a presupposition, not shared by the conditional case.

This is not to say that in virtue of having such effects 8.3 *is* false when there is no king of France. It is simply to say that in virtue of these effects, the practice of describing such an utterance as 'false' in such circumstances is likely to serve a useful function. Perhaps the same function could be performed in other ways. For example, we might have a separate mode of criticism, reserved for this kind of mistake. The intuition that some philosophers have had that failure of a presupposition renders an utterance neither true nor false is perhaps evidence for the existence in usage of some degree of separation. Certainly it suggests that it would be within our capabilities to sustain such a distinction. But this is not a concession to the analytic standpoint – the basis for an account of 'real' as opposed to 'constructed' truth. For why stop here? Why not discern a separate kind of truth for every subject matter and every critical nuance? We saw in part I that what one needs to worry about, in leaping aboard the projectivist bandwagon, is not the engine but the brakes. Getting it started is easy enough. The problem is to pull the thing up where the projectivist wants to stop.

The question arises as to whether our explanatory view is simply projectivism unconstrained – the idealism that we saw to be the projectivist's most threatening danger. I shall return to this in the concluding chapter. I want to turn now to some preliminary questions. How widespread are evaporative disputes? Does the explanatory approach allow us to draw a distinction between those utterances that admit such disputes and those that do not? And if so, is there anything to be said for the idea that the latter are the genuine statements of fact?

### 8.9 FACTUALITY: A MATTER OF DEGREE?

Evaporative disputes are a boundary phenomenon, a manifestation of the felt inappropriateness of a continued pretence that a particular disagreement is answerable to an external standard of correctness and incorrectness. The cases we have so far considered suggest that the phenomenon is a product of the local peculiarities of particular types of utterance. Hence we might expect that in a broad range of central cases, it would be unknown. This expectation might in turn suggest a characterization of the 'genuine' statements of fact: namely, that they are those indicative utterances not prone to evaporative dispute.

Interpreted correctly, this characterization would be unobjectionable. If there is a distinction in usage between those indicative utterances that do and those that do not give rise to evaporative disputes, then by all means let us mark it. But this unobjectionable interpretation is not the source of the suggestion's appeal. This lies rather in the misconceived intuition that the

suggested characterization gives us the long-sought analysis of statement of fact – a venerable confusion of taxonomy for analysis.

That aside, it seems to me that the assumption concerning usage that underlies this suggestion is significantly mistaken. There seems to be at least one potentially global source of evaporative disputes. Hence it is doubtful whether there is any class of indicative utterances conventionally immune from the patterns of usage that supposedly distinguish ‘non-genuine’ statements of fact. This global source is a form of indeterminacy of meaning. In section 8.2 I mentioned the possibility that apparent disagreements may turn out to rest on the fact that the speakers concerned are using the same term with different meanings. This possibility threatened to trivialize the issue about relativism. To defend a space for an interesting relativist position, we suggested that the motivational character of moral judgement provides a fixed component in the meaning of moral terms, around which the relativist issue can be built. But the problems raised by meaning variance are not exhausted by its abortive role in this case.

Meaning variance can affect disputes in several ways, some of them relatively uninteresting. Perhaps the dullest involves a term that is known to be ambiguous in ordinary usage, whose multiple meanings find their way to the heart of an argument – for example as to whether in the modern computerized hospital the doctors should be expected to grasp the rudiments of ‘terminal care’. Marginally more interesting is the case in which a dispute stems from the fact that one participant has simply misunderstood the meaning of a central term. This raises an issue as to whether (and if so, why) we should say, when the mistake becomes apparent, that this person’s original assertion was false. This issue seems to have something in common with the case of false presupposition. Indeed, it could be argued to be a sub-case of that of presupposition, on the grounds that assumptions about meaning can be treated as presuppositions. Alternatively – I think it is not the same point – it could be said that we should not expect, post-Quine, a sharp distinction between mistakes about meanings and mistakes about facts.

I want to concentrate on a third kind of case, however. It does not depend on ambiguity, in the normal sense, or on any mistake about meaning. It stems rather from a form of indeterminacy about meaning that seems endemic to language. In recent years philosophers have paid a lot of attention to the fact that language is finitely acquired. Each of us has acquired her or his linguistic knowledge in the few short years since birth – in so far, at any rate, as we did not have it at birth. This observation has been applied in a number of ways: by Chomsky and his followers, in arguing for the existence of innate linguistic ability; by Dummett, in his argument for an anti-realist conception of truth; and in the literature stemming from Wittgenstein’s treatment of the notion of following a rule. And a closely similar factor, concerning the finiteness of the evidence available to the radical interpreter, is crucial to Quine’s indeterminacy thesis, developed in *Word and Object*.

For our purposes, the significant point is that no finite level of experience can determine the application of a linguistic term to all possible cases, or exclude the possibility that two speakers – fully competent speakers, by existing community standards – will diverge in their application of the term concerned to some future case. It is true that this does not happen very often. The explanation of this fact, particularly in the light of finite acquisition, is one of the central problems in the philosophy of language. (Wittgenstein’s treatment of rule following in the *Philosophical Investigations* can be seen as an attack on one conception of what a solution would look like. Cf. McGinn, 1984, chapter 1.) Given however that it can happen, and cannot be excluded in advance, it seems that we live with the possibility that disagreements may turn out to

stem from the fact that participants have made different (though equally consistent) ‘projections’ of a term from a common finite base.

What happens in cases such as this? One common if often abused response is to say that the disputed point is ‘terminological’. I think the important thing is that we recognize that neither party is *mistaken*, in either of two ways: in using the term concerned with what is the wrong meaning, by previously established standards; or in a sense that would lead us to say that they had spoken *falsely*, given their own reading of the equivocal term. The dispute thus evaporates. Neither side concedes, and yet each recognizes that there is no longer a point in insisting that the case is answerable to an external standard, that only one side can meet.

These cases reflect the intuitively appealing idea that our judgements have truth values only in so far as they are determinate in meaning. An ambiguous judgement may have a determinate truth value by default, as it were, if all its possible interpretations share the same truth value. But there is no guarantee that this will be so, in general – ambiguities tend to show up when it is not so. And in any case, it is far from clear how to apply this notion when we are not dealing with an ambiguity that is finitely specifiable in terms of non-ambiguous atoms. To recognize ineliminable indeterminacy of meaning seems to be to recognize that our utterances are at best approximations to genuine (and therefore strictly truth-evaluable) statements of fact.

I hope it is no longer necessary to emphasize that the present approach does not claim to show that such intuitions are correct. We have offered no analysis that would afford such a judgement. Here, as elsewhere, our point is rather to observe that ordinary intuitions concerning the limits of factuality can be correlated with a certain pattern of usage – the existence of evaporative disputes – and to argue that this pattern can be explained in terms of a general account of truth, falsity and the function of argument.

The indeterminacy of meaning thus seems to be the basis for a form of universal non-factualism. That is, it offers a reason for departing from the pattern of usage characteristic of the statement form that may be applicable in any area of discourse. This topic-neutral form of non-factualism should not be confused with the form that Quine himself extracts from the indeterminacy thesis: the view that there are no genuine facts about *meaning*. (Cf. Quine 1960, p. 221.) Ignoring for a moment our scruples about the analytic viewpoint, we might say that the present view, in contrast, is that in virtue of the nature of meaning there can be no genuine facts about anything.

More correctly, the result suggests that the factuality of usage is always a matter of degree. Across the broadly fact-stating part of language – in English, roughly, the indicatives – the appropriateness of the statement form varies from case to case, both type to type and token to token. From the explanatory point of view there is nothing surprising about this conclusion. Only the analytic presumption leads us to expect a sharp distinction.

## 8.10 THE SOURCES OF SUBJECTIVITY

What then can be said in general about the sources of evaporative disagreement, and about its significance in other areas in which non-factualism has seemed attractive? A survey and prospectus seem in order.

Disputes evaporate when disputees decide that in virtue of the peculiarities of their own standpoints it is not appropriate for each to insist that the other is mistaken. Evaporation is a manifestation of a degree of felt subjectivity in judgement. The cases discussed above illustrate the range of potential sources of such subjectivity. On the one hand are cases in which the

source lies in the context or conversational role of the judgements concerned. Probabilistic and conditional judgements seemed to be of this kind; their tendency to yield insubstantial disagreements being attributable to their distinctive roles in discourse, and hence their sensitivity to the evidential context in which they are used. On the other hand are cases in which the source of subjectivity seems to lie more in some distinctive feature of the speakers themselves – in the idiosyncrasies of their grasp of the meaning of a relevant term, for example. (A more familiar case of this kind might be the subjectivity of judgements about secondary qualities; more on this in a moment.)

The distinction between intrinsic and extrinsic sources of subjectivity is far from sharp. The sources of moral or aesthetic relativity seem to lie somewhere between the two extremes, for example. However, I think it provides sufficient structure to enable us to locate the likely basis of evaporation and hence non-factualism itself in some of the other areas in which such an approach has seemed attractive.

On the contextual side, the modal operators are the obvious candidates. The epistemological approach to possibility and necessity can be seen as an attempt to clarify the dependence of modal judgement on evidential context; even if usually under the misguided assumption that what is called for is an analysis of modal judgement – an account of its truth conditions, or some such. Another possible candidate is causal judgement. Here, particularly, the issue is clouded by the appeal of Humean metaphysics – a route to non-factualism whose viability we disputed in part I. Metaphysics aside, however, non-factualism about causality has an appeal that rests on contextual and conversational factors. This is not the place to explore these factors in detail, but I mention two. In view of the connections between causality and probability – particularly the intuition that causes increase the probability of their effects – disagreements about single case causes seem able to evaporate when the same happens to connected judgements about probability. And the well-recognized association between causal judgement and counterfactual conditionals suggests that the sensitivity of the latter to background assumptions – about ‘the normal course of events’, for example – will also give rise to insubstantial disagreements about causality.

At the other end of the scale, I have already mentioned judgements about secondary qualities. It has long been felt that these exhibit a degree of subjectivity, attributable to their dependence on contingent features of our perceptual apparatus. The view is supported by cases of major variation between human subjects in sensory response to a given stimulus. A substance may taste bitter to some subjects yet be tasteless to others. It is interesting that our willingness to treat such a disagreement as insubstantial depends on the relative frequency of the different responses. We are inclined to say that the aberrant judgements of colour-blind subjects are indeed mistaken – that it is a *fact* that jacarandas are mauve rather than blue. But who would insist that chemical X is really bitter, when about 50 per cent of people find it tasteless? This phenomenon is easily explained from the functional viewpoint, of course; but again, it is troublesome for an account that expects a sharp underlying distinction.

What else lies at this end of the scale? I mentioned the possibility of idiosyncratic grasp of meaning – a potentially universal source of evaporative argument. My impression here is that among the determinants of meaning there might be some we would count as strongly individual – perhaps innate dispositions to categorize experience in certain ways – and others more dependent on factors external to the individual speaker, such as communal systems of convention. The latter possibility suggests a place in the scheme for another topic about which non-factualism has been popular, namely theoretical discourse. The recent popularity of scientific anti-realism stems largely from an appreciation of the difficulties of inter-theoretical

discourse. Once it is recognized that proponents of different scientific theories may lack a basis for argument, it has seemed difficult to hold onto the idea that each is really making claims about the same objective world. The present view suggests that the source of the problem is an unacknowledged commitment to the notion that there is such a thing as genuinely fact-stating discourse. If we reject this notion, we are free to say that the inter-theoretical case is simply a prominent form of a phenomenon endemic to language. Theoretical discourse is not a special case; it differs only in degree from discourse of any kind. (The anti-realist might agree, of course, but interpret the result as global anti-realism. More on this in chapter 9.)

How much of non-factualism can be explained in these terms? It is difficult to make any firm pronouncement – the field is quite large, after all – but the prospects seem to me to be good. If there is a major area of uncertainty it seems to lie in the cases in which non-factualism is motivated by a combination of metaphysical and epistemological considerations. I argued in part I that a determined factualist should be untroubled by these factors. The appropriate conclusion might therefore be that in so far as it rests on such intuitions, non-factualism is simply mistaken, and should not be expected to conform to the pattern we have identified. However, there are several less dismissive interpretations of an appeal to such factors.

The first involves the interpretation of a disputed non-factualist thesis simply as a first order claim about what there is in the world, and not also as a second order claim about the nature of discourse. The practical difference here is that the weaker thesis, unlike a typical non-factualism, will recommend that we abandon the form of discourse in question. The force of the argument of chapters 4 and 5 was that if we are to retain an area of linguistic practice, the non-factualist cannot dislodge a determined defence of a factualist interpretation of that practice. This leaves it open for someone to argue that we should discard the practice, however interpreted. Some of the appeal to metaphysics and epistemology in these contexts might be taken to be of this kind.

Secondly, it is easy to imagine that someone might appeal to metaphysical and epistemological considerations in order to establish the relatively theoretical nature of a particular area of discourse. To give a crude example, it might be said that since we don't directly perceive probabilities, probability must be a theoretical notion. Moves of this kind bring the discourse in question within the scope of a general claim about the status of theoretical language. We have seen that theoretical non-factualism can be brought within the pattern suggested by the explanatory approach. This interpretation of the appeal to metaphysics and epistemology thus brings it within the scope of our general account of non-factualism.

The last possibility is in some ways similar. There is clearly a role for metaphysical conceptual analysis in an approach of this general kind. We have already noted an example: an analysis of causality in terms of probability, if sustained, would be likely to ensure that causal judgement was similarly liable to evaporative disagreement. (This might cut both ways: the lack of such disagreement in the causal case might count against such a conceptual analysis.)

These possibilities suggest that there is a significant place for metaphysical and epistemological concerns in an explanatory approach to non-factualist concerns. But I must leave it to the reader to judge to what extent existing appeals to these factors can be interpreted in any of these ways – as, of course, to assess the approach to non-factualism as a whole.

## 8.11 THE SIGNIFICANCE OF MOOD

What then of the non-indicatives? Here, more than anywhere, usage may seem to call for the analyst's sharp dichotomy. Hence, as we saw, the appeal of the marker view. But here too, I

think the call should be resisted. It can be countered by the strategy of indicative paraphrase, which was the basis of our argument against the marker view in chapter 3. And we can now add to that argument a possible explanation for the existence of the non-indicative moods, in terms of our account of the function of truth-evaluability.

If a non-indicative is paraphrased in the way we suggested in chapter 3, it is normally true in virtue of being uttered. At any rate, we saw that the suggested indicative paraphrase is such that its truth is normally in the gift of a speaker who utters it. In so far as non-indicatives are informative, the information they carry is not normally open to dispute. This suggests that if the linguistic functions served by the non-indicatives were to be assigned to indicatives, the argument-encouraging machinery associated with the indicative mood would *usually* be redundant in these cases. The qualification is important: first because as we'll see in a moment, such machinery would not always be redundant; but more importantly because the evidence suggests that limited redundancy is often tolerated in the interests of the general benefits of the fact-stating form. After all, there are many ordinary indicatives whose truth is similarly guaranteed by utterance – 'I have not lost the power of speech', for example.

To explain why more frequent redundancy should be disadvantageous, I think we need to mention another factor: namely, that unused machinery may be a positive liability. For one thing, the argumentative mode of speech is presumably sustained at some conceptual cost. It requires that speakers hold themselves ready to assess the utterances of others by the standards reflected in judgements of truth and falsity. Where such assessment is generally inappropriate, or predetermined by the nature of the utterance, the general conceptual cost may not be worth the very occasional benefit. For another thing, the possibility of argument might itself be disadvantageous. If non-indicatives serve linguistic functions that call for specific and often immediate responses from an audience – answering a question, meeting a request, and so on – then it might well be counter-productive to use a form of language that encourages a different kind of response: namely, an argumentative challenge to the 'correctness' of the utterance in question. This is not to say that one cannot challenge a non-indicative utterance. Clearly one can, and we do – amongst other things on the kinds of grounds that would constitute a challenge to the truth of the kind of indicative paraphrases suggested in chapter 3. Ordered to complete a form, for example, we may object that we are not by law required to do so. The point is simply that there may be a social advantage in not actively encouraging such challenges.

These factors, taken together with the inevitable element of historical accident, are perhaps as much as we can expect of an explanation of the existence of the non-indicative moods. The striking feature of these moods is the fact that they are not taken to be truth-evaluable, in ordinary usage. The above factors bring this fact within the scope of our explanatory account of the role of truth and falsity. It would seem that the non-indicative moods occupy a region of language in which, on the whole, the disadvantages of the notion of truth outweigh its advantages. But the boundaries of this region are necessarily imprecise, and the sharpness of the mood distinction, if taken at face value, is liable to be misleading. The linguistic functions it masks embody no corresponding discontinuity, but rather a complex and multi-dimensional field of factors, varying by utterance token as well as by utterance type.

An artificial order is imposed on this field by the need for regularity in language. The rules of language cannot be sensitive to every nuance of the individual case. But we should not underestimate the considerable flexibility that does exist in natural language. In practice the mood distinction is far from sharp. As we have seen, English speakers will switch between indicative and non-indicative utterances for what are recognizably similar linguistic tasks – according, at least in part, to the perceived role for an idea of an objective standard. And within

the indicative mood, we have observed the striking tendency of ordinary speakers to recognize the limits of truth and falsity, to opt for tolerance, and to allow inappropriate disputes to evaporate. Usage is very much less sharply polarized than first appearances might suggest. It is therefore a point in favour of the present explanatory approach that it does not advocate or assume a sharp distinction between fact-stating and non-fact-stating uses of language.

## 9

**End of Enquiry**

We began with the conflict between non-factualism and Fregean linguistic theory – a somewhat ironic tension, we noted, given that non-factualism was a distinctive and popular product of the mid-century linguistic movement in philosophy. In this concluding chapter I want to begin by showing how this conflict seems to be resolved by the present explanatory approach to truth and factuality. This will lead us to the question as to why, on the present account, the notion of truth should be the kind of notion that it is. We encountered this question in chapter 7, and there and elsewhere have touched on the main aspects of the answer. Here I shall try to draw together the threads. As foreshadowed in chapter 2, David Wiggins’s ‘marks of truth’ provide a useful focus at this point. We want to confirm that the proposed functional account can explain why truth should have such characteristics. The third concern of the chapter will then be to clarify some of the connections and advantages of the functional approach, with respect to projectivism, quietism and other positions in contemporary debates about realism and objectivity. We’ll finish with some comments on the scope of the functional perspective in other areas, including particularly its application to the problem of the status of folk psychology.

## 9.1 TRUTH: A CONFLICT RESOLVED

In chapter 2 we looked at Wiggins’s attempt to ground an account of assertion on a substantial notion of truth, itself derived from a Davidsonian programme for radical interpretation. Our main criticism was that the programme of radical interpretation turns out to be insensitive to the distinction between fact-stating and non-fact-stating parts of language. The theorems of the radical interpreter’s so-called truth theory do not have to mention truth at all; and if they do mention it, need only the disquotational property – the equivalence principle – in terms of which the truth predicate can be extended trivially to utterances of any kind.

The impression that Fregean general semantics requires a substantial notion of truth thus rests on an uncritical acceptance of the doctrine of the primacy of assertion. If we equate *saying* with *asserting*, then of course we equate the disquotational predicate needed in a Fregean theory of meaning with the ordinary predicate ‘... is true’. However, if we keep in mind that there may be sayings that are not assertings then such an equation looks highly implausible – the ordinary predicate does not extend to non-assertoric sayings. Hence the needs of radical interpretation, or of a systematic theory of meaning of the kind it requires, seem to have little bearing on the question as to why the ordinary notion of truth should be restricted in this way.

I think that the resolution of the conflict between non-factualism and the Fregean programme for general semantics thus turns on a distinction between the disquotational and the normative properties of truth. Contemporary semantic theories exploit the disquotational property. The non-factualist, however, needs to be concerned with the normative property; for only the normative property seems able to explain the limited application of the ordinary notion of truth, and hence to ensure that truth has something to do with a distinction of language into ‘fact-stating’ and ‘non-fact-stating’ parts.

A Fregean might object that one thing in favour of a truth-conditional approach to a theory of meaning is the intuition that to know the meaning of a sentence is to know its correct use – to know under what conditions it would be correct to utter that sentence (with serious intent). This

intuition surely depends on the normative notion of *correctness*. So to the extent that it calls for a truth-conditional theory of meaning, it seems to require a normative notion of truth, rather than merely a disquotational notion.

However, I think this objection trades on an ambiguity in the notion of correct use. In the case of assertoric discourse this ambiguity is well recognized, if not always well heeded. The term ‘assertibility condition’ (or ‘condition for correct assertion’) can mean one of two things. Taken subjectively, it refers to something like the state of mind that would normally be considered to be evidenced by a competent speaker’s utterance of the sentence in question. According to the conventional model, a belief that P is thus the (subjective) assertibility condition for the sentence P; and there seems to be something to be said for the claim that to know the meaning of P is to know that this is the case. Taken objectively, on the other hand, ‘assertion condition’ refers to some external state of the world, whose obtaining would normally ‘justify’ the assertoric use of the sentence in question. This reading itself admits of various interpretations; for present purposes the important thing is that at least one interpretation has the effect of equating ‘correctness’ with ‘truth’. And again, there seems to be something to be said for the claim that to know the meaning of an assertoric sentence P is to know when its use would be correct in this sense – in other words, to know its truth conditions.

As long as we restrict ourselves to assertoric discourse, there is little incentive to keep in mind that there are these two readings of the principle that to know meaning is to know correct use. To the extent that they are distinguished, the objective notion has seemed of greater interest. But if we broaden our field of view to include the non-assertoric part of language, I think we should regard this as another case of taking for granted what we ought to explain. The availability of the subjective reading shows that to decline to rely on the objective reading is not to reject the principle itself. And in extending to non-assertoric discourse, the subjective reading has the generality that the objective reading lacks. In practice, indeed, the subjective notion of an assertibility condition has tended to receive attention in cases in which the notion of truth seems problematic, such as that of indicative conditionals.

We can thus aim initially at a theory of meaning that does not depend on a radical distinction between the assertoric and non-assertoric parts of language (regarding this rather as something that calls for explanation). Apart from its elegance, generality and explanatory potential, this approach has the advantage we discovered in chapter 2: it turns out to be a product of its opponents’ best efforts at following the alternative approach, given that truth theories do not in fact rely on the normative aspect of truth.

## 9.2 WHY IS TRUTH TRUTH-LIKE ?

This separation of the disquotational and normative aspects of truth may itself seem problematic. Why do these aspects go together? How is it that the ordinary normative truth predicate functions disquotationally? These are questions that an explanatory theory of truth should be expected to answer.

More generally, an account of truth on the proposed lines should be able to relate the features of truth as we find it to the functional role that is being claimed to explain its development in language. True, there is some flexibility in an evolutionary account of this kind. Some features of the ordinary notion may be more or less accidental. But the plausibility of any such claim seems to rest on the relative insignificance of the feature in question. The more a property seems essential to truth, the more we should expect an account of why truth should have such a

property. In particular, we should expect an account of the marks of the concept of truth, in Wiggins's sense (1980, p. 194).

In the case of the connection between the disquotational and normative aspects of truth I think the answer is fairly straightforward. To explain the disquotational property we have to explain why P and 'P is true' can in some sense be regarded as equivalent assertions – for example, that they can be taken to be expressions of the same commitment. On the present theory, to say that P is true is to allow to judgements that P a conventionally-generated property that we are all conditioned to want our judgements to have. Roughly, it is to ascribe worth to judgements that P. The point of this conventional system of value is to encourage argument. In making our case for these conventional rewards, we indulge in a process whose effect tends to be to improve our behavioural commitments. But this improvement depends on the fact that (in the main, at any rate) argument is prompted by some mismatch between individual behavioural commitments, and resolved by a resolution of such a mismatch. So the process requires that we ascribe value to those judgements we are prepared to make ourselves. Only in this way can the ascription of these conventional rewards make socially unstable the appropriate cases of mismatched commitment. In other words, the role of truth in argument requires the disquotational property – it requires that we be prepared to acknowledge that P is true in the same circumstances as we acknowledge that P.

This answer bears on the problems that might be thought to flow from the fact that ascriptions of truth and falsity themselves take the grammatical form of indicative sentences, and are thus the sorts of expressions to which 'true' and 'false' can be ascribed. We touched on this point in chapter 8, mentioning in particular the objection that it might commit us to an endless hierarchy of disagreements, corresponding to iterations of the predicates '... is true' and '... is false'. In explaining how we avoid embarking on such an escalation of ordinary disagreements, the disquotational property is part of the answer. It means, I think, that from the point of view of our behavioural commitments there is no reason to escalate. Behaviourally speaking, semantic ascent doesn't increase the stakes.

The other part of the answer, I think, is that in the case of disagreements about truth and falsity themselves there isn't the normal reason to *introduce* these normative terms. Truth comes into the picture to provide the incentive for argument. But once it is already in the picture – when we are disagreeing about the truth of some sentence – there is nothing to be gained by bringing it in again, by moving to second order truth ascriptions. There's a prize for winning the argument, but no further prize for winning the prize. This is not to say that ascriptions of truth and falsity are not themselves regarded as answerable to the kind of objective standard the concept of which their use embodies. On the contrary, as we stressed in chapter 8, it is crucial to the motivating role of such evaluations (as of others) that they be regarded as objective – that there be no easy subjectivist escape hatch. It is simply to say that in the case of truth and falsity themselves, the manifestations of this felt answerability to an external standard do not have their usual importance in argument.

This brings us to Wiggins's marks of truth. We are now in a position to explain why truth should have the characteristics that Wiggins identifies – or rather, as we'll see, why it should have those four of the five such characteristics that are properly regarded as properties of truth.

Wiggins's first mark is the normativity of truth. It consists in truth's being 'the primary dimension of assessment for sentences' – 'that property which sentences have *normally* to be construed as aiming to enjoy' (1980, p. 205). This property is central, of course, to our proposed

account of the function of truth in discourse. Unless we take truth as something to be aimed at, we will be given no incentive to argue by the claim that one of our utterances lacks it.

The second mark of truth concerns its attainability. Wiggins puts it as follows.

A second mark of the [truth] of a sentence *s* would appear to consist in *s*'s *having a content such that with respect to this content there should be a tendency, under favourable conditions of investigation, for disagreement to diminish and for opinion to converge in agreement.* (1980, p. 206)

Here the relevant point would seem to be that unless speakers subscribe to some such view of truth then even its normativity will not encourage them to argue. Heroics aside, why should we argue with someone unless initially there seems some prospect that he can be brought to share our own opinion – in other words, some prospect of agreement? A degree of attainability thus seems a necessary part of our idea of truth, if truth is to serve the role we have claimed for it in encouraging argument.

Similar remarks apply to Wiggins's third and fourth marks of truth. The third mark (1980, p. 208) is that truth be considered independent of a speaker's means of recognizing whether a statement possesses it; while the fourth (p. 211) is that true sentences be considered to be true 'in virtue of something'. These would appear to be aspects of the required objectivity of truth, on which we commented a moment ago. The power of truth as a system of incentive depends on the fact that it is not taken to be able to be bestowed at will. This is what distinguishes the incentive provided by truth from the incentive provided by a system of boos and cheers. Rival views may share the available applause, but they cannot share the truth – or so, at any rate, it is useful for us to believe.

So to Wiggins's fifth claimed mark of truth (1980, p. 211): if two sentences are true then so is their conjunction. It is difficult to take this as a property of truth, rather than as an account of conjunction in terms of truth. There is a legitimate question as to why we should have a sentence-forming operator *and* such that we are prepared to assert *P and Q* when and only when we are prepared to assert both *P* and *Q* individually. Given however that we do have such an operator, the fifth mark follows directly from the disquotational property – from the fact that we are prepared to assert that *P* when and only when we are prepared to assert that *P* is true.

### 9.3 THE TRUTH OF TRUTH ASCRIPTIONS: SOME OBJECTIONS EVADED

I mentioned above the treatment of truth ascriptions themselves by the present explanatory account of truth. I want to draw attention to some problems that trouble analytic non-factualists at this point – problems that again the explanatory approach avoids.

For one thing, the analytic point of view leaves a non-factualist with the following dilemma. Either 'P is true' is sometimes a statement and sometimes not a statement, depending on the status of *P* itself; or all such truth ascriptions have the same status. Neither choice is particularly appealing. The former entails that truth ascriptions are not the semantically homogeneous class their form suggests. While the latter implies that 'P is true' can be a genuine statement even though *P* itself is not a statement, or *vice versa*; and hence that the equivalence principle fails. From the explanatory perspective, however, the dilemma looks entirely self-inflicted – a product of the assumption that any utterance lies on one side or other of a real semantic divide.

Arthur Fine has recently raised a related objection to a treatment of truth on projectivist lines:

The [projectivist] strategy requires a thin, unproblematic discourse on which to base its projections. To project out ‘truth’, the thin discourse must be truth free. But where are we to find any such thin discourse? The redundancy property of truth makes truth a part of discourse that merits the name. In particular, if we go for acceptance or coherence theories of truth we try to build truth from certain systems of judgements (or the like). But the judgement *that P*, and the judgement *it is true that P* are the same. You cannot have one without the other. That is to say, you cannot include one in the system without thereby including the other .... Thus there is no thin discourse from which the [projectivist] explanations of truth talk could begin. (1986, pp. 169-70)

This objection is like the above one in squeezing a non-factualist account of truth between on the one side the desire for a unified theory of truth, and on the other the equivalence principle (or redundancy property). If the projectivist is prepared to concede unification then there is again an easy way out: to offer separate accounts of ‘thin’ truth and ‘thick’ truth – separate interpretations of truth ascriptions to utterances that are themselves about, respectively, ‘genuine’ or ‘projected’ facts. (This seems to be the approach taken by Blackburn 1984, though not without reservations; see particularly pp. 256–7.)

However, I think that in Fine’s case, as in the earlier one, the explanatory approach needs no such reply. Because it is committed to no prior bifurcation in language, it cannot be turned on itself in the way that an analytic theory is bound to find embarrassing. True, its functional account of the role of truth, and fact-stating discourse in general, will itself be couched in fact-stating terms. As its proponents, we will be inclined to assert its *truth*. But there is nothing more vicious in this than there ever is in speaking or writing about the origins of language. If there were no more to Fine’s objection than the observation that the projectivist must use ‘true’ (in effect, at any rate) in offering a constructive theory of truth, the point would be trivial and irrelevant. Its force therefore rests on the consequences of bifurcation, to which the projectivist is committed but our theory is not.

#### 9.4 FUNCTIONAL REALISM: NOT IDEALISM ...

What then of the suggestion that the present theory is simply idealism? As we saw in chapter 4, the idealist takes over the projectivist’s goal of accounting for the realist features of linguistic practice but aims to do it quite generally. Our approach thus concurs with idealism in rejecting the projectivist’s bifurcation between real and projected facts, arguing that there is a single appropriate account of truth, applicable wherever truth is. Given that the idealist takes over the projectivist’s constructive programme, we also agree on an explanatory as opposed to an analytic approach to truth, and perhaps on some of the details.

The main difference, I think, concerns the treatment of non-factualist claims. The projectivist aims for ‘quasi-realism’, as Blackburn calls it – for a demonstration on projectivist lines that we are entitled to talk *just as if* there were genuine facts of the disputed kinds. Blackburn’s quasi-realist seems not to countenance differences in usage between the disputed utterances and genuine statements, aiming instead for the perfect pretence. So when the idealist challenges the distinction between pretence and reality, non-factualism, having been lashed to quasi-realism, follows it down the slope.

The present theory, on the other hand, goes out of its way to draw attention to some characteristic peculiarities of the kinds of discourse with which non-factualists have been concerned. It tries to associate these peculiarities with particular variations in the factors that seem normally to account for the functional advantages of truth and related features of usage. In

the process, as well as demonstrating its ability to explain these cases, it offers some foundation for the intuitions that have usually motivated non-factualists. True, this usage-based descriptive–explanatory programme is not what non-factualists have usually had in mind. But unlike Blackburn’s quasi-realist projectivism, it does offer a refuge for the non-factualist against the push to generalize the constructive or explanatory approach to truth.

In summary, I think the present approach concurs with the idealist in seeking to globalize the projectivist’s constructive approach to certain areas of reality – the attempt to ground our conception of the world ‘internally’, so to speak, in our own patterns of reaction to our environment. Unlike the idealist, however, we find room in this for a kind of non-factualism. This aside, the term ‘idealism’ has associations the present account will want to avoid. It is apt in particular to suggest an analytic approach to truth, in terms of coherence, or some such – an approach quite different from ours and from that of the projectivist’s opponent. The present theory is best described as *functional* or *explanatory* realism. Its goal is to explain our realist practices; as well, of course, as their non-factualist variations.

In expressing even this much sympathy for idealism I am likely to be accused of denying common sense intuitions about the reality of the external world. So I want to make it clear that my views about most aspects of the external world are the same as those of the most hard-nosed realist. We each hold true much the same class of propositions about how things are. Our professional differences concern at most a tiny fraction of such propositions: those that deal with the explanation of the fact that humans hold such views. For my part, I believe that at least in so far as holding a view about the world depends on our possession of the notion of truth or correct representation, the fact that we do so is explicable in the kind of terms here suggested. Hard-nosed realists may have different views on this topic, or none at all. But this doesn’t prevent us agreeing about almost everything else (in other words, about almost everything).

In particular, it would be a mistake to think that because I have a biological explanation of the fact that we apply a notion of correctness to judgement I am unable to apply such a notion myself. That would be as absurd as supposing that a biologist who explains hunger in evolutionary terms should no longer feel the need to eat. Indeed, it would perhaps be more absurd, for whereas to hold a theory about hunger is not to feel hungry, a belief about the origins of truth is an exercise of the very biological capacity it itself concerns. So it demonstrates its author’s continued ability to exercise this capacity.

### 9.5 ... NOT INSTRUMENTALISM ...

I should perhaps emphasize also that functional realism is not a form of instrumentalism. Indeed, it is likely to reject instrumentalism on the same grounds as it rejects projectivism. Both these approaches ask us to make a distinction between two kinds of superficially factual discourse: the genuinely factual areas, on the one hand, and the instrumental or projected areas, on the other. The present theory denies that we can draw such a distinction. The best it can offer an instrumentalist is a place in the ranks of the reconstructed non-factualists; and this only if the instrumentalist can point to the relevant sort of difference in usage between the supposed instrumental discourse (theoretical judgements in science, or whatever) and the canonical assertoric form.

All the same, I think there is a nice analogy between the instrumentalist’s most powerful argument against orthodox scientific realism and an important argument for the present explanatory approach to truth. The instrumentalist’s argument is described by Fine, in the paper from which I quoted above. Fine notes that

the realist must at least allow that, generally speaking, truth does lead to reliability. For instance, ... the realist offers the truth of a theoretical story in order to explain its success at a certain range of tasks. If this offering is any good at all, the realist must then allow for some intermediate connection between the truth of the theory and success in its practice. The intermediary here is precisely the pragmatist's reliability. Let us therefore replace the realist's 'truth' with the pragmatic conception, framed appropriately in terms of reliability. Then, if the realist has given a good explanatory account to begin with, we get from this pragmatic substitution a good instrumentalist account of the same phenomena. Indeed, the instrumentalist account we obtain in this way is precisely the intermediate account that the realist has himself been using. Since no further work is done by ascending ... to the realist's 'truth', the instrumental explanation has to be counted as better than the realist one. (1986, p. 154)

The instrumentalist thus accuses the realist of a redundant metatheoretical commitment.

An analogous argument can be drawn from our account of the role of truth in encouraging argument. We saw in chapter 7 that if one begins with a conception of truth as a property of belief, and asks what the point of argument could be – why we should care whether our beliefs have this property – then at first one may think of truth as an end in itself. It becomes apparent, however, that truth could at best be an intermediate goal in one's quest for something like behavioural success. On the other hand, we have also seen that progress towards behavioural success seems to require, or at least to be enhanced by, our all *behaving* as if there were such an end as truth to judgement and enquiry. If there were no real property of truth, we would be well advised to invent one. In other words, the real or substantial notion of truth now looks redundant – unnecessary either to account for the utility of argument, or to explain our *use* of a notion of truth.

This parallel is hardly surprising. After all, the redundancy argument is the hallmark of the claim to conceptual economy, itself characteristic of the general philosophical movement to which instrumentalism and the present theory both belong. Such arguments involve some characteristic risks, however. As the quasi-realist's problems illustrate, they tend to backfire, undermining the very positions their advocates wish to preserve. Economy needs to be backed up by stability: the thin residue must be sturdy enough to stand on its own.

Fine argues, I think rightly, that instrumentalism cannot stand on its own. As he puts it, 'instrumentalism avoids the inflationary metaphysics of realism only by pumping up either epistemology or semantics' (1986, p. 170). For the semantic case, his argument is the objection to a projectivist theory of truth we noted earlier. I think that like ours, his basic objection to instrumentalism is that it involves an untenable distinction between 'genuine' and 'instrumental' areas of factual discourse.

## 9.6 ... AND NOT QUIETISM

In the face of these objections, however – on the one side to realism, and on the other to instrumentalism – Fine recommends that we be satisfied with what he calls 'NOA ..., the *natural ontological attitude*' (1986, p. 172). He says that

NOA allies itself with what Blackburn ... dismissively calls 'quietism'. Less pejoratively, NOA holds a 'no-theory' conception of truth .... NOA does not think that truth is an explanatory concept, or that there is some general thing that makes truths true. (1986, p. 175)

Fine also links NOA to Rorty's recent views on truth (on which we remarked in chapter 6), saying that Rorty's 'developing neo-pragmatism ... seems quite NOAish' (1986, p. 173, n. 16).

My use of the term 'quietism' has not been exactly Blackburn's. I used it initially for the view that declines to theorize about the fact-stating/non-fact-stating distinction, rather than for the analogous view about truth. But clearly the two views run parallel, and my sympathies are with Blackburn.

NOA is right, I think, in rejecting attempts to *analyse* truth. And it is right to dismiss recent attempts to make truth itself an explanatory notion. (We considered these attempts in chapter 6. I argued that far from supporting a realist or correspondence notion of truth, they depend on no more than the disquotational or redundancy property.) NOA is wrong, however, to conclude that there is no further role for a theory of truth. At the very least there remains the not inconsiderable task of accounting for the fact that we have such a notion at all.

The quietist is simply blind to this, a blindness perhaps explained by the tendency on all sides in these discussions to ignore non-descriptive uses of language. From this perspective we are liable to overlook the contrasts in language that draw attention to the remaining explanatory task. To notice the non-indicatives, and the issues raised by the non-factualists, is to notice not only that truth is not a universal notion in ordinary language, but also that there is considerable doubt as to the limits of its application. At this point, I think, to decline to theorize about truth would be an attitude to ontology with all the naturalness of the ostrich's famed response to challenges of a different kind.

I have urged that the most profitable approach to this explanatory task involves a conceptual retreat. We need to step backwards, asking not why truth matters but why it matters that ordinary speakers should think that it matters. We need to enquire what purpose could be served by our ordinary conviction that there is an end to enquiry. My account of the purpose of this conviction may not be the most appropriate one, but I think the stance is right. We need to examine the game rather than the goal. Concerning truth itself, in other words, I think there is more truth than usual in the pious sentiments of Grantland Rice.

For when the One Great Scorer comes  
To write against your name,  
He marks – not that you won or lost –  
But how you played the game.

## 9.7 LANGUAGE AND MIND: EXPLANATORY PROSPECTS

I want to finish with some comments on the scope and potential of the general philosophical approach exemplified by this functional theory of truth. In particular I want to draw attention to its relevance to recent work on mind and language – to the philosophical end of cognitive science. Here its significance lies in the fact that it is at the same time naturalistic and non-reductive. I think it thus offers an important and undervalued alternative to recent work – a middle way for cognitive science, between the evils of non-naturalism on the one hand and the intractability of reductivism on the other. The middle way may not be the right way, but it needs to be recognized as one of the options.

The goal here is old enough: an understanding of the ordinary concepts of language and mind; of 'truth', 'meaning', 'belief', and so on. Increasing scientific interest in the mind has merely added weight to the view that this understanding should be sought in natural terms – in a manner

compatible with the fact that we ourselves are part and product of the natural world. It is hard to object to the naturalistic constraint so stated.

So stated, however, it is neutral between reductive and explanatory accounts of the concepts of language and mind. Recent writers have generally taken it to call for the former. The task has been taken to be to give a naturalistic account of what we are talking *about* in talking of truth, belief, and the rest. Hence the philosophical industry devoted to the naturalistic analysis of these notions – an industry founded in large part on the view that the only alternative would be to give up these notions, to regard them as superseded by scientific psychology and linguistics. True, opinions differ as to the acceptability of that alternative. Not everyone agrees with Fodor that

if commonsense intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species; if we're wrong about the mind, then that's the wrongest we've ever been about anything .... We'll be in deep, deep trouble if we have to give it up. (1987, p. *xii*)

For a less cataclysmic assessment see Stich (1983), chapter 11 or Churchland (1984), chapter 2.5, for example. All the same, there is a very widespread view that the fate of the folk notions rests on the success or otherwise of the reductive enterprise.

The functional approach suggests a challenge to this view. We have seen that it offers an account of the existence of the folk notions of truth and falsity, without seeking to analyse these notions in naturalistic terms. Its naturalism is that of evolutionary biology: it explains why creatures like us should have come to talk this way. But now why stop at truth? Why not look for parallel accounts of the rest of folk psychology and linguistics?

What would such a theory of intentional psychology be like? With one main difference, I think it might be very like that of Daniel Dennett, perhaps the most notable opponent of the view that the fate of folk psychology hangs on the possibility of its reduction to a physical basis. Dennett has long been concerned to save a place for the folk notions, given (as he accepts) that there is no place for them in neurophysiology. Briefly, his view is that beliefs and desires are 'idealized fictions in an action-predicting, action-explaining calculus' (1978, p. 30; see also 1978, chapter 1, and 1981). Clearly this has the makings of an explanatory theory in our sense. As Dennett emphasizes, the practical advantages of such a calculus in ordinary life provide the basis of an evolutionary explanation of the fact that humans have developed it.

The difference is that as the quote suggests, Dennett presents himself as an instrumentalist about folk psychology. On the present view, however, this reading of such a theory is unnecessary and unfounded. A sharp realist/instrumentalist distinction would call for a sharp factual/non-factual distinction; and that, we argued, is not to be had. This suggests that things are better than they seemed for an account of Dennett's kind. The explanatory perspective doesn't call for ontological apologies – or at least not such profuse ones, the difference between psychology and neuroscience being at most a matter of degree.

Stripped of its instrumentalism, Dennett's view thus seems very much in keeping with our general explanatory approach. However, it may then seem too good to be true. If it isn't instrumentalist, if it doesn't deny that there really are beliefs and desires, then how does it evade the strong intuition that there ought to be a physical account of what such things *are*? How in other words does it distinguish itself from reductivism?

The answer will be in two parts. The first step is to defuse the physicalist intuitions – to appeal to factors in virtue of which it seems implausible that ascriptions of psychological attitudes could be equated with ascriptions of physical states. (Not surprisingly the most comprehensive accounts of these factors are those of the opponents of realism about folk psychology; see, particularly, that of Stich 1983.) The second step is then to deny that the first calls for an anti-realist view of folk-psychological attitudes. The case here draws on the argument of this book, in a way that I hope will by now be familiar. Briefly, the anti-realist construal rests on an application of an untenable fact/non-fact distinction. All areas of discourse are potentially non-factual to some degree, but the factors responsible may vary from case to case. The peculiarities of folk-psychological discourse seem to distinguish it in this respect from talk about neurophysiology, with the result that we shouldn't expect a reduction. But only a philosophical misconception about truth could support the conclusion the anti-realist wants to draw from this: namely, that only the claims of neurophysiology are genuinely factual, truth-bearing, or whatever.

This approach to the problem of folk psychology in cognitive science thus applies the explanatory approach at two levels. It applies it first to the folk notions themselves, recommending an explanatory treatment in the same evolutionary vein as our functional theory of truth. The result, plausibly, is Dennett's account of the origins of intentionality. The second application then concerns the interpretation of this theory of psychology. The issue is the factual status of folk-psychological claims, and here an explanatory view of truth affords no basis for a sharp distinction. It was our disenchantment with the possibility of such a distinction that led us to the explanatory approach.

I don't want to sound too sanguine about the prospects for explanatory accounts of language and mind. Nor do I want to suggest that intuitions about ontological and theoretical priority can simply be dismissed out of hand. It would be much more in keeping with the present approach to try to explain these intuitions and the phenomena of theory change – to explain them in terms of our general conception of the function of factual discourse. Clearly usage is not immutable in the face of such considerations, and it would be nice to be able to explain why it changes when it does change.

For the moment, however, I want simply to urge some critical attention to the terms in which the current debate about the status of folk psychology is being phrased. The questions as to whether there are 'really' beliefs and desires, as to whether ordinary ascriptions of such attitudes have genuine truth values, are themselves phrased in folk-linguistic terms. Truth, fact and reality are folk notions. Their status is as much at issue as that of the folk-psychological notions – to which, evidently, they are closely related. Hence it is a mistake, I think, to expect or assume different kinds of answers in the psychological and linguistic cases. In particular it is a mistake to allow an unexamined philosophical conception of truth to govern one's interpretation of the issue about the status of folk psychology. There is a kind of consistency constraint here. If we're physicalists about belief then we ought to interpret the question of the truth status of belief ascriptions in terms of a physicalist notion of truth. But if we take the explanatory approach in both cases then we deserve to be judged by the standards of that approach.

It thus turns out to have been important to treat truth first. To accept an explanatory account of truth is to constrain one's understanding of any other theory in the same general spirit. Explanatory theories have in the past been crippled by their failure to apply their insights to truth itself (think of the problems of projectivism, for example). The promise of the explanatory approach lies in considerable measure in its ability to explain the terms in which the traditional issues have been phrased.

**PART III**

**Thirty Five Years Downstream**

## Chapter 10

### Embracing Minimalism

The remaining five chapters, new in this edition, are an extended postscript to the original material. They discuss issues that struck me as I reread FFT after many years, trying to unpick my evolving relationship to its author. Most turn on connections between the themes of FFT and later work, by me and by others. In the new Introduction (Chapter 0) I was trying to make FFT *accessible* to contemporary readers. In these five chapters I'm trying, among other things, to make it *locatable*, by positioning its concerns and conclusions with respect to later discussions of related topics.

In Chapter 0 I noted that more than thirty years downstream from the first edition, FFT is still swimming firmly against prevailing currents. I think it is fair to add that there have been some confusing cross currents in the field over that period, terminologically among other things, and that my own trajectory and terminological choices have contributed to them. Another goal of these remaining chapters is to sort out some of these confusions.

Terminology has moved on in thirty years, and these days 'expressivism' is a common term for what in FFT I called 'non-factualism'. To avoid switching between one and the other, I'll use 'expressivism' in these new chapters, even if I'm talking about the argument and conclusions of FFT itself. I'll retain 'non-factualism' in direct quotes from FFT.

One of the issues we'll discuss is what 'expressivism' actually means. It will be important to distinguish several different contemporary uses of the term. These differences are to some extent a matter of terminological choice, but to understand the landscape, we need to see where the choices need to be made, and what the options are. Much of Chapter 11 deals with questions of this kind. As we'll see, they are not just issues for expressivists, in any of the available senses of the term. There are complementary issues about characterising the alternatives to expressivism, and the problems on that side are in some ways even worse.

The present chapter goes like this. I begin (§10.1) by connecting FFT to a long debate since it was published about the implications for expressivism of minimalist theories of truth. I touched on this issue in the new Introduction (§0.3.1), and here I offer my perspective on the issue, as it has developed in the past thirty years. I shall also identify a puzzle about the relation between two different strategies for defending expressivism that can be discerned in FFT itself and in much of my later work – I shall return to this puzzle in Chapter 11. In §10.2 I then discuss James Dreier's well-known Creeping Minimalism challenge (Dreier 2004), which will bring us to the demarcation problems just mentioned.

Next, I explore the implications of minimalism, within my general framework, in two directions. In §10.3 I discuss the relation between my brand of 'explanatory or functional realism', as I put it in FFT, and a kind of laid-back 'non-metaphysical' realism that has been enjoying new popularity. Like several recent authors, I argue that these 'relaxed realists' (McGrath 2014) are simply failing to engage with the explanatory questions that motivate expressivism. Their options, once these questions are in view, are to revert to a less relaxed species of metaphysics, to endorse a wholly implausible quietism, or to join forces with expressivists.

§10.4 and §10.5 then turn to the famous Frege-Geach objection to expressivism – an odd omission from FFT, it now seems to me, since I had proposed a response to it in my PhD thesis.

I'll describe the view I presented soon after FFT (Price 1994), that minimalism provides the expressivist with a response to the objection; and relate it to recent work, particularly by Mark Schroeder (2008, 2015) and Joshua Gert (2021).

Building on an issue raised in §10.2, §10.6 explores an analogy between a recent proposal by Schroeder and a move I have characterised in recent work as a 'new bifurcation thesis'. Finally, in §10.7, I link Schroeder's point to some ideas I have discussed in the work of Wilfrid Sellars, and close by noting a difference between Schroeder and Sellars on one hand, and myself on the other. They count as local expressivists, whereas I profess to be a global expressivist. Explaining the difference will be another task for Chapters 11 and 12.

## 10.1 Minimalism and functional pluralism

In Chapter 4 of FFT, my orthodox expressivist opponent, seeking a foundation for the factual/non-factual bifurcation, is caught between two rivals. On one side is the person I called the radical factualist, who accepts that evaluative claims (say) express pro-attitudes, but insists that such things may also be beliefs. On the other side is an equally ambitious expressivist, who invokes the projectivist move everywhere, and hence ends up, by her own lights, as a kind of idealist. What mattered for the dialectic of Chapter 4 was not whether these options are attractive, or in the end distinct. What was important was that they *look* unwelcome, by the traditional expressivist's lights, because they abolish the factual/non-factual distinction – in one case by saying that all is factual, in the other by saying that all is non-factual.

In §0.3.8 I noted that when idealism appears again, in chapter 9, I'm aware that it looks similar to my own 'global' view. But I fall back on the claim that my view admits a diversity to which the idealist is blind. This diversity is not the old factual/non-factual bifurcation, but amounts to degrees of factuality, explained in terms of the variable possibility of No Fault Disagreements, from discourse to discourse.

In FFT I was becoming aware of the attractions of a kind of deflationary realism – in Chapter 9, as we saw, I labelled my view 'functional or explanatory realism'. I wanted to regard this deflationary realism as compatible with the insights of expressivism, of course. But this possibility – deflationary realism plus expressivism – was already disputed. At the time of FFT it had already been argued that minimalist approaches to truth were incompatible with expressivism, and this claim soon became louder. The claim itself was soon contested, but my sense is that there isn't yet a clear consensus about where the contest leaves us. Some writers still view the impact of these deflationist moves as bad news for expressivism.

This is relevant here for two reasons. First, contemporary readers may be puzzled by FFT unless they understand that it was already challenging this 'minimalism is bad news for expressivism' view. Second, my own work soon took what seems in some ways to be a different direction from FFT, in response to these kinds of issues. I want to discuss that divergence, and it will be helpful to explain where it came from.

For clarity, note that there are two varieties of minimalism, or deflationism, that will be relevant here. One is deflationist about *semantic* notions such as truth and reference; the other about *metaphysical* questions, such as matters of ontology. These are often closely related, and it is a nice question whether one could really subscribe to one but not the other – more on this below – but it will be helpful to regard them as at least notionally distinct.

Beginning on the semantic side, the argument that semantic minimalism defeats expressivism appeared in several places in the years immediately after FFT was published – e.g., in Boghossian (1990), Humberstone (1991), and Wright (1992). These pieces soon generated a response from Jackson, Oppy, and Smith (1994), who proposed a defence of the coherence of expressivism, in the face of minimalism. They argued that minimalism about *truth* need not imply minimalism about *truth-aptness*, and that the latter will do for the expressivist’s purposes. Moreover, they suggest that the latter can be explained in psychological terms. Truth-apt claims are those that express *beliefs*, rather than some other sort of psychological attitude. (Jackson, Oppy, and Smith are not expressivists themselves, but merely arguing that semantic minimalism does not rule it out.)

However, FFT had considered the attempt to rest a factual/non-factual distinction on a belief/non-belief distinction, and concluded that it wouldn’t bear the weight. So the proposal by Jackson, Oppy, and Smith left me unconvinced. Moreover, it seemed to me that their paper missed a much more interesting avenue for defending the compatibility of expressivism and minimalist theories of truth. With John Hawthorne, in a piece first written in 1994, I offered what I took to be a better way of combining expressivism with semantic minimalism (Hawthorne and Price 1996). It rested on two observations. First, what is essential to expressivism (or, as we said then, non-cognitivism) is the idea that the disputed vocabularies have distinctive functions, and that other philosophies go wrong in missing this point.

Most basically, we suggest, what these positions have in common is that they characterise a linguistic *function*, or *category*, in terms of which the non-cognitivist may claim that the disputed sentences *serve a different function from*, or *belong to a different category to*, other parts of language (and in particular, to paradigmatic causal–explanatory parts of language). (Hawthorne and Price 1996)

We note that it is a ‘distinction of this sort that underpins the philosophical impact of non-cognitivism, particularly in opposition to reductionist or eliminativist moves’:

Non-cognitivists argue that these programs, and the philosophical concerns from which they arise, rest on a distinctive kind of mistake about language—on a misidentification of the linguistic category within which particular families of concepts have their home. (Hawthorne and Price 1996)

Our second observation is that the function in question need not be defined in terms of notions such as truth and belief, rendering the view immune to deflation of these notions by minimalism. We illustrate the point with three examples: Brandom on reference, Gibbard on norms, and Horwich on truth. We point out that all these views subscribe to the ‘metaphysics rests on a linguistic category mistake’ idea, and yet at the same time are thoroughly comfortable with the application of minimal notions of truth and belief, in the vocabularies in question.

In a slightly later piece I characterised the upshot of these points like this:

Noncognitivists were right in thinking that the notion of linguistic function provides a naturalistic solution to metaphysical concerns, but wrong to try to characterise the functions concerned in the terms normally used. To get things right, we need to retain the insight that different bits of language may serve different functions in a way which isn’t obvious at first sight, but set aside the usual attempts to characterise the functions concerned in terms of truth, factuality, belief, and the like. (Price 1997)

In that piece I call the approach I am recommending *functional pluralism* (noting that it is what Hawthorne and I had referred to as ‘a version of noncognitivism’).

Later again (e.g., in Macarthur and Price 2007; Price 2009a), it seemed to me more accurate to say that traditional expressivism already had the functional stories it needed, but simply made the mistake of bundling them with something it didn’t need, and which minimalism undermined. Put this way, traditional expressivism is interpreted as making two claims: a *negative claim*, to the effect that the utterances in question are not truth conditional (or something of that semantic kind); and a *positive claim*, to the effect that the utterances in question have such-and-such other (not semantically-characterised) function.

As Macarthur and I (2007) point out, semantic minimalism undermines the negative claim, which is incompatible with the deflationist’s thesis that semantic terms do no substantial theoretical work; but it leaves the positive claim intact. This means that so long as expressivism puts all its weight on the positive, functional story, it is completely untouched by the deflation of the negative claim. Certainly some of the ways that expressivists traditionally characterised their view have to be abandoned, but this clarifies and strengthens the view.

Indeed, the upshot is that semantic minimalism is bad news for the expressivist’s opponents, for it deprives them of the semantic vocabulary in which they find it natural to state their own view – i.e., a vocabulary that puts theoretical weight on notions such as truth and reference. That may be a problem for *local* expressivists, but because it recommends global expressivism, not because it recommends global factualism. (In a deflationary sense it does recommend global factualism, but this is no help to the opponent who wishes to disagree with the expressivist’s claim about the functions of particular vocabularies.)

These arguments in support of expressivism raise a puzzle about my own trajectory. In effect, I have offered *two* defences of expressivism: first, the argument in FFT, via the possibility of No Fault Disagreements; and second, the argument of (Hawthorne and Price 1996) and many later pieces, turning on the idea that all the necessary work is done by the expressivist’s positive, functional story.

What is the relation between these two lines of defence? I’ll return to this question in chapter 11. Answering it adequately will turn out to require a great deal of further clarification of the theoretical landscape. Before that, in the present chapter, I want to pursue the theme of the relationship between expressivism and semantic minimalism into the present century. In particular, I want to mention Jamie Dreier’s influential paper, ‘Meta-ethics and the Problem of Creeping Minimalism’ (Dreier 2004).

## 10.2 Creeping minimalism

Dreier characterises Creeping Minimalism as follows:

Minimalism sucks the substance out of heavy-duty metaphysical concepts. If successful, it can help Expressivism recapture the ordinary realist language of ethics. But in so doing it also threatens to make irrealism indistinguishable from realism. That is the problem of Creeping Minimalism. (Dreier 2004, 26)

Dreier doesn’t raise this as a problem for one side or the other, but for the field as a whole:

[T]he problem is not a problem *for* realists or *for* irrealists, but more a problem in meta-meta-ethics. It's not as if one side had better be able to come up with something clever to say about how to distinguish realism from irrealism or else the other side wins. It's rather that those of us who feel confident that there is some difference between the two meta-ethical camps should be concerned that we don't know how to say what that difference is. (2004, 31)

The solution Dreier proposes – drawing, as he says, on John Hawthorne's and my proposal from (Hawthorne and Price 1996), as well ideas from Kit Fine (2001) and Allan Gibbard (2003) – rests on the idea that what distinguishes expressivism is a distinctive explanatory thesis:

[T]o explain what it is to make a moral judgment, we need not mention any normative properties. So Hawthorne and Price's characterization of expressivism turns out to mesh with Fine's criterion for distinguishing irrealism, and also with Gibbard's understanding of what distinguishes his own sophisticated expressivism from realism. I'll call the integrated account of the distinction, the "explanation" explanation. (2004, 39)

I want to quote Dreier's concluding paragraph in full, in order to return to some of its themes below:

Crucial to maintaining the distinction, in meta-ethics, in the twenty-first century, between realism and irrealism is the possibility that concepts (and meanings) can differ in ways other than by their content. Or, if the difference between normative (or evaluative, or "planning") concepts and descriptive (naturalistic) ones can also be stated as a difference in content, then at least it must be a comprehensible, substantive question whether the difference in concept is explained by (or if you prefer amounts to no more than) a difference in content, on the one hand, or rather it is explained by (amounts to) something else entirely, which in turn explains the difference in content. The divide between realism and irrealism, at least in meta-ethics, rests on the substance of questions about metaphysical explanation.

Returning to these issues in (Dreier 2018), Dreier calls this the *explanationist* proposal.<sup>10</sup> It seems to me to be mainly right, but I want to flag some differences of emphasis. I also want to highlight an issue that will be relevant in chapter 11, when I explain what I take it to be to be a *global* expressivist.

Concerning terminology, I would quibble, first, with the term 'metaphysical' in Dreier's last sentence. In the disagreements in question, between expressivists and their rivals, expressivists need to be clear that they take the relevant *explananda* to be aspects of human linguistic and cognitive behaviour, rather than anything particularly metaphysical. In my own work I have often distinguished the expressivist's explanatory project *from* metaphysics by calling it anthropology, or psychology, and FFT's distinction between analysis and explanation clearly points in this direction.

As a first pass, then, expressivists differ from their rivals in simply being *interested in different questions*. In recent work, I have noted that Ramsey is one early figure who sees the importance of this distinction.

---

<sup>10</sup> He is now more sceptical that it draws any sort of sharp line.

[W]hen Ramsey comes this way he speaks of psychology, not anthropology. Here he is in ‘General Propositions and Causality’ (GPC, 1929), reflecting on a possible response to the account of causation he has just sketched – an account that we might now call expressivist, or pragmatist.

What we have said is, I think, a sufficient outline of the answers to the relevant problems of analysis, but it is apt to leave us muddled and unsatisfied as to what seems the main question—a question not of psychological analysis but of metaphysics which is ‘Is causation a reality or a fiction; and, if a fiction, is it useful or misleading, arbitrary or indispensable?’ (Ramsey 1929, 141)

Ramsey doesn't address this concern directly, but I think it is clear that his view is that metaphysics is the wrong mode of enquiry, in this case. The illuminating enquiry is what he calls ‘psychological analysis’ – an investigation into how we come to *think* and *talk* in causal terms, conducted in a manner that we do not presuppose that the helpful answer will lead us back to the objects. (In other words, we do not presuppose that the answer will be ‘We talk this way because we are keeping track of the causal facts’, or anything of that kind.) (Price 2017a, 151)

A theorist who does say that we talk this way because we are keeping track of the causal facts, or its equivalent in the ethical case, is someone who has acknowledged the expressivist's questions about language, but taken them to lead back to metaphysics. That's the possibility that Dreier has in mind when he says that in contrast, the expressivist's view is that ‘to explain what it is to make a moral judgment, we need not mention any normative properties’.

For the moment, what I want to add to Dreier's analysis is the expressivist's insistence that the linguistic or psychological questions are *where we start*. Metaphysics often starts somewhere else, such as with questions about the *nature* and *reality* of causation, moral properties, or whatever. (As Ramsey put it, ‘Is causation a reality or a fiction; and, if a fiction, is it useful or misleading, arbitrary or indispensable?’) As I said, the difference between expressivism and its rivals often shows up as a difference in which questions are taken to be *worth asking*. (FFT's distinction between *analysis* and *explanation* turns on this point.)

Returning to Dreier's terminology, I think it is easier to make these points if we say not that the differences mark a distinction between realism and irrealism, but that the characterisation of the issue in these metaphysical terms – a matter of realism *versus* irrealism – is a mistake. There is a difference between expressivism and its rivals, but the lesson of deflationism, in metaphysics as in semantics, is that these are simply the wrong vocabularies in which to describe the difference.

As we saw, Dreier himself presents Creeping Minimalism as a problem for the field, rather than particularly a problem for one side or the other. My own view, in contrast, is that there is an important sense in which the news is worse on the non-expressivist side – in fact, on the non-*global*-expressivist side. As I said above, I have argued that the effect of semantic minimalism is to suck the air out of concepts needed to formulate not just a blanket alternative to expressivism – realism everywhere, in Dreier's terms – but even a local version of this view (realism somewhere). In other words, I have taken the view not just that (as Dreier agrees) expressivism remains coherent in the light of semantic minimalism, but that no alternative to it does so. I'll say some more in a moment about a possible response to this claim, based on the distinction I drew earlier between semantic and metaphysical minimalism.

First, I want to note that there is still a tendency to see Creeping Minimalism as asymmetric in the other direction – i.e., as an objection to the expressivist side.<sup>11</sup> I think there are two aspects to this view of the situation. In part, it reflects the old argument that semantic minimalism defeats expressivism, to which Hawthorne and I were responding. Dreier agrees that that argument has been laid to rest, and his point is not to revive it.

The second factor contributing to this one-sided view of the problem of Creeping Minimalism again rests on the mistaken belief that expressivists need some distinction *of a metaphysical sort*, some sense in which they are still *irrealists* – whereas minimalism seems to lead to *realism*, albeit a minimal kind of realism. This thought is encouraged by Dreier’s own tendency to cling to a metaphysical way of framing the distinction he wants to save.

As I said, my view is that the expressivist needs to push back against this thought, insisting that she needs no such distinction. The expressivist should make it clear that she simply *rejects* the metaphysician’s realist/irrealist dialectic, rather than taking sides within it. As I noted in §0.3.8, Blackburn is often admirably clear on this point, as in this early passage.

What then is the mistake of describing such a philosophy [i.e., quasi-realism] as holding that ‘we talk as if there are necessities when really there are none’? It is the failure to notice that the quasi-realist need allow no sense to what follows the ‘as if’ except one in which it is true. And conversely he need allow no sense to the contrasting proposition in which it in turn is true. (Blackburn 1987, 57)

I have also noted elsewhere (e.g., Price 1997; 2009b, 324) that Carnap is a useful ally for an expressivist at this point. Here he is in ‘Empiricism, Semantics and Ontology’, describing and endorsing the view of the Vienna Circle:

Influenced by ideas of Ludwig Wittgenstein, the Circle rejected both the thesis of the reality of the external world and the thesis of its irrealism as pseudo-statements; the same was the case for both the thesis of the reality of universals (abstract entities, in our present terminology) and the nominalistic thesis that they are not real and that their alleged names are not names of anything but merely *flatus vocis*. ... It is therefore not correct to classify the members of the Vienna Circle as nominalists, as is sometimes done. (Carnap 1950, 215)

However, this offers a clue to the sense in which it may be too quick to take semantic minimalism to defeat Dreier’s realists *tout court*. A metaphysician who disagrees with Carnap – who allows substantial sense to the ‘external’ ontological questions that Carnap rejects – has a vocabulary to differentiate themselves from expressivists. Again, this differentiation is a matter of asking different questions – questions such as ‘Is causation a reality or a fiction’, to use Ramsey’s example again.

Notice that this possibility depends on *rejecting* deflationism on the metaphysical side. Is this compatible with semantic deflationism? On the face of it, yes, and many would cite Quine as an example. Quine’s insistence that to say that ‘Snow is white’ is true is simply another way of

---

<sup>11</sup> ‘Since James Dreier introduced it, the so-called *problem of creeping minimalism* has threatened metaethical expressivism’, as Simpson (2018, 510) puts it, for example. Similarly from Mathew Chrisman (2008, 345): ‘[A]s a form of expressivism, quasi-realism seems to be proposed as the best possible *irrealist* position. If it turns out to be indistinguishable from forms of realism, then the expressivist has again lost the debate.’

speaking about snow is certainly deflationary in a semantic sense.<sup>12</sup> So long as we combine this with the view that Quine is a champion of Ontology with a capital ‘O’,<sup>13</sup> then we have the combination of views we were looking for. In other words, we have something that professes to be Realism with a capital ‘R’, despite its endorsement of semantic minimalism.

Is this version actually incompatible with global expressivism? The global expressivist may, with Blackburn, claim to ‘make no sense’ of the capitals letters; and, with (Price 1992a, 2009b), dispute the reading of Quine on which they depend. But this hardly amounts to inconsistency. A real conflict only emerges if this self-proclaimed Quinean Ontologist, picking up the explanatory questions about language with which the expressivist begins, seeks to answer them in terms that the expressivist must disallow. This will be so if the terms concerned amount to an abandonment of semantic minimalism, but in this case the realist has taken themselves off Dreier’s map – it’s no surprise that one way to meet the challenge of Creeping Minimalism is to disavow minimalism!

This points to a dilemma for the realist, within the constraints of Dreier’s dialectic. The challenge is to find an alternative to expressivism that is compatible with semantic minimalism. Without such an alternative, Creeping Minimalism will entail global expressivism, in the way that I have proposed. But if the realist’s claims are compatible with semantic minimalism, what is to stop the expressivist from endorsing them? How could they mark something to which only the realist is entitled?<sup>14</sup>

Let’s explore this dilemma with reference to a version of explanationism recently defended by Matthew Simpson (2020). Simpson characterises explanationism as follows:

Explanationism, then, is the view that what distinguishes expressivism is how it completes claims like:

- (1) The English sentence ‘stealing is wrong’ means *stealing is wrong* in virtue of ...
- (2) Ella’s belief has the content *stealing is wrong* in virtue of ... (2020, 754)

Simpson calls Dreier’s own (2004) view *ontological* explanationism. As Dreier himself says, it takes the expressivist to claim that ‘we need not mention any normative properties’ in completing claims such as (1) and (2). In order to evade some difficulties raised by Chrisman (2008), Simpson proposes instead what he calls *subject matter* explanationism:

The subject matter view is that expressivism diverges from its rivals by not mentioning the subject matter of ethical language and thought when explaining it. More precisely it does not mention the subject matter on the right-hand side of claims like (1) and (2). (2020, 758)

---

<sup>12</sup> As Quine puts it: ‘By calling the sentence [“Snow is white”] true, we call snow white. The truth predicate is a device of disquotation.’ (Quine 1970, 12)

<sup>13</sup> ‘Ontology became a respectable subject for an analytic philosopher to pursue ... in 1948, when Quine published a famous paper titled “On What There Is.” It was Quine who single handedly made Ontology a respectable subject.’ (Putnam, 2004, 78–79) For my own criticism of this reading of Quine see (Price 1992a, 2009b).

<sup>14</sup> In (Price 2022d) I claim Crispin Wright as an ally against Dreier, in the sense that both Wright and I \see minimalism as imposing a greater burden on realism than on its rivals, for making a case that it has something distinctive to say. ‘All the onus, everywhere, is on the realist’, to repeat some words from Wright (1993) that I borrowed long ago in ‘Metaphysical pluralism’ (Price 1992a).

Now to our dilemma. Let's look at Simpson's examples of non-expressivist versions of (1) and (2).

The subject matter view will classify the following as representationalist [i.e., non-expressivist] accounts:

- (6) 'Stealing is wrong' means *stealing is wrong* because it represents/describes/pictures stealing as wrong.
- (7) 'Stealing is wrong' means *stealing is wrong* because it is true iff stealing is wrong.
- (8) '... is wrong' means ... *is wrong* because 'wrong' refers to/predicates the property wrongness.
- (9) '...is wrong' means ... *is wrong* because (the relevant) utterances of 'wrong' are (typically) caused by something being wrong.
- (10) Ella's belief has the content *stealing is wrong* because it involves a concept C whose teleological function is to track wrongness. (2020, 759, underlining added)

In claims (6), (7), and (8), however, the terms underlined are the kind of terms often taken to be swept up in the semantic minimalist's basket. But can they be read as minimalist in this context? If the realist says 'yes', the expressivist will object that she is entitled to say the same kind of thing. The realist may reply that the 'because' makes a crucial difference – the expressivist can indeed say

'Stealing is wrong' is true iff stealing is wrong

but this is not (7), which is a more substantial claim. But then the expressivist will object that her opponent has abandoned semantic minimalism, the point of which is that truth and similar notions *cannot* do substantial explanatory work. Hence the dilemma for realists (or Simpson's representationalists), trying to stake out a position both distinct from expressivism and compatible with semantic minimalism: if they invoke semantic notions at all, these will be either too lightweight to distinguish realism from expressivism, or too heavyweight to be compatible with minimalism.

However, Simpson's remaining examples, (9) and (10), suggest a way out. The notions of an utterance *being caused by some worldly circumstance*, or of a belief *having the teleological function to track some worldly circumstance*, do not seem to be the folk semantic notions deflated by semantic minimalism. On the contrary, they are notions that seem to have a respectable theoretical role. So long as realists rely on such notions, then, they can certainly avoid the charge that they are saying something incompatible with semantic minimalism. (Presumably Quine's own preferences would go in this direction, though they would take issue with the use of the notion of meaning in claims (9) and (10).)

This gets us to a useful point. If Creeping Minimalism is to leave a space for an alternative to expressivism, this alternative requires a theoretical vocabulary in which to discuss the relation between our words and their subject matter, *other than the vocabulary deflated by semantic minimalism*. This will connect to other themes, later in the chapter (§10.6 and §10.7). Among other things, I will raise the question whether the expressivist has any reason to disagree with claims couched in such a vocabulary.

This question connects to my next comment about Dreier's characterisation of expressivism. Concerning meta-ethical expressivism, Dreier's explanationism turns on the expressivist's claim

that ‘to explain what it is to make a moral judgment, we need not mention any normative properties’ (Dreier 2004, 39). But if this is what expressivism amounts to, how could we possibly be global expressivists? We need to mention *something* in an explanation, and the criterion doesn’t seem to allow us to be expressivists about that. This is what Blackburn (2013, 78), following Robert Kraut, has called the No Exit problem for global expressivism.

This suggests that while Deier’s criterion works perfectly well in the meta-ethical case, and probably in any case in which expressivism is motivated by the attractions of metaphysical economy, it doesn’t quite put its finger on the basic point. We global expressivists need something different. We need to point to a component in the expressivist’s treatment of the standard cases (ethics and probability, say), explain why that component counts as distinctively ‘expressivist’, and then show that that component never goes to zero – any assertoric vocabulary whatsoever needs to invoke some such component, for reasons recognisably related to the need in cases with which we began. More on this in Chapters 11 and 12.

So far I have been discussing Creeping Minimalism against the background of the kind of expressivism I defend in (Hawthorne and Price 1996) and later papers, but what about the distinctive features of FFT itself? How does it appear in the light of those? Three aspects of FFT are particularly relevant, I think: (i) the characterisation of expressivism in terms of the (varying) prevalence of NFD; (ii) the proposal that variation with respect to NFDs be explained in terms of differences in underlying function of the discourses concerned; and (iii), framing everything else, the characterisation of factuality as ‘downstream’ of the assertoric mode of speech. By the lights of FFT, all of these are available to mark respects in which expressivism differs from standard forms of realism. If a realist claimed that semantic minimalism enabled her to embrace these proposals, we expressivists would have no reason to shoo her away.

### 10.3 Relaxed realism

This brings us nicely to our next topic. Another manifestation of the influence of deflationary or minimalist ideas in contemporary philosophy is a cluster of views that Sarah McGrath has dubbed ‘relaxed realism’. As McGrath says, ‘relaxed realist themes are central to Dworkin’s *Justice for Hedgehogs* (2011), Parfit’s *On What Matters* (2011), and Scanlon’s *Being Realistic about Reasons* (2014)’ (McGrath 2014, 187). She says that she

call[s] this picture relaxed realism ... to capture the way in which its proponents combine a commitment to realism with a certain lack of anxiety about the status and standing of morality, despite understanding morality in ways that might naturally encourage such anxiety’ (2014, 187).

Relaxed realism is not a distinctively twenty-first century position. We can find similar views earlier, both in the normative case and in others: for example, McDowell’s ‘re-enchanted naturalism’ (1994); John Campbell’s ‘simple realism’ about colour (1993); and the kind of ‘neo-Fregean platonism’ associated with Bob Hale and Crispin Wright (2001). We might also include Strawson’s ‘catholic or liberal naturalism’ (Strawson 1985, 1), the latter term being used in the present century by writers such as Macarthur and de Caro (2010).<sup>15</sup> We might also include Arthur Fine’s Natural Ontological Attitude (NOA), already discussed in Chapter 9.

---

<sup>15</sup> Indeed, Snowdon (2001, 335) already refers to Strawson as a relaxed realist: ‘As an ontologist, therefore, we might call Strawson’s attitude “relaxed realism.” He endorses the reality of ... entities and their properties – hence, the realism – without supposing that this requires some strong unification between the different levels of thing or property – hence the relaxedness.’

Common to these views is the rejection of a certain sort of metaphysical stance (the one that encourages anxiety, as McGrath puts it). A familiar term for this rejection, used by some of these authors themselves, is metaphysical ‘quietism’. Often attributed to Wittgenstein, this kind of quietism is characterised by McDowell as the rejection of a sideways metaphysical perspective on our practices.

As I have noted, FFT agrees with these views about the attractions of *metaphysical* quietism – of a deflationary approach to metaphysical issues. Where it disagrees is in insisting on the interest and respectability of another project – the functional and genealogical project. Concerning McDowell, my own strategy (Price 2015a) has been to present him with a dilemma. Either he has to be more quietist than even he wants to be, in being unable to explain the sense in which (in his words), ‘[v]alues are not brutally there—not there independently of our sensibility—any more than colours are’ (1998, 146). Or he has to endorse what is in effect an expressivist genealogy – a ‘sideways’ explanation of how our value and colour judgements come to depend on aspects of our sensibility (different aspects, in each case).

Applying Dreier’s explanatory test, we might say that relaxed realists face a trilemma. Faced with what seem to be legitimate questions about particular discourses – why we have them, how they differ, how they relate to our sensibilities – there are three main options. In the metaphysical corner are views that appeal to the nature of the properties or entities in question (e.g., again, colours and values) to answer such questions. In the extreme quietist corner are views that simply fail to engage with such questions. And in the third corner is expressivism. The first corner seems off limits for anything worth calling relaxed realism. But that leaves a choice between an excessive quietism and expressivism itself.<sup>16</sup>

#### 10.4 The Frege-Geach argument (I): impacts of minimalism

I have described my response to the claim that minimalism about truth creates difficulties for expressivism, by making it easy to be truth-apt. As I explained, my view has long been that on the contrary, semantic minimalism counts in favour of expressivism – at any rate, of *global* expressivism – by threatening to deprive its opponents of the theoretical vocabulary in which to formulate rival accounts.

I have also long felt that semantic minimalism provides a response to another prominent objection to expressivism, the so-called Frege-Geach argument. Yet this objection remains in centre stage, taken very seriously both by expressivists and their opponents. From my perspective this looks puzzling. As I say, I had thought that it, too, would be a victim of semantic minimalism. More precisely, I felt that the notion that there is a distinctive problem here for expressivists would be a victim of semantic minimalism – more on this in a moment.

In a different sense, I am also puzzled why I didn’t mention the Frege-Geach argument at all in FFT. The answer may be that my views were in transition. I do respond to the Frege-Geach argument in my PhD thesis (Price 1981), and some of that material found its way into (Price 1994), parts of which were written in the early 1980s (Price 1983d). This material is somewhat in the spirit of Blackburn’s responses to the argument, though critical of them in some ways. But the (1994) paper also proposes a simpler response, drawing now on semantic minimalism. This material post-dates FFT, reflecting work on semantic minimalism by writers such as Horwich and Wright in the early 1990s. As I’ll explain, the key move was already in place – I had made it

---

<sup>16</sup> For similar criticism of relaxed realism, see (Ridge 2019) and (Böddeling 2019).

for a different reason in one of my early papers – but the interest in minimalism gave me a framework to develop it in the form needed for the Frege-Geach problem in general.

I'll introduce the Frege-Geach argument by reproducing Dummett's presentation of Frege's original version of it (Frege 1960), where it is an objection to the proposal that sentential negation be treated as a force-modifier indicating a distinctive kind of speech act, different from assertion. I quote the following passage from Dummett in (Price 1983b, 171–172).

It is natural at first to take denial as an activity parallel to assertion. ... The sign of negation 'not' would thus be interpreted as a sign of denial, parallel to the sign of assertion, and conveying, not assertoric, but negatory force. Frege opposes this view with a conclusive argument: namely that, while it would be possible so to interpret those sentences in which the sign of negation was the principal operator, it would be impossible to interpret the negation sign in this way whenever it occurred in a sentence otherwise than as the main operator. ... [A] sentence formed by prefacing another sentence by the sign of negation may itself serve as a constituent in a complex sentence, as, say, a clause in a disjunctive or conditional sentence. A sign which conveys force must relate to the whole sentence in which it occurs; it cannot form part of a subordinate clause. Hence the negation sign which does occur as part of subordinate clauses must be interpreted differently, namely as a functional expression (sentential operator) which contributes to the sense of the clause in which it occurs. (Dummett 1973, 316–317)

As noted in (§0.1), I respond to Dummett's appeal to Frege's argument by proposing that the force–sense parsing of speech acts might be non-unique. An utterance might be properly treated both as a denial *that P* and as an assertion *that not-P* (compatibly with denial being understood, initially, to be a different speech act from assertion).

I returned to this idea in (Price 1994), now linking the point to a deflationary attitude associated with minimalism:

[T]he upshot is that we do not have to make a choice between the view that negation indicates denial (or expresses disbelief) and the view that it indicates an assertion with negative content (or expresses such a belief). We may say both things, so long as we are dealing with a suitably minimal notion of belief. Disbelief or dissent comes first, for it is such a notion, cashed in functional terms, that accounts for the presence and utility of negation in the first place. But given that this expression of disbelief takes the minimal assertoric form, we may think of it as the expression of belief.

Putting it in Fregean terms, we might say that the sense–force boundary is not unique—we may have two (or more) ways to parse a given utterance. “Not-P” may be thought of both as a denial of the proposition that P and as an assertion of the proposition that not-P. The pragmatic account of the function of denial is not a separate component of a theory of meaning from the theory of sense, but a sub-theory, whose task is to explain how there come to be sentences with senses of a particular sort—how there come to be sentences with the sense of negative judgements, for example. (1994, 72)

Reading this passage now, I am struck by how much the stance reflects what Brandom (still several years in my future, even in 1994) was to call ‘the pragmatist direction of explanation’:

The pragmatist direction of explanation ... seeks to explain how the use of linguistic expressions, or the functional role of intentional states, confers conceptual content on them. (Brandom, 2000, 4)

Starting with an account of what one is doing in making a claim, it seeks to elaborate from it an account of what is said, the content or proposition—something that can be thought of in terms of truth conditions—to which one commits oneself by making a speech act. (2000, 12)

We saw that I had this idea about negation in (Price 1983b), so I take its presence in (Price 1994) to date from early versions of the paper, a decade earlier. As noted in §0.3.4, I had made a similar point in the debate about desire-as-belief in (Price 1989). In response to (Smith 1987) I point out that a belief that it is good that *P* can be held to have mind-to-world direction of fit with respect to the content *It is good that P*, and world-to-mind direction of fit with respect to the embedded content *P*, with no contradiction – again, the point turns on the possibility of multiple parsings of the same utterance or attitude.

The application of this idea to negation, and its relevance to the Frege argument, is also explicit in (Price 1990), a piece on the functions and possible genealogy of negation:

[I]f we allow that (an utterance of)  $\sim P$  may properly be regarded *both* as a denial with content *P* *and* as an assertion with content  $\sim P$ , then Frege's argument is powerless; for in this case the latter reading is available to explain the contribution of  $\sim P$  to complex constructions, in the standard way. The difficult task is to defend a view of assertoric discourse and its limits that permits this kind of multiple factorization and it is this task I am here setting aside.<sup>17</sup> (1990, 225)

The newer claim in (Price 1994), apparently added later – after, as I say in the paper, I had learnt the term ‘minimalism’ from Wright and from Horwich – was that this kind of multiple parsing had much more general application, and was the key to the simple general response to the Frege-Geach argument that semantic minimalism provides. The passage from (Price 1994) above continues like this:

The two parsings engage with two different aspects of a theory of meaning. The platitude that to know the meaning of a sentence is to know the conditions for its correct use has two importantly different readings (not always properly distinguished). It may be taken to refer to what I ... called subjective assertibility conditions, so that it amounts to the claim

- (4) To know the meaning of “It is snowing” is to know that it is normally appropriate to say “It is snowing” only when one believes that it is snowing.

Or it may refer to truth conditions, so that the claim is that it is correct to say “It is snowing” if and only if it is snowing; or, in the more familiar form, that

- (5) “It is snowing” is true iff it is snowing.

These claims are not incompatible, of course, and knowledge of meaning surely involves knowledge of both kinds. (5) has the form we expect of the theorems of a content-specifying truth theory—a systematic specification of the meanings of the sentences of an object language by means of sentences in the home language. As has often been

---

<sup>17</sup> I note at this point that I defend such a view of assertoric discourse in FFT.

emphasised, this enterprise needs only a thin notion of truth. It therefore applies uniformly to all minimally descriptive parts of the object language [i.e., as I put it elsewhere in the paper, any sentence ‘which is capable of being minimally true or false’].

Principle (4) on the other hand has the resources to cope with the functional perspective, which is crucial to [my] proposed reformulation of non-factualism [i.e., a formulation that stresses the expressivist’s positive story in terms of distinctive functional roles, rather than a negative story couched in semantic terms]. We may say for example that

- (6) To know the correct use of “Not-*P*” is to know that it is normally appropriate to say “Not-*P*” only when one *disbelieves* that *P*.

Note that this is not incompatible with the following instance of (4):

- (7) To know the meaning of “Not-*P*” is to know that it is normally appropriate to say “Not-*P*” only when one believes that not-*P*.

Interpreted in terms of minimal belief, (4) is true of all minimal descriptions. But it is (6) which captures what is distinctive about utterances of the form *Not-P*.

In summary, then, the proposed reformulation of non-factualism encounters no special problems with respect to the goals of a theory of meaning. To the extent that such goals are met by a content-specifying truth theory, the reformulated view coincides with the standard [i.e., factulist] account; while to the extent that such a truth theory needs to be supplemented by theses of the form of (4), this form is flexible enough to accommodate the functional perspective adopted by the reformulated view. (1994, 73)

My thought was that in general, as for negation, this gave expressivists an easy response to the Frege-Geach problem. They could have all the advantages of a Tarski-style truth-conditional theory of meaning, while retaining their positive, functional account of *how there come to be* claims with truth-conditions, or contents, of particular categories. In effect, these two different kinds of theories engage with the two different senses of the truism that ‘to know the meaning of a sentence is to know the conditions for its correct use’, as I put it at the beginning of this passage. I’ll come back below (§11.2) to this important distinction, which I had been emphasising since my thesis – it is crucial to separating the view that probability claims *express* credences from the view that they *report* credences. (In Price 1983b I take Dummett to task for paying it insufficient attention.)

Other writers at that time also argued that semantic minimalism makes short work of the Frege-Geach argument; see particularly (Horwich 1994) and (Stoljar 1993). Like mine, these claims do not seem to have found much traction, even among writers most sympathetic to expressivism. An influential critic, once again, was Jamie Dreier (1996). Dreier responds directly to Horwich and Stoljar, arguing that ‘Truth Minimalism is of no help to a Norm Expressivist in giving an explanation of truth-functional embeddings of normative sentences’ (1996, 49). He proposes that ‘the best way to understand the Embedding Problem [i.e., the Frege-Geach argument] is as a burden of explanation laid on the shoulders of Expressivists’ – effectively, the burden of explaining *why* the functional stories should find appropriate expressions in the truth-conditional form that permits embeddings. He illustrates the point by imagining that we decide to use the expression ‘Bob is hiyo’ as a way of greeting Bob. As he points out, the fact that this has subject–predicate form is no help at all in making sense of what it would be to negate it, or embed it in other complex constructions.

Dreier's point about explanatory burden seems to me to be entirely fair. I had made a similar point myself in (Price 1994), though with a twist. After offering a response to the Frege-Geach argument that descended from my thesis, somewhat in the spirit of Blackburn's, I proposed a simpler and much more direct response, appealing to semantic minimalism:

Viewed in the light of our appeal to minimal semantics in answer to the Frege point, Blackburn's approach [to the Frege-Geach argument] may ... seem unnecessarily complicated. In defence of Blackburn, however, it should be conceded that the quasi-realist program embodies an insight which is in danger of getting lost in the rush to embrace minimal semantics. Not all linguistic functions are such as to be usefully cast in terms of truth and falsity, however minimally these are conceived. Someone who wants to be pluralist about underlying linguistic functions thus owes us an account of what the truth-bearing form achieves in language, and hence an explanation, case-by-case, as to why various disparate functions should invoke it. ...

This may sound like the difficult case-by-case work we tried to avoid in invoking minimal semantics. Have we therefore advanced at all ...? It seems to me that from the non-factualist's point of view the situation has improved in one crucial respect: the question as to the general function of the truth-bearing form of language has now been raised as an issue that all sides need to address. Previously, in couching their views in semantic terms, the non-factualists effectively conceded to their opponent the latter's right to an unexamined notion of the genuinely factual (or truth-bearing) use of language. The explanatory onus thus lay almost entirely on the non-factualists' side. The new approach distributes the burden much more fairly. True, the non-factualists' opponents may not have noticed that there is a general question to be raised concerning the rôle of truth-bearing constructions in language; but this is hardly a point against non-factualism. ... The appeal to semantic minimalism does not evade this complexity, for the difficult explanatory issues remain; but it does ensure that the burden is properly spread, and that all sides take their fair share of the load. (Price 1994, 75)

## 10.5 The Frege-Geach argument (II): explanatory burdens

I want to illustrate this point about sharing the explanatory burden with reference to some recent work. Much recent criticism of expressivism has turned on its difficulties with explaining the simplest kind of sentential embedding, that of negation. Mark Schroeder (2008) offers a particularly clear account of this problem, drawing on the work of Unwin (1999). This is how Schroeder characterises the difference between the task facing expressivists and their descriptivist rivals:

An ordinary descriptivist account of why 'grass is green' and 'grass is not green' are inconsistent would point out that 'grass is green' can be true only if grass is green, and that 'grass is not green' can be true only if grass is not green, but that grass cannot be both green and not green: the inconsistency of their contents guarantees the inconsistency of the sentences. An explanation of why they are *logically* inconsistent would point out that given their structure, this feature would hold no matter what interpretation we assigned to 'grass' and no matter what interpretation we assigned to 'green'. So their inconsistency is guaranteed by the logic of 'not'.

Expressivists cannot appeal to such an explanation, because they don't want to tell us what it would be for murder to be wrong. The only resource that expressivists have to

appeal to, in order to explain the inconsistency of 'murder is wrong' and 'murder is not wrong', is the mental states that are expressed by each. So an expressivist account of their inconsistency is going to have to work by appealing to some 'inconsistent' feature of the attitudes that each expresses. And this is, in fact, the shape of all expressivist attempts to explain inconsistency, entailment, and validity. (Schroeder, 2008, 39)

To me, however, this contrast relies on letting the descriptivist off an important hook, far too easily. For a descriptivist, too, there is an important question as to why the kind of commitments we express by saying 'Grass is green' and 'Grass is not green' are incompatible. We can see this by beginning with a case in which the incompatibility has not yet been made explicit in our choice of vocabulary, and noting that there is a question about what makes that vocabulary – i.e., the use of negation to *mark* an incompatibility – an appropriate choice.

I can explain this point with an example from my paper 'Why "Not"?' (Price 1990), where my project was precisely to ask what negation is *for*. The example is intended to demonstrate the utility of an explicit linguistic marker of disagreement, based on perceived incompatibility. I imagine a linguistic community – 'Ideological Positivists' – who have banished negation from their language in an attempt to encourage positive thinking. I ask my reader to imagine that they and I are Ideological Positivists ourselves, and that we share a goal: we both want to see Fred. I think that Fred is in the kitchen, and the reader thinks Fred is in the garden.

If we both want to see Fred, your belief will lead you to the garden, mine me to the kitchen. Fred cannot be in both places; so one of us, at most, will satisfy our common desire. How much better things might have been, if we had only noticed that our beliefs were incompatible, and hence discussed Fred's whereabouts, before we set out to look for him.

It is true that even as Ideological Positivists we might realize that 'Fred is in the kitchen' and 'Fred is in the garden' are incompatible, and hence reconsider. It is also true that in the real world we might miss the incompatibility, by missing the inference from 'Fred is in the garden' to 'Fred is not in the kitchen'. But the advantage of a sign of denial is that it gives us a perfectly general means of registering and pointing out the incompatibility. Think how it might go for us as Positivists:

Me: 'Fred is in the kitchen.' (Sets off for kitchen.)

You: 'Wait! Fred is in the garden.'

Me: 'I see. But he is in the kitchen, so I'll go there.' (Sets off.)

You: 'You lack understanding. The kitchen is Fred-free.'

Me: 'Is it really? But Fred's in it, and that's the important thing.' (Leaves for kitchen.)

Your problem is to get me to appreciate that your claims are incompatible with mine. Even in such a trivial case, we can see that it would be useful to have a device whose function was precisely to indicate that an incompatible claim was being made: precisely to deny an assertion or suggestion by somebody else. It seems that this is what negation gives us. (1990, 224)

For present purposes, it doesn't matter how plausible this is as an account of the function and genealogy of negation. The important thing is what it illustrates about what Schroeder's ordinary descriptivist is missing, in helping themselves to the account he describes of why 'grass is green' and 'grass is not green' are inconsistent – an account that simply appeals to

the logic of negation. At one level, it misses the task of explaining why claims such as ‘Fred is in the kitchen’ and ‘Fred is in the garden’ are incompatible (a question we can express, once we have negation, as that of explaining why each implies the negation of the other). At a deeper level, it misses pretty much every question we might think of about how creatures like us come to avail ourselves of propositions and logic, in the sense that the ordinary descriptivist is taking for granted. (Somewhere in between, it is missing my question about the point of negation itself.)

In other words, it seems to me that factualists as well as expressivists have an explanatory mountain to climb, before they are entitled to help themselves to logical machinery, in the move that Schroeder allows to his ordinary descriptivist. Once they have climbed it, then by all means let them appeal to the logic of negation – but that’s true as much for the expressivist as for the factualist, in my view (if indeed there is any surviving distinction between the two).

This is the even-handed picture I was recommending in (Price 1994). I took it that I had made a start on climbing the explanatory mountain, both in the proposal presented in FFT, Part II, and in my papers on negation, especially (Price 1990). I was committed to the view that the right way to climb it was by a single path, with no need for separate trajectories for descriptivists and expressivists. But I had no detailed route to offer, beyond the beginnings just mentioned, and I may well have underestimated the difficulties ahead. In a moment I’ll describe some interesting recent work by Joshua Gert. It makes a very promising assault on a crucial face of the mountain, starting from a similar position. First, a little more about responses to Dreier.

As we have seen, my approach associates the point of truth and falsity, and of negation itself, with disagreement, and links the point of disagreement itself to the question as to whether different speakers are ‘in the same boat’ – in other words, whether, for the particular kind of commitment or speech act in question, there is any value in alignment, across a speech community.

Horwich applies something like this thought to Dreier’s ‘hiyo’ example.

I am sympathetic to the following suggestion of Allan Gibbard's [Gibbard, 2003, 65–75]. When someone genuinely accepts something and someone else accepts its negation there is a sense of substantive disagreement; there is conflict, a clash, a feeling that the other person is somehow in bad shape. But the "hiyo" rule would not give rise to these phenomena. I, while accosting Bob, think and say "Bob is hiyo"; you, with no intention or prospect of accosting him think "Bob is not hiyo"; but beneath these words there is no real tension. Nor does my pronouncement go hand in hand with a belief that you would be well advised to think as I do. (Horwich 2005. 86)

In other words, in my terms, greeting someone is not normally the kind of speech act for which there is any value in pressure to alignment, across a speech community. But it is not hard to imagine how it might become so. Imagine that greeting becomes a highly ritualised matter, subject to many norms and taboos. It’s crucial to get it right, and to greet someone when and only when it is appropriate. If, further, this ‘rightness’ is constant across our particular speech community, then it is not hard to see how the kind of conflict that Horwich speaks about here would creep in – and how the means to express it, in such a way as to call attention to disagreements and expose our ‘greeting commitments’ to the rigours of the game of giving and asking for reasons, would prove useful. (The obvious place to look for analogies would be in expressions of matters of taste, which are on the assertion spectrum, but towards the relaxed end of it – No Fault Disagreements are easily tolerated.) In this way, Dreier’s ‘hiyo’ example is smoothly accommodated by the framework of FFT.

What does all this mean for the Frege-Geach argument? I think that my view that semantic minimalism assists expressivism in dealing with it is in pretty good shape, though it's important to note that much work remains to be done. The crucial thing about semantic minimalism is that it allows us to regard much of the machinery of truth-conditionality, and the like, as common property, equally accessible to expressivists and descriptivists (though far from free from explanatory commitments for either – on the contrary, the same mountain looms over both).

This is how I concluded in (Price 1994):

[T]he most central issue to which the above considerations direct our attention seems to me to be this one: *What does assertoric discourse do for us?* It is possible to distinguish a number of sub-issues here. What are the concepts of truth and falsity for? (What function do they serve in the lives of a linguistic community?) What is the significance of the linguistic constructions that apparently depend on truth – conditionals, for example? Again, how does it help us to have them? And is there a category of genuine judgements, as opposed to commitments more generally? (1994, 76)

I noted that it is the last issue that distinguishes me from Blackburn:

In Blackburn's terms, I am someone who extends the quasi-realist project "all the way down." Blackburn's difficulty seems to me to be that non-global quasi-realism is in danger of being self-refuting: the better the quasi-realist does locally, the less reason there will be not to "go global". (1994, 76, n. 12)

'[T]he great theoretical significance of the Frege point', I say, 'is that it directs our attention to these long neglected issues.' (1994, 77) That still seems to me a fair assessment.

In many quarters, these issues have remained neglected. Much of the progress in the field has been in clarifying the challenges faced by expressivism, rather than in tackling the deeper, more even-handed, issues that I had in mind here. However, I want to conclude this section by describing some recent proposals by Joshua Gert (2023). These proposals are explicitly even-handed, in the present sense. If they work, they take us further up the looming explanatory mountain than anything else that I know.<sup>18</sup>

Gert approaches these topics via Schroeder's (2008) discussion of the challenges that expressivists face in making sense of negation, and via Schroeder's proposals – ultimately unsuccessful, in Schroeder's own assessment – for dealing with these issues in an expressivist semantics. Gert modifies Schroeder's proposal in a way that avoids Schroeder's reliance on, as Gert puts it, 'a heavyweight, representational notion of proposition' (2023, 108):

In developing my account, I will initially be making use of a strategy pioneered by Mark Schroeder in his tour-de-force, *Being For*. Although Schroeder himself regarded his effort as falling short – indeed, as something very much like a *reductio* of the expressivist program – I will argue that it is possible to travel with him only halfway, and then to move in a simpler direction. In particular, it is possible to part ways with Schroeder after he makes his main theoretical innovation – the introduction of the attitude he calls 'being for' – but before he uses it to develop what he calls 'biforcated attitude semantics'. It is at

---

<sup>18</sup> Though see also (Incurvati and Schlöder 2023).

the latter point that he begins to rely, implicitly, on a heavyweight, representational notion of proposition: ... I will suggest that it is Schroeder's unquestioning acceptance of a representational notion of proposition that causes him to take a path that leads to technical problems that my view avoids. (2023, 108)

Gert's alternative to Schroeder's notion of proposition builds, as he says, on an FFT-like proposal about the function of disagreement:

Briefly put, the idea is that for some classes of mental states it is useful if members of a linguistic community can apply pressure to each other's tokenings of them – often because this is a way to get them to converge on some optimal state within the relevant class. (2023, 114)

Gert argues that this approach 'yields an independently motivated account of an attitude that turns out to function much like Schroeder's being for' (2023, 108). He also argues that it has implications

related to recent "bilateralist" proposals according to which denial is not to be explained in terms of assertion and negation, but, on the contrary, assertion and denial are used to explain negation. (2023, 108)<sup>19</sup>

'[I]n a parallel fashion', Gert continues, 'any given predicate [equally, say, 'wrong', or 'green'] will be associated with *two* distinct attitudes: one associated with assertion, and another with denial' (2023, 108) With these two 'paired attitudes' already in play, Gert is able to appeal to them to yield 'a bilateralist semantics for negation' – a 'toggling' account of negation, as he also calls it. He proceeds to argue that such an account of negation solves what's left of Schroeder's embedding problem, when the heavyweight notion of proposition is replaced by the notion that emerges from the dialogue-based account of assertion and denial.

In a little more detail, Gert argues that the dialogue-based semantics 'offer[s] a representation-free account of belief, whether descriptive, normative, or something else' (2023, 108). This account does give rise to an analogue of the embedding problem for negation discussed by Schroeder; Gert calls it 'the predicate negation problem'. But he shows that this version of the negation problem is solved by the toggling account of negation. And he goes on to argue 'that the predicate negation problem does not generalize to other logical operators', and to propose 'semantics for conjunction, disjunction, and the existential and universal quantifiers' (2023, 108).

Summing up, Gert lists the three main innovations of his proposal:

(i) combining Schroeder's main innovation – the notion of being for – with a "toggling" semantics for negation similar to bilateralist proposals, (ii) constructing a semantics for non-normative language that parallels Schroeder's semantics for normative language more closely than Schroeder's itself does, and (iii) offering an attitudinal semantics for the logical connectives that takes advantage of the toggling account of negation. (2023, 133)

He concludes:

---

<sup>19</sup> Gert notes antecedents of this bilateralist view both in my work and in Ramsey (1927).

Neopragmatists can explain the applicability of first-order logic to assertoric discourse without having to rely on contested and metaphysically heavyweight notions of representation, proposition, belief, or property. (2023, 133)

I hope that this outline makes it clear what progress Gert's proposals represent, by the lights of the sketch of the explanatory task I had in mind in (Price 1994).

## 10.6 The new bifurcation thesis

I criticised Schroeder (2008) for letting descriptivists off an explanatory hook too easily, but his perceptive discussion of the options for expressivism is in many ways highly congenial, from my point of view. Schroeder, too, is interested in the implications of expressivism for broader issues – among them, the nature of belief. In this context, he makes a suggestion that seems very much in the same spirit of one of my own proposals in (Price 1994) and more recent work – a move that I have referred to recently as a ‘new bifurcation thesis’ (Price et al 2013, 35). In Schroeder's version as well as in mine, the proposal is motivated in part by the Frege-Geach point.

In (Price 1994), as in FFT, I offer a two-level picture. At the top level, assertoric language is held to have an interesting global function – roughly, that of encouraging useful convergence between the commitments of the members of a speech community. This sits on top of a lot of lower-level functional diversity among the commitments themselves (including, of course, the kind of differences of function to which expressivists typically appeal).

Against this background, I suggested that we could usefully distinguish two notions of ‘description’ (not the best term, perhaps). The first was an inclusive, top-level notion, encompassing the entire domain, characterised in terms of something like aptness for deflationary truth. While the second was a lower-level functional notion, applicable only to a subclass within the domain:

I shall use the term “minimal description” for any utterance which is capable of being minimally true or false. The suggestion is thus that within the class of minimal descriptions, we may find sub-classes of utterances serving a range of different linguistic functions. ... Let us now suppose that one of the functions served by some minimal descriptions is ... to signal the presence of certain conditions in the physical environment of a speaker ... . [W]e thus have a distinction between the semantic ... notion of minimal description, and the functional notion of natural description (or physical signalling). (Price 1994, 67–8)

I then went on to propose that traditional expressivist claims could be charitably reconstructed within this framework:

My suggestion is then that the non-factualists' central thesis may be thought of as the claim that in certain cases we systematically confuse minimal descriptions for natural descriptions. Moral judgements (or whatever) are minimal descriptions, but are not natural descriptions. (1994, 68)

However, I pointed out that the difference is that traditional expressivists normally drew this distinction in semantic terms, whereas here it is explicitly functional, ‘the semantics on both sides of the distinction being agreed to be of the minimal sort.’ (1994, 68)

In recent work (Price 2011a, Price et al 2013) I have used different terminology for a similar distinction. I have proposed that we distinguish two ‘nodes’ in contemporary theory about representation, in philosophy, cognitive science, and related fields. Each of these nodes is itself a cluster of notions, at least in the sense that the defining features of the node may themselves be developed in a number of different ways. Bracketing the latter kind of diversity, my proposal is that we see the nodes themselves as distinct notions, rather than different aspects of the same single concept of representation.

The first node – *e-representation*, as I call it – involves the world-tracking conception of representation, associated, in biological cases, with the idea that the function of an evolved representation is to co-vary with some (typically) *external* environmental condition. (In some cases the relevant piece of the world is within the skin, as for pain or thirst.) There are familiar non-biological examples, of course: the needle in the fuel gauge and the level of fuel in the tank, the barometer reading and air pressure, and so on. What unites such cases, biological and non-biological, is that some feature of the representing system either does or (in some sense) is intended to vary in parallel with some feature of the represented system.

Thus e-representation emphasises the *system–world* links. The second node – *i-representation* – gives priority to *internal* connections between one representation and another. By this criterion, a token counts as a representation in virtue of its position in some sort of cognitive, inferential or functional architecture – its links, within a network, to other items of the same general kind. The notion is flexible enough to accommodate several different kinds of network: causal–functional, inferential, even computational.<sup>20</sup> The important thing is that the notion be divorced from any external notion of representation, thought of as a system–world relation.

These two notions of representation may have quite different uses, for various theoretical purposes. In particular, by the time we get to human language (from simpler forms of biological representation, in either sense), there may be no useful external notion of a semantic kind – no useful general notion of a relationship that words and sentences bear to the external world, that we might identify with truth and reference. This is a conclusion that a semantic minimalist has already come to from another direction, of course. The impression that there are such external relations can be regarded as a kind of trick of language – a misunderstanding of the nature of the disquotational platitudes. But we can hold this without rejecting the internal notion – without thinking that there is no interesting sense in which mental and linguistic representation are to be characterised and identified in terms of their roles in networks of various kinds.

This distinction allows us to draw a bifurcation in e-representational terms, while preserving homogeneity in i-representational terms. We can say that all declarative utterances are i-representations, but that only a subset of them are also e-representations. As I have pointed out, this captures something of the intuitions that originally motivated local expressivists such as Blackburn, while remaining compatible with what I term global expressivism, or global pragmatism. In my preferred way of developing things, we again have global pragmatism at two levels: at the top level, a Brandomian inferentialist story about assertoric discourse as a whole; and at the lower level, a range of pragmatic stories about the distinctive functions of particular discourses, *e-representation being just one pragmatic function among others*.

Now to Schroeder. Where I distinguished two notions of *description*, or more recently two notions of *representation*, Schroeder distinguishes two notions of *proposition*. Or rather, he distinguishes two

---

<sup>20</sup> It might also be social, and so not internal to any single speaker. The sense of ‘internal’ that matters involves looking to the network, not to the world. (I am grateful to Josh Gert here.)

roles that propositions are commonly thought to play in philosophy, the first as ‘the entities which are the objects of [‘propositional’] attitudes and the bearers of truth and falsity’, and the second entities that ‘mark out distinctions in reality, between ways that the world might be’ (2015, 88). Schroeder adopts the policy of reserving the term ‘proposition’ for whatever plays the former role, and using ‘representational content’, ‘for the entities, whatever they are, which mark out distinctions in reality, and are associated with metaphysical commitment’.

Terminological differences aside, this proposal seems to me to be interestingly analogous to mine. It allows Schroeder to grant to expressivists the right to genuine moral beliefs, beliefs in the truth of genuine moral propositions (though not, of course, moral beliefs with representational contents – only genuine descriptive beliefs have those). In the same way, my (1994) version allows expressivists minimal descriptions but not natural descriptions. And, as I just noted, my more recent version allows a Blackburnian expressivist to say that moral claims and beliefs are i-representational but not e-representational.

Schroeder notes that his view implies that although ‘paradigmatic beliefs, such as the belief that grass is green [are] simultaneously relations to propositions and to representational contents’ (2015, 89), in other cases the two things can come apart. There may be no representational content whatsoever, or there may be an embedded representational content, playing a different role from the main propositional content. If ‘P’ is a regular descriptive sentence, for example, then the believe that it is good that P involves a relation to the proposition that it is good that P, and a different relation to the embedded representational content P. Again, this seems to me strikingly analogous to the move I made in the 1980s for negation and for desire as belief, of recognizing that meanings may need to be parsed in more than one way.

There are some differences between Schroeder’s viewpoint and mine, of course, despite these structural similarities. More on these differences in a moment. Before that I want to note another point of analogy, and introduce another comparable view, that of Wilfrid Sellars.

### 10.7 Escaping the minimalist’s net

Schroeder notes that his notion of representational content is a theoretician’s notion, not a notion in pre-theoretical use. As a result, it is immune to the deflationary pressure on notions such as truth and belief – pressure that renders them fully applicable even in the vocabularies the expressivist wants to treat as non-descriptive. Here Schroeder explains the point with respect to an expressivism about epistemic modals and conditionals:

[S]aying that ‘Jack might be in Buellton’ does not express a representational content is *very different* from saying that it cannot be true or false, that it is not something that you can believe, or that it does not express a proposition. And that is because, since the notions of truth, falsity, belief, and propositions are all either *pre-theoretical* notions or are (in the case of ‘proposition’) under direct pre-theoretical *constraints*, there are pre-theoretical constraints which require us treat them as applying even to sentences involving, for example, conditionals, epistemic modals, and the words ‘true’ and ‘false’. But in contrast, the notion of a ‘representational content’ is a *theoretical* notion—motivated by the very considerations which motivate expressivism about epistemic modals and conditionals. (2015, 132)

This point, too, resonates with my view of the landscape. It also resonates with what I take to have been Wilfrid Sellars’ view. In (Price et al 2013, Ch. 8, §5) I discuss Sellars’ long engagement with the factual/non-factual distinction and related issues. In particular, I relate my distinction

between *i*-representation to *e*-representation to Sellars' distinction between two notions of truth. On one hand, in Sellars' view, there is a 'generic' notion of truth, defined in terms of what Sellars calls *semantic assertibility* ("S-assertibility"). This is the notion we find in ordinary use, and it is applicable, as Jim O'Shea puts it, to 'all kinds of propositions, whether they are empirical, mathematical, or moral claims' (O'Shea 2007, 144). Sellars' account of it is expressivist and deflationary, in contemporary terms. On the other hand is a notion that Sellars characterises in terms of 'picturing', and relates to the representations animals possess of their environments. This notion is much narrower in scope. Sellars takes it to be applicable only to what he calls 'matter-of-factual' language. Sellars stresses that the two notions are quite distinct: 'Picturing is a complex matter-of-factual relation and, as such, belongs in quite a different box from the concepts of denotation and truth.' (Sellars 1968: 136). He draws similar distinctions with respect to related notions, including that of 'fact' itself. For these, too, he distinguishes their ordinary use from a narrower theoretical use.

This is my diagnosis, and my attempt to relate Sellars' discussion to my own view:

What is happening ... is that a cluster of notions – what we might loosely call the *semantic* notions – are being pulled in two directions, one inclusive and one exclusive. ... [W]e have seen Sellars making this point with respect to the notions of 'descriptive', 'fact', 'proposition' and 'true' itself. In all these cases, he ends up saying, there's a *generic* notion application to declarative statements of all kinds and a *local* notion applicable much more narrowly – to the matter-of-factual, as Sellars puts it.

My response to this fundamental terminological tension has been to see it as reflecting the fact that all these notions are trying to serve two quite different masters. I have suggested we get a much clearer view of the landscape by making this explicit – by recognising that we have two quite distinct notions or clusters of notions in play, misleadingly being forced together by our failure to recognise the distinction and to modify our terminology accordingly. My terms *e-representation* and *i-representation* were my attempt to mark this distinction. (Price et al 2013, 166)

I go on to give my own gloss on the sense in which Sellars' two notions of truth belong in different boxes:

Sellars puts the difference in terms of the idea that picturing is a natural relation between *objects* – linguistic items considered as objects, on one side, and objects in our environment, on the other – whereas truth in the generic sense is a 'pseudo-relation', to be understood in terms of its inference-supporting role within the language game. I think we can emphasise this distinction even further ... by noting the different theoretical stance we employ in each case.

In the case of the generic notion, we are interested in a notion we find in use in ordinary language. To the extent ... that the explanatory, pragmatist approach recommends itself in such a case, our theoretical focus will be on the use of the notion. We will be asking, in effect, 'What our creatures like us *doing* when they use this notion? Why do they have it in their language in the first place?'

In the case of picturing, however, our focus, as Sellars himself always stresses, is first-order, matter-of-factual and highly theoretical. There is no reason whatsoever to imagine that the notion we find ourselves investigating will be in play in folk usage. And our theoretical interest is in the relation *in the world*, not in the use of certain terms in ordinary language. (Price et al 2013, 167)

The conclusions I take to be shared by Sellars, Schroeder, and myself are these. Deflation of the *folk* semantic notions need not imply that there can be no role for *theoretical* notions that, at least at some superficial level, seem to be working the same territory; but if we want to avail ourselves of this option then we need to keep our eye on the distinction, and not confuse one for the other.

This is not to say that Schroeder, Sellars and I are completely on the same page. For one thing, Schroeder and Sellars seem to be local expressivists. They admit notions of ‘matter-of-factual’ language (Sellars) or ‘descriptive’ language (Schroeder), taking these to be distinguished from the vocabularies for which expressivism provides the appropriate theoretical framework. Whereas I claim to be a global expressivist. One might ask how this is possible, if I acknowledge that some vocabularies are e-representational, and want to regard this as at least in some sense analogous to the narrow notions that Sellars and Schroeder employ. This is an excellent question, to which I’ll return in Chapters 11 and 12. (I’ll argue that Schroeder’s own preferred formulation of expressivism helps to point the way to an argument for regarding it as global in scope.)

Relatedly, Schroeder’s viewpoint takes metaphysics more seriously than I am inclined to do. As he says, he uses the term ‘representational content’, ‘for the entities, whatever they are, which mark out distinctions in reality, and are associated with metaphysical commitment’ (2015, 88). This may mark a difference between Schroeder and Sellars, whose narrow notions of truth and content are firmly grounded in the biological world (much in the spirit of the work his student, Ruth Millikan). Famously, Sellars scholars disagree about the value of this aspect of Sellars’ view. Left-wing Sellarsians want to dismiss ‘picturing’ altogether. Elsewhere (Price et al 2013, Chapter 8, §5, Price 2017b) I have proposed a reconciliation, reading picturing as a pragmatic notion of e-representation, and recruiting Sellars to the ranks of the global expressivists. Not surprisingly, this has the effect of making the differences between Sellars and Schroeder greater than those between Sellars and me.

Finally, a remark to connect the discussion of these last two sections to §10.2. In §10.2 I argued that if Creeping Minimalism is to leave a space for an alternative to expressivism, this alternative view requires a theoretical vocabulary in which to discuss the relation between our words and their subject matter, other than the vocabulary deflated by semantic minimalism. In effect, it requires that we split the terminology in which, on the surface, we seem to be talking about the relation between word and object. Some of the terminology – perhaps all of the *folk* semantic terminology – goes into the minimalist box, labelled ‘Too Thin for Theoretical Purposes’. While the rest of the terminology – perhaps entirely parts of it invented by theorists – gets saved for theoretical work, including the expressivist’s opponent’s work of accounting for the use of some parts of our language in ‘realist’, ‘factualist’, or ‘representationalist’ terms.<sup>21</sup>

Clearly, this two-boxing strategy is analogous to the one we have been discussing in the last two sections – one that I have identified in different forms in Sellars and Schroeder, as well as in my own distinction between e-representation and i-representation. In my version, as I’ve said, I have wanted to take the view that e-representation itself can be a pragmatic, functional notion, the sort of thing that an expressivist properly appeals to in answering explanatory questions about some assertoric vocabularies. If so, then it won’t after all provide what the expressivist’s opponent was looking for, in the face of Creeping Minimalism. So this issue, too, remains on the table, as we turn to the concerns of the remaining chapters.

---

<sup>21</sup> In practice, of course, the same term may have both a folk and a theoretical use, and we need to be prepared to pull them apart.



## Chapter 11

### Cambridge Expressivism

The four remaining chapters are mainly in survey mode. This chapter asks what we should take expressivism to *be*. We will be interested in defining characteristics, routes of approach, and terminological choice points. In Chapter 12 I'll explain why I take expressivism to be properly a global programme, applicable throughout the assertoric domain; but I'll also discuss in some detail one particular application, that to causal concepts. And in Chapters 13 and 14, finally, we'll look over the fence, and consider this kind of expressivism's relation to several other views, most of them philosophical neighbours, in one sense or another.

The present chapter begins with the puzzle held over from §10.1, concerning the relation between two different strategies for defending expressivism in FFT and in much of my later work. One strategy turns on the idea that different assertoric vocabularies have different underlying functions, the other on the phenomenon of No Fault Disagreement (NFD). Are these approaches in tension with one another, and if not, how do they relate? These questions will lead us deep into issues, some terminological and some theoretical, about how we should characterise expressivism. The landscape in the literature is confusing, both conceptually and terminologically. I'm conscious of having contributed to this confusion, not least by using a variety of different terminology myself.

In this chapter I'll try to sort things out. The main work of the chapter will be to pull together my proposed general recipe for expressivism, combining several ingredients and philosophical motivations. It is intended to be a recipe with very broad application – in fact, *global* application, for reasons I'll return to in Chapter 12. However, I'll highlight points at which there is an option to use 'expressivism' in a narrower way, choosing a different term for the general case.

Because my use of 'expressivism' is in some ways idiosyncratic, it might be helpful to give it a distinguishing label. If so, then to acknowledge its debt to figures such as Ramsey, Wittgenstein and Blackburn, as well as to my own Cambridge connections, early and late, I propose we call it *Cambridge* expressivism.<sup>22</sup>

#### 11.1 Two ways of defending expressivism

In Chapter 10 I noted that the comparison between FFT's defence of expressivism and my own later trajectory raises a puzzle. On the face of it, I have offered *two* lines of defence. FFT defends expressivism by pointing to the phenomenon of NFDs. My later work defends it by pointing out that the important part of expressivism is the positive, functional story it offers us about various discourses, and that this is quite untouched by semantic minimalism. (Indeed, it is reinforced by it, since minimalism closes down the option of substantial *semantic* accounts of function, which is the traditional rival to expressivism.) What is the relation between these two lines of argument? Is there any tension between them? Indeed, are they actually defending the same thing in the first place?

---

<sup>22</sup> The name makes me think of Cambridge Burnt Cream, a version of *crème brûlée* supposedly invented at Trinity College. According to the Trinity website (Trinity College, 2021) this story 'almost certainly has no basis in fact'. That seems an excellent precedent for a view that aims, in the spirit of FFT, to do exactly the reverse – in other words, to provide a basis *for* factuality!

The functional, explanatory dimension is not missing from FFT altogether. On the contrary, it frames the approach of the positive part of the book. (In §9.4, for example, I characterised the approach as ‘functional or explanatory realism.’) It is true that my proposed explanation of the assertoric or factual language game *as a whole* is on the front foot, but FFT is also clear that there are underlying functional differences *between* different discourses – differences with the potential to explain their variable susceptibility to NFDs.

Thus FFT had the materials for a ‘purely functional’ defence, putting the weight solely on the explanatory ambitions of expressivism. Where it would have ended up is nicely described by Blackburn, in passages in which he reads Wittgenstein as a kind of global quasi-realist:

The focus of theory is the nature of the commitment voiced by one adhering to the proposition, and the different functional roles in peoples’ lives (or forms of life, or language games) that these different commitments occupy. ... You may end up, that is, saying that these assertions describe how things are with values, probability, modality, and the rest. But the way you arrive at this bland result will be distinctive, and it will be the bit that matters. (Blackburn 1998a, 166–167)

This position may be bland at the level of linguistic appearances, but it is far from bland as a piece of philosophical theory – indeed, it accuses its factualist rivals of something like theoretical blandness, or failure to notice interesting differences, in their blindness to the functional–explanatory layer. In many ways, it would have been a very satisfactory conclusion for FFT.

Why didn’t FFT go that way? A charitable interpretation would be that I was more ambitious. I felt that the level of linguistic appearances didn’t need to be quite so bland. On the contrary, *we could find differences there, too, if we took the trouble to look*. We could find NFDs, popping up for different reasons in different language games. And we could correlate those differences both with intuitions about comparative ‘factuality’, or lack of it, and with variations in the underlying functional stories. Relatedly, I may have felt – rightly, as I would still say – that direction of explanation alone doesn’t get us to the intuitive sense of ‘gradation’ of factuality. In FFT I wanted gradation, in place of bifurcation, so I needed the additional story about its surface manifestations in NFDs.

However, another possibility is that I felt – wrongly, I would now say – that I had to take this extra step, if I was to defend expressivism at all. I may have taken too seriously my own argument that various underlying functions were compatible with deflationary realism, and simply failed to see that direction of explanation still saved the day for expressivist intuitions. Late in §8.10 there are a few paragraphs that discuss the issue as to whether the approach in terms of NFD captures all of non-factualism, or whether there are other cases in which non-factualism can be grounded on the ‘metaphysical and epistemological factors’ discussed in Chapter 4. The discussion there doesn’t reach a firm conclusion, but I certainly note the argument of Chapter 4, that a determined factualist can accommodate those factors. These days I would be inclined to push such a factualist harder on the explanatory questions.

At any rate, I want to argue in this chapter that these two routes to defending expressivism are not in tension, and that the route I foregrounded in FFT, in terms of NFDs, adds something significant – something worth bearing in mind in contemporary discussions of expressivism. Indeed, there are some signs that contemporary expressivism is moving in this direction. In Chapter 13 I will discuss the connection between the framework of FFT and the so-called New Relativism of writers such as John MacFarlane (2003, 2005a, 2005b) – links highlighted by Shapiro (2014), who notes the analogies between the linguistic phenomena to which New

Relativists appeal and FFT's NFDs. Schroeder, too – reversing a suggestion by MacFarlane himself – has suggested that expressivism might be 'relativism done right' (2015, 25).

To understand the relationship between these two routes to expressivism, and to prepare the ground for considering these issues in Chapter 13, we need a great deal of clarification about what we mean by expressivism. Most of the remainder of this chapter will be devoted to distinguishing several ingredients of, and motivations for, contemporary expressivist views, and to flagging the need for terminological choices at several points. At the end I'll come back to these issues about expressive function and NFD.

As a taster, these are the main ingredients that we will be discussing, in addition to the proposal of FFT itself: a use-based approach to meaning; a programme presenting itself as an alternative to contemporary reductive metaphysics; an explanatory programme, proposing to account for the *existence* and *practical relevance* of the particular concepts or vocabularies in question – typically the former in terms of the latter, in some way; a focus on contingent *features of speakers* – typically features of practical or 'pragmatic' significance – that play characteristic roles in expressivist accounts of particular vocabularies; and a kind of *perspectivalism*, with these contingent features playing the role of the perspective from which the users of that vocabulary speak. I will link this last ingredient to the Copernican metaphor familiar from Kant, noting how well it characterises the sense in which expressivism provides an alternative to metaphysics. What we took to be a feature of reality, in need of metaphysical investigation, is revealed to be an artifact of our perspective.

## 11.2 Expressivism as use-first semantics?

We have observed that traditional expressivism made two claims: a *negative claim*, to the effect that the utterances in question are not truth-conditional, or something of that 'semantic' sort; and a *positive claim*, that the utterances in question have such-and-such other (not semantically-characterised) function. I noted that semantic minimalism deflates the negative claim, but leaves the positive one intact – and that that's the one that matters. Indeed, we saw in the previous chapter that far from undermining expressivism, semantic minimalism threatens its factualist opponents, whose own positive story about the function of the utterances concerned would be expected to appeal to such semantic notions.

In early versions of expressivism, including the ethical and probabilistic examples discussed in FFT, the positive claim took a particular form. The function of the utterances concerned was said to be that they 'expressed' psychological states of distinctive kinds – pro-attitudes and credences, say, in these cases. Hence, in this tradition, the use of the label 'expressivism'.<sup>23</sup> The important thing about these psychological states is that they *not* be characterised, *at least initially*, simply as beliefs. The qualification is important. We have seen that the expressivist may be happy to say – downstream, as it were – that 'It is probable that P' expresses a belief that it is probable that P. But what it is to express such a belief is explained – upstream, in the same metaphor – by pointing to credences, and saying that this kind of talk begins life as an expression of an attitude of that kind.<sup>24</sup>

---

<sup>23</sup> The term seems to have been introduced explicitly by Gibbard (1986, 473), who says of his own view: 'Such an analysis might be called "expressivistic." According to the analysis, claims about what it makes sense to do express a state of mind: the speaker's acceptance of a system of norms.' Gibbard cites Blackburn's (1984, 167–171) discussion of 'expressive analyses' as a precursor.

<sup>24</sup> Gibbard puts this important point like this.

Mark Schroeder (2008, §2.4; 2015, Ch. 1) helpfully distinguishes several quite different things that might be meant by *expressing*, in this kind of account. He argues that some of them – including a causal version offered by early emotivists, and a Gricean intentional account initially offered by Gibbard (1990) – are unsatisfactory. He recommends instead what he calls ‘assertability expressivism’. This notion depends on a distinction between truth conditions and what Schroeder calls ‘semantic correctness conditions’ – which, as he says, should be thought of as *use* conditions.

Every sentence in the language is associated with conditions in which it is *semantically correct* to use that sentence assertorically. But we are not to think of those conditions as *truth*-conditions so much as *use*-conditions. After all, a speaker who is mistaken about who is president of the United States, believing it to be John McCain, is not making a *linguistic* mistake by asserting, the president of the United States is John McCain. She is making a *mistake*, but not a *linguistic* one, since she really does believe that John McCain is president. So semantic correctness conditions capture the conditions under which a speaker would be making a *linguistic, semantic* mistake in the use of some sentence, which are not the same as its truth-conditions.

Moreover, according to assertability expressivism ... these assertability conditions are not assigned to sentences *derivatively*, by means of their truth-conditions. Rather, sentences get their truth-conditions from their assertability conditions ... (2015, 50–51)

Schroeder is quite right, in my view, to call attention to the importance of this distinction, which is the one I drew in this passage from (Price 1994), quoted in §10.4:

The platitude that to know the meaning of a sentence is to know the conditions for its correct use has two importantly different readings (not always properly distinguished). It may be taken to refer to what I ... called subjective assertability conditions, so that it amounts to the claim

- (4) To know the meaning of “It is snowing” is to know that it is normally appropriate to say “It is snowing” only when one believes that it is snowing.

Or it may refer to truth conditions, so that the claim is that it is correct to say “It is snowing” if and only if it is snowing; or, in the more familiar form, that

- (5) “It is snowing” is true iff it is snowing. (Price 1994, 73)

I had drawn the same distinction even earlier, discussing Dummett’s views on meaning:

---

¶¶ If the account is to qualify as expressivistic, it can’t specify the state of mind expressed by citing its content. The state of mind expressed by ‘Stealing is wrong’ can’t be specified, in the explanation, as anything like ‘believing that stealing is wrong’. The idea is to explain the meaning of the word via explaining the states of mind that constitute believing things couched with the term. We can’t, then, legitimately start out relying on an understanding of what it is to believe things so couched. We can’t just say that there’s a general relation of *believing* a proposition, and that believing that stealing is wrong is standing in this relation to the proposition that stealing is wrong. (2015, 212)

Gibbard goes on to say that he now regards himself as a *global* expressivist in this sense, although, as he says, ‘my sense of the term “global expressivist” may not be Price’s’ (2015, 211). But there’s not much difference between us, in my view, at least on this point; see (Price 2023a) for discussion.

[I]t is doubtful whether the view that the meaning of a sentence is determined by its assertion conditions—by *when* it may be correctly asserted—need offer this as a revised account of what it is *that* a person who makes an assertion is claiming to be the case. The alternative is to say that although the content, or sense, of an assertion is ultimately determined by its assertibility conditions, it does not *state* that these conditions hold. 'Eric is flying' states that Eric is flying, and not that it is assertible that this is so, even if what it is to state that Eric is flying is ultimately to be understood in terms of *when* this may correctly be stated. (1983b, 163)

Indeed, the distinction was already crucial to the credal expressivism I defended in my PhD thesis. As I emphasised there, it is crucial to distinguish the claim that it is correct to say 'It is probable that P' *when* one has a high credence that P from the idea that in saying 'It is probable that P', one is saying *that* one has such a credence.

Accordingly, I am pleased to find myself on the same page with Schroeder in thinking that this distinction provides a helpful way to characterise expressivism, or at least one important ingredient of it.<sup>25</sup> The expressivist proceeds by giving use conditions – crucially, use conditions that are not, in the first instance, simply a matter of saying that a statement P is appropriate when the speaker believes that P.<sup>26</sup>

In the canonical cases of probabilistic and evaluative claims, the expressivist's use condition typically refers to a distinctive kind of *psychological* state – credences and pro-attitudes, respectively. The 'expressive' terminology might lead us to assume that all such use conditions refer to psychological states, and Schroeder himself argues that this will be the case. He appeals, as he says, 'to the idea that assertability conditions give the rules under which speakers will count as making *linguistic mistakes*' (2015, 51).

If assertability conditions are not to fail due to an error of the speaker's about something other than language, then they will have to be something to which speakers have a special kind of *access*. And so it makes sense to suppose that they will be conditions of the speaker's own mind. And that explains why assertability conditions always mention a mental state. (2015, 51)

This seems to me a little quick, but rather than trying to unpack possible points of disagreement, I shall simply appeal to Schroeder's view as a clearly-marked exemplar of one path from a terminological fork.

What is the other path? It agrees, of course, that expressivism proceeds by giving use conditions – use conditions that are not initially characterised in terms of belief. But it admits a much broader conception of use conditions, in which mental states need not play the central role. A good example of such a view is the kind of deflationary theory of truth defended by Horwich (1990). Like an expressivist about probability or value, the deflationist disavows the traditional project of investigating the *nature* of truth, taking it to rest on a mistake about the function of ascriptions of truth and falsity. Like such an expressivist, again, the deflationist will not say (except in the deflated, downstream sense) that the use condition for 'P' is that one believes that it is true that P. But nor will she offer any alternative psychological state at this point. On the

<sup>25</sup> I don't claim to have been free of the faults that Schroeder attributes to earlier notions of expression.

<sup>26</sup> As we just observed, and as we saw in Chapter 10 that Schroeder (2015) himself now suggests, the expressivist may be able to say that with everybody else at the end of her account ('downstream'), but has to begin somewhere else. The contrast is with views that *begin* with the belief, or with its content. Cf. the remarks by Gibbard (2015) in fn. 3 above.

contrary, the use identified for truth ascriptions is a matter of their logical or inferential role of truth ascriptions (e.g., in allowing generalisations such as ‘Everything that Ramsey said is true’).

The contrast with expressivism on Schroeder’s side of the fork is a matter not just of allowing non-psychological use conditions, but of a different conception of what a use condition *is*. We can make this clear by introducing another recent model, that of Michael Williams (2010, 2013). Williams proposes a general model for use-based accounts of meaning, developing ideas he finds in Wilfrid Sellars. He calls the central component ‘an explanation of meaning in terms of use (an EMU)’ (2010, 321).

Williams’s EMUs involve three kinds of clauses: intra-linguistic *inferential* (I) clauses, *epistemological* (E) clauses (roughly, language entry and exit rules for the vocabulary in question), and *functional* (F) clauses. Williams illustrates the model with several examples, including a deflationary theory of truth of Horwich’s sort. In this case the component clauses of the EMU are as follows:

(I-T) A *material-inferential* (intra-linguistic) component. Excepting sentences that generate paradox, the inference from ‘Snow is white’ to ‘It is true that snow is white’, and vice versa, is always good; the inference from ‘Grass is green, to ‘It is true that grass is green’, and vice versa, is always good, and so on...

(E-T) An *epistemological* component. Such inferences are *primitively* acceptable (*a priori*). They are ‘free’ moves in the discursive game.

(F-T) A *functional* component. The truth-predicate is important exclusively as a generalizing device.

In this meta-theoretical analysis, the I- and E-clauses of an EMU specify the inferential patterns, or proprieties, that competent uses of ‘true’ display (or respect) in their use of ‘true’. These clauses are fundamental in that they neither receive nor need any deeper *theoretical* explanation. They do, however, both invite and receive a *functional* explanation from the F-clause. This clause explains why we have a concept that answers to the use-patterns given by I-T and E-T. So although there is a sense in which Horwich is quite right to say that the rule of use indicated by MT is ‘the source of *everything else* we do with the truth predicate’, there is an equally good sense in which *what we do with the truth-predicate* is the source of the predicate’s distinctive use-properties (or proprieties). (2010, 323)

Williams emphasises that between the I and E clauses, on one side, and the F clauses, on the other, there is a crucial distinction:

The I- and E-clauses, on the one hand, and the F-clauses, on the other, are concerned with aspects of use that must be kept separate. Roughly, the inferential and epistemological properties (or proprieties) captured by the I- and E-clauses concern *how* certain vocabulary-items are (to be) used. These clauses fix meaning in the sense of ‘conceptual content’. By contrast, the F-clauses capture what an item conforming to such properties/proprieties can be used (is useful) *for*: they capture ‘meaning’ in the sense of pragmatic significance. Failure to keep these distinctions in mind is the source of the temptation to think that, when deploying certain vocabulary-items susceptible of minimalist analysis, we aren’t *really* saying anything but *only* doing something. (2010, 325)

In other words, Williams is pointing out that expressivism needs to concern itself with the use, or function, of various vocabularies *in two quite distinct senses*. One sense, associated with the EMU's I and E clauses, is a matter of *how* the terms in question function, or how they are (correctly) used. The other, associated with the F clause, is a matter of what having terms that work in this way 'does for us' – roughly, the difference it makes to creatures in our situation to have this vocabulary in our repertoire.

This is not the kind of distinction that requires us to make a choice. On the contrary, Williams's point is that expressivists should be interested in both notions of use and function (while keeping them distinct). Williams goes on to propose that missing this distinction is the source of 'the local expressivist's *ur*-mistake' (2010, 325) – the mistake that prevented early expressivists from seeing, for example, that expressions of affective attitudes could nevertheless have cognitive content.

Let's note four differences between Williams's approach and Schroeder's. First, Williams calls our attention to this distinction between two notions of use, and insists that expressivists need both notions. Second, there is nothing in the EMU directly comparable to Schroeder's assertability conditions. Third, and conversely, the EMU's I and E clauses, describing the inferential and language entry/exit properties of the judgements in question, add elements missing from the assertability conditions account. And the fourth, there is nothing in Williams's model corresponding to Schroeder's requirement that assertability conditions involve psychological states in a central way.

The first three of these contrasts may reflect some substantial disagreements about the necessary elements of a use-based expressivism, though my instinct is to seek consensus by amalgamation. The last, it seems to me, is more a place for terminological choice. The issue of terminology deserves a section to itself.

### 11.3 Terminological choices

With EMUs on the table, it is clear that we face a terminological choice. We can use 'expressivist' in a narrow sense, as Schroeder recommends, for cases in which the use condition does involve a psychological state, finding a different term for views such as Horwich's theory of truth. Or we can use it in the broad sense, applying it also to views such as Horwich's. As usual in philosophy, the choice of terminology is less important than the recognition that a choice has to be made, one way or the other. Without such recognition, there's a real risk that we talk past each other.<sup>27</sup>

Whenever I have described my own view as global expressivism, I have (of course) been using the term in the broad sense. This seemed a natural usage, especially for someone interested, as I have been (e.g., in Price 2011a), in the convergence between uses of 'expressivism' in the Humean tradition associated with writers such as Blackburn, and the Hegelian tradition associated with Brandom. What Brandom means by the term is global in any case. What Blackburn means by it is something that works much better if thought of as a global view, in my

---

<sup>27</sup> Gibbard (2003, 80) also draws a distinction between broad and narrow expressivism, but means something different. His terminology rests on a distinction between a narrow class of 'straight attitudes', and a broader class of 'attitude-laden judgements' – roughly, judgements whose contents get to be explained in terms of straight attitudes. A narrow expressivist about a class of claims takes them to express straight attitudes; a broad expressivist says that 'attitude-laden statements express attitude-laden judgements', but not necessarily straight attitudes.

Note also that even for the narrow view in my sense, there is a further terminological choice. Do we restrict the term 'expressivism' to cases in which the psychological state is not an ordinary belief, so that it *contrasts* with what Schroeder calls descriptivism. Or do we count descriptivism as a special case of expressivism? Gert (2023) makes the latter choice, but it seems to me closer to the standard uses of the term to make the former.

view, not least because that removes an internal instability. Neither Blackburn nor Brandom feels tied to psychology, as the narrow use of the term would require.<sup>28</sup>

Still, I've also been conscious of a tug in the other direction, a feeling that we stretch the relevant notion of expression a bit far, if we break the link with psychology. What are the alternatives? Sometimes I have relied on a distinction between 'use-based theories of meaning' and 'content-based theories of meaning' (and their respective advocates, 'use theorists' and 'content theorists'; see Price 1993). These terms have the advantage of being more immediately informative than 'expressivism'. In the concise terminological spirit of 'knowledge-first epistemology', I now want to add 'use-first semantics' and 'content-first semantics' to the mix. I say 'semantics' rather than 'theories of meaning' because it is less of a mouthful, and because Schroeder and others already characterise expressivism itself as a 'semantic' programme. As a result, however, another terminological issue needs to be flagged. In FFT and elsewhere I often use 'semantic' as a label for matters of truth, reference and the like; whereas the term 'use-first semantics' calls for a broader conception, of course.<sup>29</sup>

That point aside, I hesitate on two grounds to recommend 'use-first semantics' as a *replacement* for 'expressivism'. First, I think it overstates the degree to which expressivism is (solely, or even mainly) a semantic programme, ignoring other central aspects – more on these as the chapter develops. Second, it risks overstating the degree to which use-first semantics and content-first semantics are *rivals*. I think it is crucial to bear in mind that there may be more than one important project in the territory we loosely call semantics, or the theory of meaning. Arguing that some part of the territory is use-first, even globally so, need not be incompatible with thinking that there is also an important role for content-first or truth-conditional semantics. (More on these issues in Chapter 13.)

Another alternative to 'expressivism' is 'pragmatism'. I have often referred to my view as a kind of global pragmatism – indeed, as Cambridge Pragmatism (Price 2017a, 2017b). This choice avoids the 'tug to psychology' of 'expressivism', but has a different defect. 'Pragmatism' is very widely used for other views; in some cases, arguably, for conflicting views. 'Expressivism' treads on fewer terminological toes. Yet the link with pragmatist themes is undeniable. Whatever else we say, we should recognise that expressivism is a kind of pragmatism.

There may be a case for a neologism. Some writers, notably Gert (2018, 2023a, 2023b)), following Williams (2010), now use 'neo-pragmatism' for expressivism in the broad sense. This is a helpful proposal, in my view. True, 'neo-pragmatism' is also used in other contexts, but I am not aware of any that risk the kind of terminological tribalism that 'pragmatism' itself sometimes encourages. (If necessary, after all, we could call ourselves *expressivist* neo-pragmatists.) So I offer 'neo-pragmatism' as an alternative here, to anyone who feels that 'expressivism' won't stretch as far as I want it to go.

There is also a terminological issue on the other side of the fence, for the expressivist's 'factualist' opponents – and indeed for a *local* expressivist's account of what she takes to lie on the 'descriptive' or 'factual' side of the line, where she no longer takes expressivism to be the appropriate approach. In effect, this is the Creeping Minimalism issue again. Minimalism tends to make talk of *description*, *facts*, and the like, common property, at least in their deflated senses. But this means that terms such as 'descriptivism' and 'factualism' itself are in danger of losing

<sup>28</sup> This may not have been true of Blackburn's early work, but by the 1990s Blackburn had certainly come to recognise that deflationary theory of truth, such as Horwich's, was grist for the quasi-realist's mill.

<sup>29</sup> Partly for this reason, it would be attractive to follow writers such as Chrisman (2015) and Ridge (2014) in describing expressivism as a form of *metasemantics*.

their grip on the anti-expressivist views they are intended to characterise. The same is true of ‘realism’ – again, remember Creeping Minimalism.

In this case, the terminological problem reflects the fact that unless one abandons minimalism in some way, there is no viable alternative to expressivism (except the kind of quietism mentioned in §10.3, which simply declines to address the explanatory questions that motivate expressivists). As we saw in Chapter 10 (§10.2, §10.6–7), this means that a non-global expressivism requires a *non-deflated* word–world relation of some kind, in order to characterise the non-expressive side.<sup>30</sup> I noted a structural analogy in this respect between my own notion of e-representation and proposals by Schroeder and by Sellars. In Schroeder’s case, for example, his use of the term ‘descriptive’ rests on the theoretical notion of *representational content*. But that means that in his usage the term ‘descriptive’ is itself theoretical, and doesn’t mean simply what it means in ordinary usage. The expressivist’s factualist opponents can attempt to build themselves a platform in this way, but they can’t at the same time claim simply to be defending common sense.<sup>31</sup>

In much of my own recent work, especially since (Price 2004c), I have often referred to the view I take my global expressivist to be opposing as ‘representationalism’ (or ‘Representationalism’, with a capital R). This term, too, hardly wears its meaning on its face. Indeed, in ‘One cheer for representationalism?’ (Price 2010) I distinguish three grades of commitment to representationalism, in addition to the outright rejection we associate with Rorty. Those distinctions needn’t concern us here, but it is relevant here that my use of the term ‘Representationalism’ is linked to a particular route to expressivism – a different route, in some ways, from the semantic path we have considered so far. This route begins with the attractions of expressivism as a solution to what are often thought of as metaphysical puzzles. In order to have this face of expressivism clearly in view, let me next explain what this route involves, and how I use the term ‘Representationalism’ in this context.

#### 11.4 Expressivism as an alternative to metaphysics

In the years after FFT was first published, one of my major interests was to try to exhibit the benefits of a wide-ranging expressivism as an alternative to popular programmes in reductive metaphysics. This project was in the spirit of the deflationary, pluralistic realism of Chapter 9 of FFT. Being based in Australia, I had distinguished metaphysicians as colleagues and philosophical neighbours – and, famously in one case, as regular visitors. The attempt to be clear about how I disagreed with giants such as Armstrong, Jackson, and Lewis was often an important motivation.

My first new piece in this deflationary vein was ‘Metaphysical pluralism’ (Price 1992a), written in 1989. Here my targets were those contemporary metaphysicians who felt the need to bulk-up a bare, Humean conception of the world, in order to provide (as some of them put it) ‘truth-makers’ for the claims of modality, mentality, or some other difficult case. These philosophers thought of themselves as defending a conception of a single World or Reality – just a richer or bulkier world than that envisaged by their Humean opponents. I called this kind of view *additive monism*, and went on to argue that it is surprisingly difficult to distinguish from the kind of Wittgensteinian linguistic pluralism I had come to in FFT. To make a case for their monism – i.e., for the view that their truthmakers live in a *single* world – these metaphysicians need to rely

<sup>30</sup> For an interesting recent attempt to do this, from a broadly pragmatist perspective, see Zalabardo (2023). I discuss Zalabardo’s approach in (Price 2024).

<sup>31</sup> I noted in Chapter 10 that in my own case, I don’t take this move to provide a *successful* challenge to global expressivism. More in Chapter 12 on my reasons for this view.

on the claim that the different vocabularies are all in the same line of work; that they are all in the business of ‘representing the world’, or something of that kind. The paper argues that a substantial basis for this claim would amount to a foundation for the distinction between factual and nonfactual discourse, and draws heavily on FFT in arguing that this is not a straightforward matter. It also criticises Blackburn’s quasi-realist for relying on such a distinction, suggesting instead that we abandon it, and ‘extend ... quasi realism "all the way down".’ (1992a, 42)

I characterise the upshot like this.

[I]t may be helpful to emphasize that discourse pluralism is not an irrealist position. The pluralist accepts with all sincerity that there are moral states of affairs, possible worlds, numbers, or whatever. What he or she rejects is the additive monist’s attempt to put a further metaphysical gloss on such existential claims. (That gloss turns out to depend on a semantic distinction of questionable standing, or so the pluralist argues.) Without the gloss, discourse pluralism sits quite happily with a non-metaphysical or “minimal” realism. (1992a, 49)

I argue that discourse pluralism is the default position, the onus being on the metaphysical monists to defend any alternative to it. And I claim to find an ally in Crispin Wright, who had heard a version of the paper at a conference at the University of Queensland in 1989, and from whom I was learning to speak of minimalism. After noting a difference of opinion with Wright about Dummett’s anti-realism, I continue as follows:

That issue aside, I take the following remark of Wright’s to be in much the same vein as my emphasis on the economical advantages of discourse pluralism: “Anti-realism thus becomes the natural, initial position in any debate. It is the position from which we have to be shown that we ought to move. All the onus, everywhere, is on the realist” (1993: 69). Substitute “pluralist” for “anti-realist” and “non-pluralist” for “realist,” and these are my sentiments exactly. (1992a, 48)

So it is clear that I already thought of minimalism as good news for what I would now call global expressivism.<sup>32</sup>

A little later in the 1990s, another local piece of metaphysics became a particularly useful foil. This was Frank Jackson’s ‘Canberra Plan’, dubbed as such by John Hawthorne and I in our (1996) piece, written in 1994. I soon borrowed from Jackson the terminology of *location* or *placement* problems. Placement problems are the familiar philosophical puzzles about the place of moral properties, probabilities, causation, meaning, and so on, particularly in a natural world. In all such cases, I argued, it was crucial to distinguish between two different questions we can ask in the face of such a puzzle. Taking causation as our example, we can ask ‘What is causation?’, or we can ask ‘What are we humans doing when we use terms or concepts such as *cause* and *effect*?’ The former question is in *material* mode, the latter in *linguistic* or *psychological* mode. This is in the spirit of the distinction I drew between *analysis* and *explanation* in FFT, of course.

As the familiar examples of ethical and credal expressivism make clear, the expressivist takes the linguistic or psychological question to be the fruitful one, and takes the relevant placement problem to have been addressed, once it has been answered in a certain way. In what way? Recall Dreier’s explanationist proposal for the distinction between expressivism and (what he calls)

---

<sup>32</sup> I reconsider this comparison between my views and Wright’s in (Price 2022d); see also Chs. 13 and 14 below.

realism. This is Dreier's version of the moral expressivist's crucial claim: '[T]o explain what it is to make a moral judgment, we need not mention any normative properties' (2004, 39).

Contrast this to what I call a metaphysical approach to a placement problem. The metaphysical approach either takes the *material* question to be the one that needs to be answered – 'What is causation?', 'What is goodness?' – or, accepting the *linguistic* starting point, seeks an answer in terms of reference, truthmakers, or some such. In other words, in the latter case, it seeks an answer that provides something for the language in question to be *about*.

In my discussion of the work of Jackson and the Canberra Planners, I sought to do three main things: (i) to call attention to the under-recognised role being played by semantic presuppositions, in Jackson's programme and other parts of contemporary metaphysics; (ii) to argue that these presuppositions are problematic; and (iii) to make it clear how expressivism provides an alternative response to placement problems.

Concerning point (i), I introduce the term '(big R) Representationalism' for an assumption – or, as I say, a 'proto-theory' – that I take to be presupposed by much contemporary work in metaphysics. As I put it, 'the proto-theory says that our statements "stand for", or "represent", aspects of the world' (Price 2011c, 5). I note that many contemporary naturalists combine this proto-theory with the assumption that all the available referents or truthmakers for our claims must lie in the natural world, in anywhere at all, and that it is this 'combination that leads to the puzzles to which they devote so much philosophical energy' (2011, 5).

I claim that a recognition of the role of the proto-theory – Representationalism – is important for understanding what many contemporary naturalists take themselves to be doing. Even more importantly, as I put it, 'it reveals an interesting vulnerability in the Naturalist's own position':

By the Naturalist's own lights, the proto-theory ought to count as an hypothesis about what it is right to say about language itself, from a naturalistic standpoint. If it turned out to be a bad hypothesis – if better science showed that the proto-theory was a poor theory – then the motivation for the Naturalist's [metaphysical project] would be undermined. But it would be undermined from within a scientific view of language and its place in the world. In that sense, the undermining wouldn't be an anti-naturalist conclusion – on the contrary, it would depend on convicting some self-styled naturalists of sub-optimal science. (2011c, 5)

Exploiting this vulnerability brings us to point (ii) in my case against the Canberra Plan. I offer three main arguments against the proto-theory, or at least against the assumption that it can play the role it is often being asked to play in contemporary metaphysics. The first argument appeals to the attractions of semantic minimalism. The second relies on points I borrow from Stephen Stich (1996), to the effect that no scientifically-respectable theory of reference seems likely to bear the weight required by metaphysics. And the third argues that such presuppositions are self-undermining, in the context of a global programme such as Jackson's. (See Price 1997; 2004c; 2009a; Menzies and Price 2009.)

As I note, my metaphysical opponent can try to avoid these difficulties by staying firmly in material mode, which is what Stich himself recommends. Concerning the question of eliminativism about folk psychology, for example, Stich argues that the right question is not whether the terms 'belief' and 'desire' *refer* to anything, but simply whether there are beliefs and desires. I argue that this option breaks the analogy between the Canberra Plan and Lewis's original argument for physicalism, because the parallel requires that semantic relations play the

role that causal relations played in Lewis's original case. That line of argument needn't concern us here (see Menzies and Price 2009 for details). What matters for present purposes is that expressivism makes the opposite choice, *staying in linguistic mode but rejecting the Representationalist proto-theory*.

In these familiar contexts, in other words, expressivism stands out in combining two commitments. In approaching placement problems it combines an insistence that the linguistic or psychological question *comes first*, with a rejection of approaches that seek to answer that question *in referential terms*. Or as Macarthur and I (2007, 233) put it (calling the view pragmatism rather than expressivism):

PRAGMATISM = LINGUISTIC PRIORITY without REPRESENTATIONALISM.

What is the relation between this route to expressivism and the use-first semantic view? Michael Williams explains how it leads us to the same place:

*Linguistic priority* is a methodological maxim. When dealing with metaphysical issues, don't start by asking about (say) the nature of values: examine what is distinctive about evaluative language. *Anti-representationalism* is a more substantive commitment. Representationalists explain the (proper) use of vocabulary-items in terms of their meanings, and explain meaning (at least of non-logical vocabulary) in terms of semantic (word–world) relations, such as reference. By contrast, anti-representationalists eschew the use of semantic notions as explanatory primitives. All vocabularies – semantic vocabulary included – are to be characterized (or explained) functionally, in terms of their use-properties. Oversimplifying a bit, meaning does not explain use: use explains meaning. (Williams 2010, 318)

There's another message I've been keen to stress concerning this metaphysics-avoiding face of expressivism. It is that clarity about what we take our philosophical project to be is indispensable. This message is intended for all sides in the relevant debates. For our part, we expressivists can't be clear about what our view is until we are clear about the questions it takes to be worth asking, and how they differ from those typically asked in metaphysics.<sup>33</sup>

In this spirit, I have criticised Brandom for sometimes failing to make clear that his own proposals fall on the expressivist side of the line (see, e.g., Price 2011a). I have also noted (see Price 2017a, 160–161) that there is live debate about whether David Lewis's work on knowledge is actually an account of *knowledge*, or of *knowledge ascription*. Jonathan Schaffer (2015) has argued for the latter interpretation, which, as I point out, can be regarded as an expressivist view. The fact that such prominent writers can leave it obscure what question they took themselves to be asking is an indication that the line needs marking with much greater clarity, in my view.

All the more reason, then, to try to be clear about what expressivism is. I said at the beginning of the chapter that I wanted to compile a recipe, with all the key ingredients listed. Let's see where we stand.

## 11.5 The ingredients of expressivism (I)

At this point our recipe for expressivism has three main ingredients in play, and another on the table, still in search of its place in the mix. The first of the active components is the use-first approach to meaning. Our comparison of Schroeder and Williams has given us some sense of

---

<sup>33</sup> This was already the message of the distinction between analysis and explanation in FFT, of course.

what the options are, concerning this ingredient. Their approaches seemed in many ways complementary, and we might hope to add some of each to our mix.

The second active ingredient is a programme that presents itself as an alternative to reductive metaphysics – a programme motivated in the same way by placement problems, but combining an insistence on linguistic priority with a renunciation of the ‘representational’ moves that lead from there *back* to metaphysics.

The third ingredient – closely linked, obviously, both to the first and the second – is an explanatory programme. It aims, roughly speaking, to account for the *existence* and *practical relevance* of the vocabularies in question; typically the former in terms of the latter, in some way. Why do creatures like us employ these terms and concepts? And why do these terms and concepts exhibit distinctive links to various aspects of our practical lives? In the case of credal and evaluative expressivism, for example, the latter question amounts to what FFT calls the confidence and approval problems, respectively. We could unpack this ingredient further by distinguishing the explanatory tasks, but let’s think of it as a single package.

In different ways, all of these three ingredients contribute elements of functional distinctness, and hence give content to what I described as my second, post-FFT, route to a defence of expressivism. Sitting on the bench beside this amalgam, we have a fourth item, viz., the framework of NFDs from FFT. We are aiming to come back to the question as to whether it belongs in the same recipe, and if so, how it gets incorporated.

To take us in that direction, I want to introduce another ingredient. It involves a focus on *features of speakers* – typically features of practical or ‘pragmatic’ significance – that play characteristic roles in expressivist accounts of particular vocabularies. These features are already in the mix, individually as it were, when we examine any of the first three ingredients in any particular case. But by considering them collectively, as a coherent category, we will be able to understand the relationship between NFDs and the rest of the recipe.

In order to make a case that this new ingredient is a legitimate part of the mix, and deserves attention in its own right – and in order to identify some key steps in the evolution of the mature expressivist recipe, some of them steps that call for terminological choices – let’s go back to the beginning.

## 11.6 Three dogmas of expressivism

Early expressivism often took its inspiration from Hume’s distinction between the operations of reason and taste.

[T]he distinct boundaries and offices of reason and of taste are easily ascertained. The former conveys the knowledge of truth and falsehood: The latter gives the sentiment of beauty and deformity, vice and virtue. The one discovers objects as they really stand in nature, without addition and diminution: The other has a productive faculty, and gilding and staining all natural objects with the colours borrowed from internal sentiment, raises in a manner a new creation. (Hume 1998, 163)

In FFT (§5.1) I argued that expressivists need to be careful with this point, at least if they want to read it as providing a psychological grounding for the bifurcation thesis. For one thing, Hume himself is here characterising belief, or reason, in terms of what I called their *semantic* properties – their link to truth and falsity. Within the dialectic of Part I, this was presented as a difficulty for

an expressivist seeking to ground the factual/non-factual bifurcation in psychology. If Hume is to be our guide, the psychological distinction seems to be grounded in semantics.

Even more importantly, tying expressivism to the *sentiments* makes it blind to the possibility of an otherwise exactly parallel theory, tied to other kinds of mental attitudes. In other words, it makes it blind to its family relationship to – if not to the very possibility of – a view such as credal expressivism. This is a bad kind of blindness in several senses. Theoretically, it misses something that seems to have a claim to be an important category in linguistic theory. Strategically, it isolates the meta-ethical expressivist from important potential allies.

To coin a phrase, let's call blindness to this continuity between ethical and non-ethical cases the first dogma of expressivism. I should be clear that what I take to be dogmatic is a failure to recognise a parallel between the evaluative and non-evaluative cases, or the independent existence of the latter. It would not be dogmatic, at least not in the same sense, to recognise the parallel but to choose for some reason to use the term 'expressivism' only for the evaluative case. (Allan Gibbard perhaps qualifies as someone who makes this choice, in much of his work.)

Let me give two examples of how easy it has been for recent writers to slide into this dogma, even if they actually know better. In the original edition of Blackburn's own *The Oxford Dictionary of Philosophy* (Blackburn 1994), 'expressivism' is defined as a '[t]erm used for those theories of ethical discourse that contrast ethical sentences with expressions of belief' (Blackburn 1994, 127). Blackburn himself was already a card-carrying expressivist about probability and causal necessity, by that point, of course.

My second example is from Mark Schroeder, writing in 2008. Schroeder notes some of the breadth of expressivism, at that point, but suggests that it simply reflects the broad footprint of normativity, in areas of philosophical interest.

Blackburn and Gibbard's efforts have gone so far toward making expressivism a respectable view that expressivism is now being widely applied across the 'core areas' of philosophy. As philosophers discover normative dimensions of the subjects of inquiry of epistemology, the philosophy of language, and the philosophy of mind, the standard questions of metaethics come to apply in those areas, and the standard views in metaethics are naturally brought to bear. So Hartry Field (2000), for example, has defended expressivism about the *a priori* and evaluative concepts in epistemology more generally. Matthew Chrisman (2007) has defended expressivism specifically about *knowledge*, and David Velleman and Nishi Shah (2005) now advocate expressivism about what it is for a mental state to be a *belief*, in contrast to something you merely accept. Expressivism is also a respectable view about the *content* of both sentences and mental states, and has even been defended about logic (Brandt 2001).

It is natural that philosophers would apply expressivism to all of these topics—the topics seem to involve normative issues, metaethics is the study of the metaphysics, epistemology, philosophy of language, and philosophy of mind with respect to normative issues, and expressivism is considered to be one of the main options in metaethics. (Schroeder 2008, 6)

Schroeder may be right that in some, even all, of the particular cases he mentions here, the route to expressivism goes via normativity, and expressivism about *that*. But that's not true of all the examples in the traditional Humean portfolio, or its twentieth century extensions: causal necessity, probability, lawlike generalisations, and so on. Humean expressivism about those topics begins with psychological states which are not on the side of sentiment and affect in the

first place. In my terms, the first dogma of expressivism stems from blindness to these facts about the historical origins of the view.

Hume's distinction between the 'offices of reason and of taste' also encourages what we could now label the second dogma of expressivism. This is what Williams called 'the local expressivist's *ur*-mistake' (Williams, 2010, 325), that of failing to see that expressions of evaluative attitudes might nevertheless have cognitive content. As Williams points out, this is a failure firmly rebuked by Wilfrid Sellars in the 1950s. As Sellars puts it:

[T]he core truth of "emotivism" is not only compatible with, but absurd without, *ungrudging* recognition of the fact, so properly stressed (if mis-assimilated to the model of describing) by "ethical rationalists," that ethical discourse as *ethical discourse* is a mode of rational discourse.

It is my purpose to argue that the core truth of Hume's philosophy of causation is not only compatible with, but absurd without, *ungrudging* recognition of those features of causal discourse as a mode of rational discourse on which the 'metaphysical rationalists' laid such stress but also mis-assimilated to describing. (Sellars 1957, 285)

In granting the 'core truth' of emotivism, Sellars aligns himself with Humean expressivism. His easy analogy between the case of causation and that of ethics shows that he is free of the first dogma (and thinks that Hume himself was free of it, apparently). And in insisting in both cases that expressivism does not mean expelling the discourse from the office of reason, Sellars rejects the second dogma.

As we now know, Sellars' point was to be picked up or rediscovered by leading expressivists of later generations, including both Humeans such as Blackburn and Gibbard, and Hegelians such as Brandom. Thanks to their work, the second dogma has lost some of its bite, though it still yaps loudly at contemporary heels.

Hume's productive metaphor stands up quite well, I think, even when we rid ourselves of these first two dogmas. It still directs our attention to the distinctive aspects of our psychology relevant to the kinds of judgement in question, the basis from which we project our 'new creations'. Thus the credal expressivist takes 'It is probable that P' to be an expression of a high credence that P. In talking about 'the probability that P' we have, at least in a loose sense, raised a new creation. Avoiding the second dogma opens the way to treating this new creation as something of an appropriate shape – an object of belief. However, the metaphor might mislead us into thinking that this is somehow automatic, whereas on the contrary, there's a lot of explanatory work to be done here. (Recall our discussion of the Frege-Geach point in Chapter 10.)

Let's say that by calling attention to these distinctive aspects of our psychology, the Humean model identifies a feature of the speaker that underpins, or *grounds*, the vocabulary in question. This focus on *features of speakers* that underlie vocabularies is going to be the fifth ingredient in our developing recipe for expressivism.

To develop our terminology further, let's say that such a feature constitutes the *standpoint*, or *perspective*, from which such a speaker speaks. In making aesthetic judgements we speak *from the perspective* of a creature disposed to sentiments of certain kinds. In speaking of probability, we speak *from the perspective* of a creature with credences, or partial beliefs. Creatures who didn't have such psychological states – because their beliefs were binary, on/off matters, say – simply

wouldn't be equipped to play this linguistic game. (More on this kind of consequence in a moment.)

Now to the third dogma. It is the assumption that such a ground will always be psychological. Once again, the key to ridding ourselves of this dogma is to note a family resemblance between these psychologically grounded applications of expressivism and views that are otherwise closely analogous, but not *psychologically* grounded. Someone in the grip of the dogma is likely to appeal to their linguistic intuitions at this point, objecting that the link to psychology is part of what we mean by expressivism. Rather pushing head-on against such intuitions, let's approach the point indirectly, via the metaphysics-avoiding face of expressivism.

## 11.7 Generalised perspective

As we've seen, one of the key motivations of expressivism is that it is ontologically conservative, in places where that seems to matter – where naturalist intuitions confront the puzzles of placement problems, for example. This characteristic immediately tugs us in the direction of a broader conception of expressivism. It is easy to find this kind of explanatory package – proposals claiming to explain our *talk*, without appealing to additional *ontology* – well beyond the narrow expressivist's psychologically-grounded cases.

Once again, deflationary views of truth provide an obvious example. I want to introduce a very different example, that of indexical vocabulary. I choose it, in part, because it offers such a clear view of what a non-psychological ground might look like. Here is Brandom, pointing out how natural the corresponding explanatory move seems in this case:

[C]onsider someone who is puzzled about what is represented by indexical and demonstrative vocabulary. Are there indexical and demonstrative *facts*, over and above those expressible in non-indexical terms? If not, why aren't indexical terms freely interchangeable with nonindexical ones (as the phenomenon of the essential indexical ... shows they are not)? If so, what are these peculiar items? (One might imagine here some naturalistic analogue of the theologians who worry that a deity who is not spatio-temporally located could not think the sort of indexical and demonstrative thoughts we express using 'here', 'now' and 'this'.) The fact that we can formulate rules sufficient to specify the correct use of indexicals (at least for ordinary, spatio-temporally located speakers) – including the uses that are demonstrably not interchangeable with the use of any non-indexical terms – entirely in non-indexical terms should be enough to dispel any concern that there is something spooky or mysterious going on. Some thought like this seems to be behind Wittgenstein's stories about the use of terms such as 'pain' and 'rule'. If the practices themselves are all in order from a naturalistic point of view, any difficulties we might have in specifying the kind of things those engaged in the practices are talking about, how they are representing the world as being, ought to be laid at the feet of a Procrustean semantic paradigm that insists that the only model for understanding meaningfulness is a representational one. (Brandom 2013, 86–87)

Clearly, standard expressivist views of probability or value would work equally well as Brandom's Wittgensteinian examples, as analogies for the move that Brandom is describing in the indexical case. In all these cases, a key advantage of the move is to avoid the question that confronts rival 'ontology-first' accounts – the question 'What are these peculiar items?', as Brandom puts it.

The parallel between the indexical case and familiar expressivist cases goes a lot deeper than this. As the work of Anscombe (1975), Perry (1979), Lewis (1979), Ismael (2007) and others has

shown us, indexical beliefs have a distinctive *pragmatic* roles. They connect to our practical life in particular ways, ways that any proposal to introduce ‘indexical facts’ would be obliged to explain. If there is a real ‘now’, for example, as some philosophers of time maintain, why should beliefs about it be relevant to action? Indeed, how would we *know* when the relevant fact obtained – ‘How do we know it is now now?’, as Braddon-Mitchell (2004) says.

These questions are the analogue of what in Chapter 4 I call the *confidence* and *approval* problems, for factulist accounts of probability and value, respectively. In those cases the expressivist avoids these problems by *beginning* with the feature of speakers of immediate practical relevance – credences and pro-attitudes, behaviourally characterised – and building out from there.<sup>34</sup>

Similarly in the indexical case. If we begin with the features of speakers of practical relevance – their location in space and time, and their own identity – and build those into the use conditions for indexical claims, then there is no practical gap to be bridged. As I noted in §0.3.3, Brad Weslake and I (Price and Weslake 2010) have called the requirement to which this move provides a response the *Practical Relevance Constraint* (PRC). Weslake and I are discussing the case of causation, but the constraint has much more general application, as these examples show.

In all three cases, then – or five if we count the temporal, spatial and personal indexicals separately (the differences between them will be relevant when we return to No Fault Disagreements) – we have a common structure. There is a feature of speakers or their circumstances, a feature of practical relevance to them in some way, that needs to be mentioned in an adequate account of the use conditions of particular terms or vocabulary. *Only* by mentioning that feature can we understand the practical relevance of judgements employing those terms or vocabulary, in the lives of those speakers. The feature in question will typically show up on both sides of Williams's distinction between two kinds of use – i.e., both in the ‘how to use’ rules that correct use of the vocabulary requires, and in the ‘what’s it useful for’ sense, where we are interested in what the vocabulary does for us.

I have already noted that the indexical examples make the attraction of the explanatory move particularly obvious, and that they are cases in which the relevant grounds are non-psychological. A third advantage is that the gap between indexicals and the familiar psychologically-grounded cases offers a rich vein of mixed cases. In particular, there are opportunities for analogous accounts of much of our temporal vocabulary, and of temporally-oriented modalities such as causation and chance. As I have argued at length elsewhere (Price 1996, 2001, 2004b, 2007, 2011b, 2023b; Price and Weslake 2010), these are all topics best explored in terms of the role of the concepts concerned in the lives of particular sorts of creatures – agents located in space and time, strongly oriented in time, acting to achieve their ends in conditions of epistemic uncertainty. In all these cases, in my view, expressivist accounts offer their characteristic advantages. They easily meet the Practical Relevance Constraint, and they avoid unnecessary and explanatorily ineffectual metaphysics. (More details on the case of causation in Chapter 12.)

Not surprisingly, these claims are much more controversial in some cases than in others. Even in the temporal cases, they are more controversial for some aspects of our ordinary conception of time than for others. It is easier to make a case that our sense of a distinguished present moment is perspectival, a reflection of own temporally ‘located’ standpoint, than it is to make the analogous point about the direction of time – the idea, roughly speaking, that time differs from space in being intrinsically ‘directed’.

---

<sup>34</sup> Does it follow that *only* an expressivist view could do so? Chapter 4 claimed not, proposing that it is open to factualists to take these links to be analytic. But in the richer conception of this chapter, it is arguable that any view of this kind ought to count as expressivist.

My own view is that direction, too, is a perspectival matter. As Boltzmann put it in the 1890s: ‘For the universe, the two directions of time are indistinguishable, just as in space there is no up and down.’ (1964, 446–447) Boltzmann was merely entertaining the possibility at that point, noting that it would be a natural thing to say in the context of a particular cosmological proposal. However, the idea doesn’t seem to have been proposed explicitly before this, which makes the remark a significant step in intellectual history – especially so if the proposal turns out to be correct, as I think it does.

Indeed, in my view (see Price 2011b) it is very hard to make sense of *what it would be* for time to have an intrinsic direction, at least if we want to have some prospect of connecting that fact to our ordinary dealings with time, in physics and in ordinary life. In comparison, it is very easy to explain why temporally-oriented creatures such as ourselves – all of us, as it happens, sharing the *same* orientation – should come to think of this as an entirely objective feature of their environment, just as our ancestors did with up and down. Recognition of the perspectival element has its usual Copernican advantages, avoiding a need for structure *in the world* by explaining the appearance of structure as an artifact of our viewpoint.

As I say, this view about the direction of time is controversial. Not all writers on the topic, by any means, agree with me about the bleak prospects for non-perspectival accounts. I mention the case here with two lessons in mind. First, I think it provides a very clear idea of the *potential* of the general recipe – in particular, its potential for avoiding difficulties that arise when issues are addressed in other keys. Second, it illustrates some significant points about the contemporary philosophical landscape. In a case like this part of the argument for expressivism rests on criticism of rival approaches. These rival approaches are typically defended by writers who, as specialists in fields such as metaphysics and the philosophy of physics, are likely to have little familiarity with the ins and outs of expressivism. For their part, most expressivists will have little sense of the issues within metaphysics and philosophy of physics, and hence will be poorly placed to evaluate the potential advantages of their own methodology in those fields. So the two sides tend to fail to engage – thereby missing, in my view, insights of considerable importance from both points of view.

In Chapter 12, as a more detailed illustration of the general method, I will outline the case for an expressivist view of causation. It is similarly controversial – perhaps even more so, given the central role of causation in various metaphysical projects. The two cases are related, of course, because causation is temporally directed. When we return to this topic in the next chapter I will introduce some remarks from Ramsey, that provide a characteristically perspicuous sketch of a perspectival view of causation. Ramsey’s remarks begin with the following questions about time:

It is, it seems, a fundamental fact that the future is due to the present, or, more mildly, is affected by the present, but the past is not. What does this mean? It is not clear and, if we try to make it clear, it turns into nonsense or a definition. ... [W]e think there is some difference between before and after at which we are getting; but what can it be? (Ramsey 1929, 145)

Ramsey proceeds to treat this as a psychological question: ‘What then do we believe about the future that we do not believe about the past?’ (Ramsey 1929, 145. His answer appeals to the epistemic standpoint of a deliberating agent, oriented in time. Again, much more on this in Chapter 12.

## 11.8 Pragmatic grounds and the Copernican shift

For the moment, these cases serve to emphasise what a broad church I take expressivism to be. What unites them is the Copernican shift, the recognition that what we took to be a distinctive kind of *fact* is best explained as a manifestation of a distinctive kind of *standpoint*. This common structure doesn't mean that there is a one-size-fits-all argument for applying the move. On the contrary, the details will vary from case to case – both the details of the expressivist proposal itself, and of the options for rival views. Still, unless one has some sense that there is a common pattern, it is easy to miss the expressivist option altogether, at least in some cases. So the big picture matters. It is crucial to notice the analogies between moves in metaethics, on the one hand, and moves in the philosophy of time, chance, or causation, on the other.

Some readers may feel that this is far too broad a church to lay claim to the label 'expressivism'. Once again, however, we need to distinguish the terminological point from the issue about the shape of the landscape. For the moment, I am interested in explaining what I take the commonalities in the landscape to be. With those in view, we can come back to the issue of terminology. We can also come back to the question of whether there are important boundary lines in the landscape that my sketch has missed. (I will be proposing one such line myself.)

In previous work (Price 2007), I have also explored the analogy between the causal case and some familiar cases in which descriptive concepts reflect the perspective of the speaker. When we call things hot or cold, tall or short, foreign or local, we are often relying on our own temperature, height, or tribe as a point of reference. If we formulate those comparisons as parts of use conditions, in obvious ways, then we have more cases that conform to this broad expressivist model. They provide further support that this wide range of cases offers for ridding ourselves of the third dogma of expressivism – for acknowledging that the interesting 'natural kind' for linguistic theory does not mark a boundary between psychological and non-psychological grounds, any more than it marks a boundary between affective and other sorts of psychological states.

In all of these cases, I think, we can usefully retain the metaphor of speaking *from a perspective* – of understanding a vocabulary as a 'giving voice to' a particular situation, characteristic, or circumstance of the speaker. Such vocabularies are linguistic games that creatures are equipped to play, in virtue of the fact that they possess or occupy the situation in question. In all these cases, moreover, we need to insist on the same distinction as in the probability case. Speaking *from* a situation or perspective need not mean speaking *about* that situation or perspective. The feature in question is *backgrounded*, as we might say.

This attention to the distinctive features of speakers and their situations on which expressive vocabularies depend, is what I had in mind above (§11.5), when I said that I wanted to identify one more ingredient in the expressivist recipe. It is an ingredient already implicated in the earlier ingredients, as all of our examples make clear. But because it is such a distinctive feature of the expressivist package, it is worth singling out for attention in its own right. As we'll see, singling it out in this way makes it easier to explore the connection between NFDs and the other ingredients of our recipe, and also to explain the case for global expressivism.

There doesn't seem to be an established label for this element of the expressivist package. In (Price 2004b, where I explored the perspectival metaphor in more detail, I called such features the 'contingent grounds' of particular concepts or vocabularies. In (Price 2013, 48) I called it a 'practical stance' – as I put it 'a practical situation or characteristic that a creature must instantiate, if the language game in question is to play its defining role in her life.' More recently still, I have used the term 'pragmatic grounds':

Expressivism links particular assertoric ‘vocabularies’ to particular ‘pragmatic grounds’ – i.e., to the practical features of speakers on which the use of a particular vocabulary depends. In the moral case, for example, the pragmatic grounds are (in the simplest version of the view) the affective attitudes that moral claims are taken to express. (Price 2019, 146)

I note there that a good framework to develop this idea is that of Brandom (2008): ‘When Brandom asks what one has to be able to *do*, in order to *say* particular things, this is an enquiry about the pragmatic grounds of a discourse, in my terminology.’ (Price 2019, 146, n. 27)<sup>35</sup>

As we’ve seen, the notion of pragmatic grounds enables us to make sense of the idea that vocabularies might simply be *inaccessible* to speakers who lack the relevant features or capacities. In the indexical case, Brandom mentions the thought ‘that a deity who is not spatio-temporally located could not think the sort of indexical and demonstrative thoughts we express using “here”, “now” and “this”’. It is easy to find more down-to-earth versions of the same kind of concern in other cases. As I put it in an early piece, for example:

Particular topics of conversation seem to be inaccessible to speakers who lack an insider’s view of the subject matter concerned. The familiar examples involve sensory deficiencies: discourse about music may be inaccessible to the tone deaf, wine talk to the anosmic, the finer points of interior decorating to the colour blind, and so on. The traditional secondary qualities thus provide the obvious cases of concepts which seem to exhibit this form of subjectivity—this dependence on specific and quite contingent human capacities. (Price 1991b, 80)

In a similar spirit, Dummett asks us to imagine creatures who are ‘mere observers and not agents at all—a kind of intelligent tree’ (Dummett 1964, 338). Dummett suggests, plausibly in my view, that some ingredients of our ordinary notion of causation would be inaccessible to such a creature. In our present terminology, agency constitutes the pragmatic grounds for at least a core part of our causal concepts.

However, this kind of all-or-nothing variation between speakers isn’t the most interesting possibility, for our present purposes. A more limited kind of variation is the key to No Fault Disagreement. Before we turn to this, let’s revisit our recipe for expressivism, with pragmatic grounds in the picture.

## 11.9 The ingredients of expressivism (II)

In §11.5 I described our developing recipe for expressivism like this. We had three ingredients already in the mix: use-first semantics; the idea of expressivism as an alternative to metaphysics, especially in the face of placement problems; and an explanatory programme, aiming, as I put it, to account both for the *existence* and *practical relevance* of the vocabularies in question. Sitting on the bench beside these ingredients, waiting to be associated with them, we had the framework of NFD from FFT. I said that our goal was to return to the question as to how it relates to the other three ingredients.

---

<sup>35</sup> If there is a difference between me and Brandom here, it is that not all pragmatic grounds in my sense are a matter of *doing* something. In some cases, including the indexical case, it is more a matter of *having* certain characteristics, or of *being* a certain way. However, I think we have the same notion of practical prerequisites in mind: conditions without the instantiation of which, a speaker simply doesn’t have what it takes to employ a particular vocabulary. In my terminology, the term *grounds* is intended to have this practical sense – in particular, I don’t have rational or metaphysical grounding in mind.

In pursuit of that goal, we have been focussing on the *features of speakers* – typically features of practical or ‘pragmatic’ significance – that play characteristic roles in expressivist accounts of particular vocabularies. These were already in the picture implicitly, as it were, being present in some form in each of our examples of how the expressivist recipe is used in practice. What we have done is to give this element a name, and put ourselves in a position to identify its characteristic role in the many different applications of expressivist techniques.

I think we can now see that pragmatic grounds are a vital ingredient, in two senses. First, the key difference between all the different kinds of loaves that can be baked to an expressivist recipe is the difference in the pragmatic grounds incorporated at the beginning. (This is simply the natural generalisation of the observation that the key difference between credal and evaluative expressivism is the difference between credences and pro-attitudes.) Second, if there could be a vocabulary baked to a non-expressivist recipe, it would have to be marked by *lack* of pragmatic grounds. As we’ll see in Chapter 12, the impossibility of such a thing is a key part of the case for global expressivism.

I’ve argued that the picture that emerges is a kind of perspectivalism. To bring a vocabulary within the scope of the programme is to treat it as a case in which the users of the vocabulary speak ‘from a location’ – *qua* occupiers of some particular pragmatic grounds. Reference to those grounds will show up at some point in the use conditions of the claims in question, and will typically, perhaps always, be relevant to both the two senses of use or function Williams distinguishes: both the *how* and the *why* of the vocabulary, so to speak.

I emphasise again that in advocating this very general recipe for expressivism, I am not insisting that the details will work the same way, in all cases. On the contrary, I think it likely that there will turn out to be important sub-categories – perhaps groupings of vocabularies in which use conditions work in some distinctive way, or which all employ pragmatic grounds that are themselves grouped in a distinctive category (e.g., of affective attitudes). I intend the model to allow for sub-structure of that kind, and I will propose one internal distinction myself, relating to the structure of NFDs.

In many cases, if not all, these references to the pragmatic grounds will serve as the basis to an alternative to metaphysics – that is, as an alternative to the view that the distinctive character of the vocabulary is to be understood in terms of a distinctive kind of subject matter (the important philosophical project being to understand the nature of that subject matter). Expressivism explains the distinctive character in terms of something closer to home, namely, the feature of speakers identified as the vocabulary’s pragmatic grounds.

As I noted, the Copernican metaphor is a good fit for the expressivist’s central move. Expressivism avoids the need to postulate structure *in the world* by explaining our *talk* of the structure in question as a product of our own standpoint. ‘That’s what the world looks like from this location’, as it were. As a bonus, it makes short work of the practical relevance problems that often plague its rivals. In typical cases, the expressivist takes her rival’s metaphysical postulates to be both *redundant* (not needed to explain the talk) and *ineffectual* (not capable of explaining practical relevance).

### 11.10 Two comparisons: the manifest image and secondary qualities

There are two familiar philosophical frames we might apply to this picture, though with some reservations, in each case. One is Sellars’ (1962) distinction between the scientific and manifest

images, in the general sense in which it is now widely applied. (Sellars himself applied it specifically to the scientific and manifest images of ourselves.) Expressivism can be seen as a recipe for *explaining* the manifest image, case by linguistic case. The qualification needed is that we want to hold open the possibility that our image of the world is manifest all the way down, if expressivism turns out to be necessarily a global view.

The second familiar frame is the distinction between primary and secondary qualities. This calls for a different qualification. Some proposals for generalising the notion of a secondary quality turn out to conflict with expressivism, as presented here. I'm thinking of the programme of so-called response-dependent realism (RDR). In our present terms, RDR tries to represent pragmatic grounds as part of the *content* or *truth conditions* of the vocabularies in question, rather than as part of a use condition. In Chapters 12 and 13 we'll come back to this proposal, and to the questions how it differs from expressivism, and what is at stake. For the moment, we simply need a caution. If we want to link generalised expressivism to an expanded realm of secondary qualities, we need to be careful what we take ourselves to be expanding. But with that caution firmly in place, I am happy to associate expressivism with this idea, and indeed with Putnam's suggestion, characterised as pragmatism, that Kant holds everything to be a secondary quality (Putnam 1981, 60–61). If nothing else, that's a convenient taster for global expressivism.<sup>36</sup>

Some readers may feel that associating expressivism with the manifest image, or with secondary qualities, obscures an important divide. Certainly, the content of the manifest image, and of descriptions in terms of secondary qualities, may be to some extent perspectival. But surely they involve factual matters, and vocabulary suitable for saying how things are. Wasn't the whole point of expressivism supposed to be that it wasn't like that?

This concern might come in two flavours. One would be motivated by an attachment to the idea that expressivism has something to do with a belief–desire distinction, and the thought that the manifest image and secondary qualities lie on the belief side – they have a mind-to-world direction of fit. This conception of expressivism was our first dogma, or terminological choice point. As I said, I have no deep disagreement with someone who chooses to restrict the use of the term in this way. But the programme that I have been characterising is explicit that it does not mark an important boundary at this point.

The second version of the response would link to the second dogma, that is, to the view that expressivist vocabularies cannot also be cognitive matters – the early expressivist's *ur*-mistake, as Williams put it. Again, the relevance here is largely terminological. If someone insists that the label 'expressivism' be tied for life to the *ur*-mistake, then we could choose a different term for the recipe we are now describing. (As I said earlier, 'neo-pragmatism' would be my next choice.) But this ruling seems a little harsh on those self-styled expressivists who have been using the term for decades, without making the *ur*-mistake!

Linking our general recipe for expressivism to these familiar tropes – the scientific versus manifest images, the primary–secondary distinction, and the idea of perspective – has the advantage of linking it also to familiar ideas of subjectivity, of a failure to speak 'from nowhere'. I want to embrace those links, but at the same time to reaffirm two points from FFT.

---

<sup>36</sup> Putnam himself describes the view like this: 'If *all properties are secondary*, what follows? It follows that everything we say about an object is of the form: it is such as to affect us in such-and-such a way.' (1981, 61) Global expressivists will *not* put things this way, most importantly because they want to put our reactions in the *background* in use conditions, rather than in the *foreground*, in the content or truth conditions of what is said. More on this distinction when we turn to RDR in Chapters 12 and 13.

The first point is my usual one. The analogy with these familiar distinctions should not stand in the way of the possibility that expressivism turns out to be a global view. The second point concerns order of priority. We might begin with a conception of fully objective judgement, and treat these sources of subjectivity as ways of falling short of it. As I say in FFT, I think that gets things back to front. In §7.6, I asked ‘[w]hat determines whether a disagreement about a particular sort of judgement should always be treated substantially’ – that is, as not subject to No Fault Disagreement? I responded as follows:

The following answer might suggest itself: a disagreement is or at least should be treated as substantial if it is a dispute about a matter of fact – if it answers to an objective reality, in virtue of which one party must be mistaken, as long as they disagree. As will by now be apparent, I think that this gets things back to front. I suggest that judgements of a certain kind are factual just to the extent that we do treat all disputes involving such judgements as substantial. Evaporative disagreement marks the limits of what we might characterize as the factual pattern of usage. In other words, the crucial feature of the factual pattern is that it requires us to say that of any pair of conflicting judgements, at least one must be false. (FFT, §7.6)

Applied to our present concerns, this means that I want to take our truth-norm-driven practice of disagreement to be *constitutive* of the relevant notion of objectivity, and to interpret our sense of loss of objectivity in the perspectival frameworks we have just been discussing in terms of this conception of the ideal.<sup>37</sup> Explaining how this works will lead us back to the ingredient we have yet to incorporate in our recipe, and to the issue held over from the beginning of the chapter – in other words, the relation between the framework of FFT and the rest of the expressivist recipe. Before that, we need two further distinctions.

### 11.11 Two ways for speakers to differ

Let’s distinguish two ways in which speakers may differ with respect to pragmatic grounds. We have already mentioned the difference between a speaker who possesses the ground in question and one who does not. Let’s call this an *extra-modal* difference. The speaker without the ground is simply not in a position to use the vocabulary at all, except in a kind of indirect sense, employing a use condition aligned to the circumstances of another speaker – one who does possess the pragmatic ground.

For present purposes, however, a second kind of difference is more important. I’ll call it *intra-modal* difference. It occurs when two speakers both possess what is recognisably the same kind of pragmatic ground, but differ with respect to it in some way. It shows up very clearly in the case of indexicals, though to very different extents with respect to personal, spatial and temporal indexicals. At one extreme, take Ramsey’s example: ‘I went to Grantchester’, ‘I didn’t’ (Ramsey 1929, 137). The two people involved both ‘have what it takes’ to use the first-person pronoun – they are both people! But they have it differently, so to speak, because they are different people. This is intra-modal variation.

In the case of (singular) first person indexicals, intra-modal difference is an obvious default condition. Conversational roles are normally clear, so that there’s no danger of discovering that what we took to be a single first-person voice is actually two. At least from the inside, it’s hard to

---

<sup>37</sup> Zalabardo (2023) offers a similar criterion for what it is to speak in objective mode, with the intention of representing the world, as he puts it. As I explain in (Price 2024), I find that proposal congenial, but would argue that it needs to come by degrees.

make sense of an apparent disagreement resolved by the discovery that you and someone else have been mistakenly thinking of yourselves as one person. (This case comes close. You find yourself disagreeing with what you take to be your own reflections about yourself, written some time previously. ‘I’m not a scatterbrain’, you think; ‘Why did I ever say that I was?’ You then discover that you are reading someone else’s diary, by mistake.)

For the spatial indexical, the situation is more mixed. Speakers are often sufficiently closely co-located that they occupy the same relevant grounds for these purposes – they share the same *here*, as it were. But there are counterexamples, too, and new-fangled means of conversing at a distance have made them more frequent.

For the temporal indexicals things are different again. Technology has not yet given us two-way conversation at a temporal distance, though one-way communication is commonplace. We have had it since we have had the ability to remember what our fellow speakers said in the past. When we take their sayings into account in the present, we don’t take their temporal indexicals literally, as it were, but adjust for the temporal shift. (More recently, in a case linked to the ability to converse at a spatial distance, we have learnt to adjust to the fact that not all speakers share the same clock time, or season.)

With a bit of help from science fiction, we can imagine how two-way conversation at a temporal distance might go. (Imagine that Zoom adds it in a future upgrade.) In ordinary conversation we could still proceed as though we shared the same pragmatic grounds – i.e., in this case, the same temporal location – as our fellow speakers. But we could make sense of cancelling the assumption, if an apparent disagreement turned out to rest on the fact that two speakers were at different temporal locations. Again, this would be intra-modal variation, but now with the characteristics needed to produce No Fault Disagreement. As FFT explained, two things are crucial. In the terminology we have now introduced, they are, first, that in ordinary circumstances we can proceed as though we share the same pragmatic grounds as our fellow speakers; and, second, that we be prepared to cancel the assumption, if an apparent disagreement turns out to rest on a difference in our pragmatic grounds.

With a bit more help from science fiction, we may be able to imagine the same thing for apparent disagreements about the *direction* of time. This would require an even more surprising update to Zoom, but anyone who imagines that conversation would be impossible in such a case does not know Murray Macbeath’s poignant paper, ‘Communication and time reversal’ (Macbeath 1983).

### 11.12 Two trajectories for No Fault Disagreement

The above examples illustrate how intra-modal differences in pragmatic grounds can give rise to NFDs, but they lack a feature that was important in at least some of the cases considered in FFT. To put the point in the terminology of FFT, the above examples are cases in which it can turn out in the course of a conversation that we are *not* in the same boat – we are not at the same place, or the same time, or oriented the same way in time, as we had been presupposing. Since we are not in the same boat, however, there need be no pressure for anyone to revise their view (except terminologically, if necessary, to avoid confusion).

Cases like probability are different. Here, we take ourselves to be involved in a common project. In effect, it is that of choosing the appropriate betting odds for various communal gambles that confront us. So it isn’t an option simply to recognise that we differ (because we have been using

different evidence, say), and to leave it at that. That doesn't resolve our common problem. It doesn't set a course for our common boat (or bet).

To introduce some more terminology, let's say that this is the difference between *divergent* and *convergent* trajectories for NFDs. In both cases we are inclined to say that the other party is not at fault, in the way initially assumed. They haven't got their facts wrong, as it were. But the way this gets handled is different. In the divergent case we can each continue to do our own thing, recognising that we are not at odds with one another (at least not about 'the facts'). In the convergent case, on the other hand, the conversational project calls for a shift, on one or both parts. We weren't *wrong*, but our common project cannot tolerate continuing intra-modal variation. We need to 'accommodate' in some way, to borrow Lewis's (1979) term.

This was the feature that made the NFDs of FFT unlike what happens in the indexical case. I noted in Chapter 0 that in work prior to FFT I had argued that the occurrence of NFDs in the probability case could not be explained by taking them to make implicit reference to the speaker's own evidence. That doesn't get the dispute behaviour right, because it doesn't explain the pressure to agree.

In FFT I describe the difference between the probability case and the indexical case like this:

There is an important difference between the two cases, however. Imagine for example that you phone me and say 'It is snowing here today.' I will disagree with you if and only if I think it is not snowing where you are; and agree if and only if I think that it is snowing there. But in a probabilistic case I may disagree with you even if I think your evidence does support your assessment of the probability at issue; and agree with you even if I think that it does not. In other words, a disagreement may exist in the probability case, even if both parties acknowledge that neither has incorrectly assessed his or her own evidence. There is no analogous possibility in the usual indexical case. (FFT, §7.8)

This was material derived from (Price 1983a), which also proposed an explanation of the difference:

Whether an indexical belief is advantageous to a person holding it often depends on where or when or by whom it is held. For a person fond of swimming, for example, the consequences of assent to 'It is warm here' are likely to be appropriate in Canberra in summer, but disastrous in Cambridge in winter. But there is no corresponding dependence of the consequences of a given partial belief on the evidence on which it is based. If I'm confident that Proper Name will win the 3:15, and bet accordingly, then I'll lose my shirt when he comes in last whether I've followed his form for years or picked him out with a pin. (It won't carry weight with my bookie that given *my* evidence, the horse was almost a certainty.) (1983, 404–405)

In effect, the proposal is that there is a good pragmatic reason, deriving from an instance of the Same Boat Property, why NFD should be convergent, in the probability case.

Still, I think that FFT doesn't draw the distinction between the divergent and convergent cases as clearly as it should. It would also have been helpful to have on the table at various points in my more recent work. Jenann Ismael has long taken me to task for failing to draw some distinctions related to this, between different sorts of perspectivity involved in examples such as near and far,

and local and foreigner, on the one hand, and core expressivist cases on the other (Ismael, 2016). I will return to her discussion in Chapter 12.

For the moment, it seems we are dealing with two phenomena, the second building on the first. The first is the perspectivity revealed by intra-modal variation in pragmatic grounds, real or imagined, as well as by extra-modal variation. This is the subjectivity of ‘failing to speak from nowhere’. In ‘Causal perspectivalism’ (Price 2007), I illustrated the impact of the discovery of this kind of subjectivity with a familiar example.

[T]hink of the discovery each of us made when, minds broadened by travel, we realised that foreigners themselves use the very same concept, but apply it to us! What we learnt (at that unsettling moment) was that the distinction between them and us, foreigners and compatriots, isn’t as objective as we’d assumed. It is a distinction drawn ‘from a perspective’—that of a local speech community, embedded in a tribal population. There are objective divisions in the world, of course, but not the asymmetric distinction that each side sees from where it stands. God sees us as Afghans, Zimbabweans, and many things in between, perhaps, but not as locals and foreigners. So, the reality of foreigners notwithstanding, there’s a sense in which foreignness is a less objective matter than we used to think. (2007, 250–251)

Calling attention to the role of pragmatic grounds in such a case is the Copernican shift. Very often, at least, it accompanies other elements of the expressivist package, including the avoidance of unnecessary and pragmatically ineffectual metaphysical commitment, when what we took to be an independent feature of reality is revealed to be a manifestation of our perspective. In the intra-modal cases it provides a source of NFDs, real or imaginable, though often of a kind whose natural course would be divergent.

So much for the first phenomena. The second builds on this one, with the addition of a new factor. It involves cases in which our joint conversational projects require that we *eliminate* intra-modal variation in the pragmatic grounds of our claims – that in some way or other, we resolve our NFD by shifting our pragmatic grounds (e.g., by pooling our evidence).

However, the lines here seem unlikely to be sharp. There are two lines to consider, the first between divergent NFDs and simple at fault disagreements, and the second between divergent and convergent NFDs. In both cases, we should expect it to be to some extent a conventional matter whether we allow our conversational practice to tolerate difference, without judging either party to be at fault.

For example, imagine disagreements about the timing of noon, by speakers located at different longitudes. That issue, a new one created by railways and the telegraph, resolved itself into conventions that allow no fault between time zones, but not within a single time zone.

However, the message of FFT is that it is not merely a matter of convention. There are considerations to do with the function of particular discourses that recommend, for example, what we have now labelled as convergent NFDs. Even to the extent that it is a matter of convention, such conventions need not be fixed in stone. Whether we can diverge depends on what our projects are, and with whom. An awareness of the role of our own perspective is a sometimes unsettling incentive for tolerance, in many cases.

Pulling these threads together, we have distinguished between *extra-modal* and *intra-modal* variation in pragmatic grounds, and between *divergent* and *convergent* forms of the latter. In order to

understand the relationship between an expressivism couched in terms of pragmatic grounds and the framework of FFT, we need to consider the implications of both of these distinctions for the phenomenon of NFD.

Taking extra-modal variation first, this would not be expected to give rise to NFDs, because a speaker who lacks the relevant pragmatic ground is simply not in a position to participate in the discourse in question in the first place. If agency is a pragmatic ground for causal discourse, for example, then we wouldn't have NFDs about causal matters with Dummett's intelligent trees. This is not because they would agree with us, but because, not being agents, they don't have what it takes to go in for causal talk at all (unless by piggybacking on a speaker who does possess the ground, which reduces any variation to the intra-modal case). Nevertheless, as this case illustrates, a focus on extra-modal variation, real or imagined, may be a revealing route to an understanding of the pragmatic grounds of particular forms of speech. To this extent, there is certainly a strand in the case for perspectival expressivism that doesn't align with the possibility of NFDs.

Turning to intra-modal variation, this now divides into four main cases, resulting from two distinctions: the first between convergent and divergent trajectories, in the above sense, and the second between actual and merely 'imaginable' cases. Taking the convergent cases first, embedded perspective may manifest in actual NFDs such as probabilistic cases, and in merely imaginable cases in which the possibility of NFD resulting can't be excluded in advance, and in which a convergent convention would apply. Turning to divergent cases, we again have actual cases, such as result from the discovery that two parties to a conversation do not have the same 'here' in mind, and merely possible cases, such as the discovery that two parties do not share the same direction of time.

In all these cases NFD, real or imagined, is a manifestation of the role of the perspective. However, divergent cases leave open the possibility of saying 'It was simply a terminological point, we were talking past each other.' The question therefore arises as to whether the divergent/convergent line should be taken to mark a line between genuinely factual claims and others. My inclination is to say that while this line does indeed mark an interesting distinction, we should think of it as characterised in the terms just described. We should resist the urge to try to make it do duty for a further distinction, that we have no reason to think that we actually need.<sup>38</sup>

### 11.13 FFT and Cambridge expressivism

FFT proposed an account of what the assertoric, 'factual', mode of speech as a whole is 'for' – it helps us to align commitments across a speech community, something presumably useful for social creatures. Within this framework, it sought to make sense of the idea that some claims are less than fully factual, in terms of NFDs – cases in which the normal normative game of challenge is seen as cancellable, by participants. There was the beginnings of an attempt to explain these lapses from full factuality – i.e., susceptibility to NFDs – in terms of peculiarities of some of the examples considered. Moreover, FFT argued that this NFD-based approach was the *only* way to defend the expressivist programme, in the light of what it argued to be the impossibility of drawing the factual/non-factual distinction in the way that conventional (local) expressivists had presupposed.

The recipe developed in this chapter has come to the same territory – i.e., thinking about how disagreement works in particular cases – by thinking about the dependence of vocabularies on

---

<sup>38</sup> We'll touch on this issue again in §12.3.

their pragmatic grounds. I have characterised the resulting view as a generalised perspectivalism, Copernican in something like Kant's sense. We are now interested in understanding what happens when these two routes converge.

On the one hand, I think it is fair to say that FFT offered would-be expressivists a rather meagre diet, compared to the multigrain recipe we have been describing in this chapter. FFT wasn't blind to the other ingredients, but it did maintain that a focus on NFDs did essential work in keeping expressivism from the clutches of determined factualist opponents, once the bifurcation thesis had been abandoned. That conclusion was too pessimistic. The positive sides of expressivism – its explanatory programme, its approach to accounts of meaning, and the alternative it often provides to metaphysical rivals struggling with the Practical Relevance Constraint – all these survive the kinds of accommodation with factualism on which the argument of Part I was based. In effect, this is the lesson of Creeping Minimalism. Expressivism has no reason to be embarrassed by looking much like realism, *so long as it maintains its explanatory credentials*. It is true that local expressivism has a different reason to be concerned about the failure to find an adequate ground for the bifurcation thesis, but that's an embarrassment it incurs by underreaching, not by overreaching.

This means that the NFD-based defence of expressivism is less central than FFT took it to be. Together, the other ingredients provide a robust product, quite capable of distinguishing itself from philosophical rivals. NFDs certainly provide revealing evidence of key elements of the expressivist package – evidence, in particular, of the role of pragmatic grounds – but their absence is not evidence of the *absence* of such elements. We see this clearly in cases of extra-modal variation. The difference between us and the intelligent trees does not reveal itself in an NFD, but their difficulty in making sense of causal talk is a reflection of their lack of the pragmatic grounds on which such talk depends.

Nevertheless, investigation of NFDs, both real and imagined, seems a valuable route to an understanding of pragmatic grounds.<sup>39</sup> We can 'triangulate' on pragmatic grounds, by thinking about the circumstances in which we would be inclined to say that an apparent disagreement was not a real disagreement – that the other party was not at fault. We can ask, what does it mean about our relative situations, when the case goes one way or the other? In one sense, we can see this as a matter of exploring the surface symptoms, in order to understand the underlying condition in which we are interested. But we have to be careful. According to FFT, factuality itself is a surface matter, in this sense. There is nothing more to factuality than the linguistic games of agreement and disagreement. In thinking about NFDs, we are thinking about the limits of the factual, in the only sense there is, even though we learn about what underlies the games and their limits.

I conclude that my two lines of defence of expressivism are not in tension, but nor do they line up perfectly. It would be a mistake to try to do everything 'on the surface', in the sense of FFT. But it would also be a mistake in at least two senses to ignore the framework of FFT. NFDs are an important guide to pragmatic grounds; and pragmatic grounds and the rest are nothing at all without a story about how we get from there to the assertoric mode of speech. (It is a mistake to stay too much on the surface, but we need to be clear about what lies beneath the surface, and not take it to be determinate factual contents.)

---

<sup>39</sup> Could there be a case in which NFD rested on something other than intra-modal difference in pragmatic grounds? I want to say 'no', and argue that the kind of difference in circumstances that could support NFD would automatically qualify as pragmatic grounds for the claims in question, but I'm not sure how to make this sharp.

## Chapter 12

### Generality and Causality

The previous chapter offered my broad-brush depiction of the expressivist landscape. Readers inclined to seek the devil in the details may have found it ‘maddeningly big-picture’, as a reviewer once said of one of my earlier attempts to cover this ground.<sup>40</sup> With such readers in mind, and to show in one central case what an excellent *deal* lies in the details, I’ll begin this chapter by focussing on one small piece of the landscape. I’ll present my case for expressivism in one important application, that of causation.

Causation is an important application in several senses. It is a central concern of contemporary metaphysics in its own right, of course, not least of the kind of metaphysics that seeks to take its lead from science. The place, nature, and status of causation in science are lively and contested topics. At the same time, causation is also a central element in popular philosophical approaches to other topics – causal accounts of reference, for example, or causal–functional accounts of mental states. So a successful incursion here takes expressivism to the heart of contemporary metaphysics. From my personal perspective, it is also a case in which I bring to the topic a fortuitous conjunction of interests. As I noted in the previous chapter, it is easy for potential applications of expressivism to lie off the radar of relevant specialties, from various angles.

In recent work, I have linked the case for expressivism about causation to that for expressivism about probability. There are striking and little-recognised parallels, in my view. These parallels are interesting, in part, because the factors in question are much more visible in the probability case. They turn on what I have been calling the Practical Relevance Constraint – in the probability case, the need to explain why beliefs about probability should guide decision making (e.g., choice of betting odds), in the way we take them to do so. This is FFT’s ‘confidence problem’, of course. As I noted earlier (§0.1), this issue is well recognised in the probability case, famously discussed by leading writers. In §12.2 I’ll explain how I take that discussion to map over to the causal case, where it remains almost invisible.

As I said in Chapter 11, I take expressivism about topics such as causation and probability to be interestingly analogous to more familiar cases of indexical and perspectival concepts. The latter are merely the obvious cases of something much more widespread, in my view. In §12.3 I’ll respond to some insightful criticism of my use of this analogy by Jenann Ismael. Her sympathetic critique of my claimed ‘causal perspectivalism’ will help me to clarify the view. (Ismael also notes analogies between causation and chance.) To the extent that Ismael and I disagree, much of the issue concerns meta-questions about what we take the most interesting questions in the vicinity to be. As I say in §12.4, this feels like a familiar disagreement from my perspective. I illustrate this point with an example from an early paper (Price 1991b) in which I was trying to distinguish myself from so-called response-dependent realists.

This provides a transition to the last topic of the chapter. Leaving the details of the causal case behind, I’ll zoom out again, and add a different sort of elaboration to the big picture. I’ll explain why I take expressivism to be necessarily a *global* view, extending all the way to the edges of the canvas of declarative language.

The title of the chapter is a respectful and grateful nod to Ramsey’s ‘General propositions and causality’ (Ramsey 1929; hereafter GPT), mentioned briefly in the two previous chapters.

---

<sup>40</sup> MacFarlane (2014), reviewing (Price *et al* 2013).

Remarkably, as we'll see, this single short piece is deeply relevant at both the scales we'll be considering. Ramsey not only points us in the direction of the inescapable generality of expressivism, but also puts his unerring finger on the key detail of the causal case.<sup>41</sup>

### 12.1 The case for causal expressivism

In FFT §8.10 I mention causation as one of the potential applications of the framework discussed in Chapters 7 and 8 – one of other potential ‘sources of subjectivity’, as I put it. The source of NFDs I seem to have had in mind is a kind of contextualism. Like counterfactual judgements, causal judgements are often sensitive to what we ‘hold fixed’ in considering alternative possibilities. Do we say that the spark caused the fire? Yes, probably, if we can hold fixed a normal atmospheric composition. No, perhaps, if the fire happened in a chamber in which sparks are frequent but oxygen normally absent. In the latter case, we may well say that it was the presence of oxygen that caused the fire. It is easy to imagine NFDs beginning on some such basis.

When FFT appeared, however, I was already exploring a different ‘source of subjectivity’ for causation. In our present terminology, I was interested in the idea that a crucial part of the pragmatic grounds for causal vocabulary lies in the fact that we are *agents*, creatures who manipulate their environments in pursuit of their goals. My interest in these topics had developed in the 1980s in conversations with my late contemporary, friend, and Australian colleague, Peter Menzies (1953–2015). Those conversations eventually produced a joint paper, ‘Causation as a Secondary Quality’ (Menzies and Price 1993), written in the early months of 1990. Menzies and I defend the view, as we put it, that ‘the ordinary notions of cause and effect have a direct and essential connection with our ability to intervene in the world as agents’ (1993, 187).

Menzies and I interpret this view as the claim that causation is a secondary quality, in a generalized sense of that term. We argue that this interpretation yields responses to four common objections to agency-based accounts of causation: that they confuse the epistemology of causation with its metaphysics; that they are vitiated by circularity; that they cannot make sense of causal relations between events which are outside the control of any agent; and that they make causation unacceptably anthropocentric. In each case, we point out, there is an analogous objection to a Lockean view of colour, and familiar responses to these objections map over to the causal case. We conclude that ‘provided one is prepared to countenance the possibility that causation has the character of a secondary quality, the standard objections to the agency theory appear no more forceful than the analogous arguments against orthodox accounts of colour’ (1993, 201).

Menzies and I acknowledge the common intuition that causation is ‘more objective’ than colour or taste, but claim that the agency view can make sense of this intuition, *as a matter of degree*. In the terminology of the previous chapter, our point is that – contextual factors aside, at any rate – intra-modal variation in the relevant pragmatic grounds is much less common in the case of the (idealised) agentive capacities than it is for sensory matters.

I think it is fair to say that although this paper defended important territory for an expressivist view of causation, in our present terminology, it didn't make a positive case for it. The lack of a positive case wasn't just a matter of the paper's defensive stance. Menzies and I were deliberately non-committal about the best framing for an agency view. As I have noted elsewhere (Price

---

<sup>41</sup> Both points turn on the temporal character of human thought and agency, a theme I explore at greater length in (Price 2023b).

2017c, 74), our instincts were somewhat different, Menzies' more realist than mine. At least at that time, his preference was to fit the agency view of causation, along with other secondary qualities, into the then-popular framework of response-dependent realism; whereas mine was for what I would now call expressivism. (More on the differences between these two approaches below.)

My own attempts to make a positive case for the agency view began in another paper from that period, 'Agency and probabilistic causality' (Price 1991a). This paper, too, owed much to Menzies. Some years earlier Menzies had pointed out to me that my optimistic Humean projectivism about causation was going to have to deal with arguments from his Stanford PhD supervisor, Nancy Cartwright. Then, as now, Cartwright's 'Causal laws and effective strategies' (Cartwright 1979) was widely seen as offering a powerful argument for causal realism. My (1991a) paper was the beginning of a long campaign to show, as I would now put it, that an agency-based expressivism offers a superior explanation of the facts to which Cartwright herself appeals.

Cartwright frames her argument against Russell's well-known lecture, 'On the Notion of Cause' (Russell 1913). Russell's piece is a classic defense of causal *irrealism*. It is the paper in which Russell declares that the 'law of causality, like much that passes muster in philosophy, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm' (1913, 1). One of Russell's arguments for this famous conclusion rests on the *time-asymmetry* of causation – the fact that effects are supposed to occur *later* than their causes (or at least *not earlier* than their causes). Russell claims that there is nothing to ground such a time-asymmetry in fundamental physics. Indeed, he proposes that the apparent difference between past and future is just a manifestation of a fact about us:

We all regard the past as determined simply by the fact that it has happened; but for the accident that memory works backward and not forward, we should regard the future as equally determined by the fact that it will happen. (Russell 1913, 20–21)

As I'll explain, I think that Russell was half right about this. His mistake was to pick the wrong feature of us. What matters is agency (working forward), not memory (working backward).

For her part, Cartwright begins with the distinction between laws of association and causal laws.

Laws of association are the familiar laws with which philosophers usually deal. These laws tell how often two qualities or quantities are co-associated. They may be either deterministic—the association is universal—or probabilistic. The equations of physics are a good example: whenever the force on a classical particle of mass  $m$  is  $f$  the acceleration is  $f/m$ . Laws of association ... tell how often two qualities co-occur; but they provide no account of what makes things happen. (1979, 419)

Noting Russell's contention that, as she puts it, 'laws of association are all the laws there are, and that causal principles cannot be derived from the causally symmetric laws of association', Cartwright goes on to argue 'in support of Russell's second claim, but against the first.' (1979, 419) In other words, Cartwright agrees with Russell that '[c]ausal laws cannot be reduced to laws of association' (1979, 419), but maintains that they cannot be dispensed with. Why not? *Because not all laws of association provide effective strategies for making things happen*. Barometer readings are reliably associated with prevailing weather, but moving the needle of one's barometer is an ineffective strategy for making it rain.

Cartwright argues that it is an objective matter which strategies are effective and which ineffective. Such an objective matter calls for explanation, one that it seems that laws of association simply cannot provide. Hence, Cartwright concludes,

causal laws cannot be done away with, for they are needed to ground the distinction between effective strategies and ineffective ones. ... [T]he difference between the two depends on the causal laws of our universe, and on nothing weaker. (1979, 420)

As we saw, however, Cartwright also agrees with Russell that what we find in the equations of physics are time-symmetric laws of association. Where this leaves her is concisely described by Harty Field (2003). Field agrees that Cartwright ‘makes a compelling case against Russell’s view that we should do without causal notions.’ However, he continues,

Cartwright herself draws a much stronger conclusion, a kind of *causal hyper-realism*, according to which there are causal facts *that outrun the totality of ‘noncausal facts’* (i.e. the facts that could be expressible in some language without using causal terminology). (Field 2003, 443)

Field finds this kind of realism unattractive, for familiar reasons.

Evidently there is some sort of causal fluid that is not taken account of in the equations of physics; just how it is that we are supposed to have access to its properties I am not sure. (2003, 443)

He might have added that it is mysterious why something remote from physics is *relevant to us*, in the way that Cartwright’s argument needs it to be. Nevertheless, he recognises the force of the considerations that lead Cartwright to this point:

[D]espite the implausibility of the hyper-realist picture, we have a problem to solve: the problem of reconciling Cartwright’s points about the need of causation in a theory of effective strategy with Russell’s points about the limited role of causation in physics. This is probably the central problem in the metaphysics of causation. (2003, 443)

For my part, I think it is better to say that it is one of *two* central problems, the second being the issue of the *time-asymmetry* of causation, which is central to Russell’s own critique. This is actually a bundle of two concerns: first, accounting for the difference between cause and effect; and second, explaining why this causal ‘arrow’ (from cause to effect) typically runs past-to-future, and not *vice versa*. Accounting for these features of causation is a further difficulty for the hyper-realist view, but it is a problem for all views, including views that want to regard causation as something relatively non-fundamental. However, a great strength of the agency view is that it offers a ready solution to this problem, as well to the problem identified by Field.

As I say, my argument for this conclusion begins in (Price 1991a). I say there that my project is to cover ‘similar ground to Cartwright’s [1979 paper], but in the opposite direction’ (158). I meant that Cartwright was right, in my view, that probabilities derived from laws of association don’t give us what we need in rational decision. But the solution is to move ‘inwards’, towards different, perspectival probabilities that are the proper concern of a deliberating agent, not ‘outwards’ in the direction of what Field calls hyper-realism. In other words, I thought that Cartwright was missing something important about the distinctive epistemic standpoint of agents, something that had the potential to explain the difference between effective and

ineffective strategies. (Strategies are strategies *for agents*, after all, so the agent's perspective is always in play, wherever this distinction can be drawn.)

My argument relied on two things. The first was an analogy Cartwright herself notes between her argument and that of the early proponents of so-called Causal Decision Theory (CDT). Since the mid-1970s, writers such as Allan Gibbard and Bill Harper had argued that no merely *evidential* decision theory could yield the correct recommendations in the famous Newcomb Problems (Nozick 1969; Gibbard and Harper 1981). Instead, these writers offered decision theories making explicit appeal to causal dependencies, or to related ideas, such as counterfactual conditionals. But the decision problems to which Cartwright appeals, in arguing that laws of association cannot always yield effective strategies, are themselves similar to Newcomb Problems. So Cartwright and the defenders of CDT are on a similar page. Both argue that to get rational decision right, we need to appeal to causation.

Secondly, I relied on an argument I had proposed a few years earlier (Price 1986), defending Evidential Decision Theory (EDT). (Again, it was Peter Menzies who had explained to me the challenge that CDT presented to my naive Humeanism.) Price (1986) argues that critics of EDT had missed a crucial point: evidential probabilities *judged from a deliberating agent's point of view* are properly different, sometimes, from the probabilities we derive from laws of association and similar statistical evidence.<sup>42</sup> Provided EDT uses these 'agential' evidential probabilities, I claimed, it agrees with CDT where it needs to (especially in the more realistic 'medical' kind of Newcomb Problem, that are most similar to Cartwright's examples). In other words, I argued, decision theory simply doesn't need to appeal to notions of causal dependency, in the way that the proponents of CDT had argued. In (Price 1991a) I present this defense of EDT again, and apply it to Cartwright's argument. I argue that it gives us precisely the difference we need between the probabilities that ground effective strategies and those that derive from laws of association. And it does this with no additional metaphysics. *All the work is done by the distinctive epistemic standpoint of the deliberating agent.*

I hope it is clear how this proposal fits the template for expressivism described in the previous chapter. It identifies a feature of speakers, in this case an aspect of the epistemic standpoint of any agent; and it argues that this feature is able to explain the *practical significance* of causal beliefs, on the supposition that it sits in the use conditions of causal judgements (playing the role of pragmatic grounds, in our present terminology). Indeed, the case fits the simple model of probabilistic and evaluative expressivism, at least at first pass. It identifies a psychological state, in this case a particular sort of conditional credence (among other things, one whose antecedent concerns an action currently under consideration). It takes that state to be behaviourally

---

<sup>42</sup> This analogy from (Price 1991a, 163–164) gives a flavour of the argument.

Here is a case which is in some ways analogous. Consider Scrooge, who has never received a Xmas present. He has heard of the pleasure of unexpected gifts, and as the festive season approaches would dearly love to experience that pleasure for himself. Now the crucial thing about an unexpected gift is that its prior probability be low. However, Scrooge reasons that since he has never been given a Xmas present, any present he receives will have low prior probability. He concludes that he can send himself a little token of his self-esteem, safe in the knowledge that when it arrives it will be completely unexpected.

Obviously the trick only works if Scrooge is able to ignore certain relevant information which is at his disposal. To find his self-sent present unexpected, he must ignore what he knows about its history – what he knows in virtue of which, unlike presents in general, it does not have low prior probability. Scrooge's contemplated action is attractive in the light of the belief that the gift will be unexpected when he receives it; but to perform the action would be to undermine that belief.

characterised, so that the link to decision is automatic. And it begins its account of the judgements in question with the proposal that they ‘express’ such psychological states.

As in other cases, getting from that point to full-blown causal judgement is a non-trivial matter. Among other things, it involves the quasi-realist’s project. But the crucial link with action has been secured right at the beginning. This is the expressivist’s explanation of the practical relevance of causal judgements, here offered as an alternative to metaphysical postulates of (ironically) rather doubtful efficacy.

In the early 1990s I also argued that this agency-based approach gives the best account of the feature of causation that troubled Russell, its time-asymmetry (Price 1992b). And, crucially, I soon experienced what Donald Davidson called the Ramsey effect. In other words, I realised that what I needed for an agency-based expressivism about causation, *including a more general form of the crucial insight about agentive probabilities*, seemed to be already in Ramsey. From that point, my challenge was to figure out exactly what Ramsey was getting at – as usual, Ramsey was ‘too brisk for most philosophers’, as Hugh Mellor (1991) once put it – and to try to explain it and its relevance in my own words. My first attempt to do so was a paper called ‘The direction of causation: Ramsey’s ultimate contingency’ (Price 1992c).

The phrase ‘ultimate contingency’ comes from Ramsey’s remarkable paper ‘General Propositions and Causality’ (GPC), which I have already mentioned in Chapters 10 and 11. GPC is a thirty-page manuscript dated September 1929, just four months before Ramsey’s death. Over the years, I have come to see it as an *ur*-text for modern expressivism in several senses – one of them the foundations it provides for causal expressivism.

GPC begins with a discussion of the logical form of unrestricted generalisations, such as ‘All men are mortal’. Ramsey argues against his own earlier view that a sentence of this form should be treated as an infinite conjunction, and says that ‘if it isn’t a conjunction, it isn’t a proposition at all’ (1929, 134). In other words, his claim is that these generalizations – *variable hypotheticals*, as he calls them – are not propositions, but do some other kind of linguistic job: ‘Variable hypotheticals are not judgments, but rules for judging: If I meet a  $\Phi$ , I shall judge it as a  $\Psi$ ’ (1929, 137). In our present terminology, in other words, Ramsey is proposing an expressivist view of unrestricted generalisations.

Ramsey spots an important difficulty for views of this kind. If variable hypotheticals are not judgments but rules for judging, why do we disagree about them – why do we say ‘yes or no to them’, as Ramsey puts it? As he says: ‘The question arises, in what way [a rule for judging] can be right or wrong?’ (1929, 134) Ramsey meets this challenge head-on, discussing various senses in which we can disagree with a claim of this general form. And he insists that we shouldn’t be surprised by the fact that we can disagree about something that isn’t a proposition. On the contrary: ‘Many sentences express cognitive attitudes without being propositions, and the difference between saying yes or no to them is not the difference between saying yes or no to a proposition.’ (1929, 137). At these points, Ramsey is anticipating Sellars’ (1957, 285) insistence that an expressive vocabulary can also be a rational vocabulary, as well as the quasi-realist project of explaining how this can be the case.

About two-thirds of the way through the manuscript, Ramsey turns to our present concerns: the issues of the difference between past and future, and the direction of causality.<sup>43</sup> In a few short

---

<sup>43</sup> As I note in (Price 2022a, 2023b), Ramsey may have got this puzzle from Eddington. Eddington’s book *The Nature of the Physical World* was published in November 1928, and was soon a bestseller, reprinted several times by the summer of 1929. Ramsey certainly read it; his notes survive among his papers. Eddington (1928, Ch. 14)

paragraphs he dissects his way to the core of the problem, and shows us what he takes to lie at its heart.

It is, it seems, a fundamental fact that the future is due to the present, or, more mildly, is affected by the present, but the past is not. What does this mean? It is not clear and, if we try to make it clear, it turns into nonsense or a definition ...

[W]e think there is some difference between before and after at which we are getting; but what can it be? There are differences between the laws deriving effect from cause and those deriving cause from effect; but can they really be what we mean? No; for they are found *a posteriori*, but what we mean is *a priori*. [The Second Law of Thermodynamics is a posteriori; what is peculiar is that it seems to result merely from absence of law (i.e. chance), but there might be a law of shuffling.]

What then do we believe about the future that we do not believe about the past; the past, we think, is settled; if this means more than that it is past, it might mean that it is settled *for us*, that nothing now could change our opinion of it, that any present event is irrelevant to the probability for us of any past event. But that is plainly untrue. What is true is this, that any possible present volition of ours is (for us) irrelevant to any past event. To another (or to ourselves in the future) it can serve as a sign of the past, but to us now what we do affects only the probability of the future.

This seems to me the root of the matter; that I cannot affect the past, is a way of saying something quite clearly true about my degrees of belief. Again from the situation when we are deliberating seems to me to arise the general difference of cause and effect. We are then engaged not on disinterested knowledge or classification (to which this difference is utterly foreign), but on tracing the different consequences of our possible actions, which we naturally do in sequence forward in time, proceeding from cause to effect not from effect to cause. We can produce A or A' which produces B or B' which etc. ...; the probabilities of A, B are mutually dependent, but *we* come to A first from our present volition. ... In a sense my present action is an ultimate and the only ultimate contingency. (GPT, 145–46)

It is clear that Ramsey intends to deal with both the apparent difference between past and future, and the difference between cause and effect, in terms of our epistemic standpoint as deliberating agents. On the former matter his answer is like Russell's, in that he attributes the apparent difference between past and future to a difference in us, not a difference in the world; but unlike Russell's, in that the feature he picks is agency, not memory. As Ramsey himself puts it, his proposal is a matter of 'psychological analysis', not metaphysics. (As we saw in Chapter 10, he appears to wave aside the objection that the important questions might be a matter of metaphysics.)

Ramsey's intentions aside, does the proposal work? Indeed, what is the proposal, precisely? In particular, what does Ramsey mean by 'ultimate contingency'? In my view, what Ramsey has in mind here is a generalisation of a familiar conflict between knowledge and free choice. It doesn't make sense to take yourself to be choosing between options one of which you already know to obtain. The point emerges in the oddity of saying something like, 'I know I'll vote for Biden, but I'm still making up my mind.' You take away with one hand the authority you bestow with the other, much as in Moore's paradox, when you say 'P, but I don't believe that P.'

---

discusses causation and the apparent difference between past and future, including the puzzle why causation seems to work only in one direction (even though the underlying physics is time-symmetric). Eddington doesn't have a solution, but he ventures two possible approaches. Ramsey's phrasing in GPC suggests that he is briskly dismissing one, while putting his finger on the crucial element needed for the other. See (Price 2022c) for further details.

Ramsey is generalising the point from knowledge to credence, so that it becomes the claim that an agent cannot take herself to have authoritative evidence about what she is going to do, as she deliberates. In other words, it is what has since become known as the thesis that *deliberation crowds out prediction* (DCOP). As Isaac Levi puts it, ‘deliberation crowds out prediction so that a decision maker may not coherently assign unconditional probabilities to the propositions he judges optional for him’ (Levi, 2000, 394). It follows from DCOP that from an agent’s point of view, a contemplated action must be regarded as probabilistically independent of anything to which she does assign an unconditional credence, even if *other people* (or she herself *at other times*) could legitimately take something of that kind as evidence about her choice, or *vice versa*. In my view, this independence from prior evidence is what Ramsey means by ultimate contingency.

But DCOP is controversial. Defended in its contemporary form by Spohn (1977), Levi (1986, 2000), and others, it has been criticised by Rabinowicz (2002), Joyce (2002), Ahmed (2014), and Hájek (2016). In recent work, motivated in part by the desire to find secure foundations for Ramsey’s ‘ultimate contingency’, Yang Liu and I have set out to clarify DCOP, and to defend it (or a version of it) against these criticisms (Liu and Price 2019, 2020). We link what we take to be the heart of DCOP to the so-called ‘transparency’ of first-person present-tense reflection on one’s own mental state. This is the idea that if asked whether you believe it is raining, you check the weather, not your own epistemic condition. (We rely on Richard Moran’s insightful (2001) discussion of this principle.)

Putting the pieces together, this rather subtle exploration in philosophical psychology claims to fill out what Ramsey means by ‘ultimate contingency’, and hence to provide foundations for the distinctive agential evidential viewpoint needed in my response to Cartwright. In effect, it characterises the pragmatic grounds of causal judgement, understood in expressivist terms.

For the response to Cartwright, the crucial point is that evidential dependencies are *different* from an agent’s perspective, as she considers possible actions, than they are from a third-person perspective. As Ramsey puts it:

What is true is this, that any possible present volition of ours is (for us) irrelevant to any past event. To another (or to ourselves in the future) it can serve as a sign of the past, but to us now what we do affects only the probability of the future. (GPT, 145)

This formulation can be improved in two respects, I think. First, it is not only past events with respect to which deliberation breaks evidential dependencies. If I am deliberating about whether to move the needle of my barometer, I should not take my choice to provide evidence about present or future weather, or vice versa. Second, Ramsey has not given us a reason to be sure that *everything* in the past is off limits in this way. As I have argued at length elsewhere, the possibility that some of it is not turns out to be interesting in making sense of quantum theory (see Price 1996, Price and Wharton 2015).

So a better way to put things is to say that deliberation breaks *some* evidential links, and to leave the location of the boundary as a matter for empirical investigation. What we are investigating, in effect, is the structure of evidential dependencies that survive from a point of view of an agent that takes its own actions to have no evidential ‘upstream’, within the domain of interest. We get to the crucial distinction between effective and ineffective strategies not by *adding something* to bare associations in the ontological realm, but by *taking something away* from the import of those associations in the epistemic realm. Ramsey’s insight is to see how deliberation always does the subtraction for us.

As we saw, Ramsey himself links his ‘psychological analysis’ of the agent’s viewpoint to Russell’s and Eddington’s issues: the time-asymmetry of causation and the apparent difference between past and future. I have returned to these issues, too, again aiming to vindicate Ramsey’s account. In particular, Brad Weslake and I have defended this approach against other recent proposals (Price and Weslake 2010). I have also tackled the issue of the direction of time directly (Price 2011b, 2023b), defending the Boltzmann view: ‘For the universe as a whole the two directions of time are indistinguishable, just as in space there is no up or down.’ (Boltzmann 1897, 412) As I noted in Chapter 11, I take this to be another application of the same expressivist perspectivalism.

In both cases, the direction of time and the direction of causation, the most interesting rival views try to account for the asymmetries concerned in terms of entropy and thermodynamics. As Ramsey knew from Eddington, this is the one significant place in which modern physics is *not* time-symmetric, in the sense that Russell takes to be a problem for a realist view of causation. However, Ramsey dismisses the second law of thermodynamics as a potential basis for the difference between past and future on the grounds that it is not sufficiently *a priori*. Contemporary proponents of such approaches would no doubt argue that the low entropy past has the status of a law of nature, which is good enough; and that Ramsey’s ‘psychological analysis’ is a poor sort of substitute for *a priori* truth. My view is that we do better to settle for a good explanation of a kind of perspectival *a priori* – *a priori* for creatures like us, as it were – than to hold out for something more fundamental. Once again, this is the progressive, Copernican programme, in my view.<sup>44</sup>

I have mentioned the key role that collaborating with Peter Menzies on ‘Causation as a Secondary Quality’ played in my thinking about these topics. Happily, our paper turned out to be timely. The following decade was very fruitful for the investigation of links between causation and manipulation, producing important work from Judea Pearl (Pearl 2000), Peter Spirtes and his collaborators (Spirtes *et al* 2000), and Jim Woodward, amongst others. In Woodward’s case, this work led up to his acclaimed book, *Making Things Happen* (Woodward 2003). There, and in more recent papers, Woodward devotes space to distinguishing his view from earlier manipulability approaches, including particularly that of Menzies and myself. Among other things, he stresses the importance, in his view, of basing an account of causation on the notion of what he calls *intervention*, rather than on human agency.

Despite some remaining differences with Woodward (see Price 2017c), these were very welcome developments, from my point of view. The move to ‘interventionism’ in the theory of causation is a massive shift in the direction of expressivism, in my sense. And Woodward’s careful development of his account seems to me precisely the kind of work that expressivism needs, to *really* fill in the details. In particular, he pays careful attention to the question as to how creatures like us make a transition from a purely egocentric sense of their own agency to an objectified causal model, which will be a crucial part of the project whose beginnings we find in Ramsey, Sellars, and Blackburn.

Moreover, Woodward himself is clear that he is not in the business of providing a metaphysics of causation, in a traditional sense. He emphasises that interventionism does not provide a reductive account, because intervention itself is a causal notion. For this reason, and because Russell’s, Cartwright’s and Field’s issues have not gone away, the expressivist insights seem to me to be as relevant as ever. As I put it recently, I think that Woodward’s work can be read as a

---

<sup>44</sup> For my objections to this thermodynamics-based approach, see (Price 1996) and (Price and Weslake 2010), in the case of causation; and (Price 2011b) in the case of the direction of time.

masterly vindication of the viewpoint ‘whose colours Peter Menzies and I nailed to the mast.’ (Price 2017c, 97).

I said earlier that one of the interesting things about this case is that causation is such a central notion in metaphysics, both as a subject of enquiry in its own right, and as an ingredient in much else. It is a particularly crucial piece of territory. However, my sense is that this centrality translates not into particularly careful defense of the territory from expressivist rivals, but into a difficulty in understanding how there could be such rivals in the first place. In the face of this I have tried to make progress sideways, by calling attention to analogies between the case of causation and that of chance, and exploiting the fact that there is a very well-known pragmatist strand in mainstream metaphysics in the latter case. Before we return to more general issues, let me describe this analogy.

## 12.2 Expressivism and probabilistic subjectivism

One way to characterise my disagreement with Cartwright is in terms of a so-called Euthyphro question about causation. Is it the causal connection between A and B that makes it rational to do A to achieve B (or, in Cartwright’s own terms, that makes doing A an effective strategy for achieving B)? Or does the connection go in the other direction? Does the rationality of doing A to achieve B somehow *constitute* or *ground* the fact that A causes B? Cartwright argues for the former view, I argue for the latter. Borrowing some terminology from the case we are about to discuss, we might call Cartwright’s view *objectivism*, and mine *subjectivism*.

The analogous question is much more well-known in the case of chance, or objective probability. Is it the fact that it is probable that P that makes it rational to hold a high credence that P? Or is the fact that it is probable that P somehow *constituted by*, or *grounded in*, the fact that it is rational to be confident that P? Here the labels ‘objectivism’ and ‘subjectivism’ are already familiar.

In this case the debate goes back a long way. One of Ramsey’s objections to Keynes’ theory of probability as a logical relation seems to provide an early example. As Ramsey puts it, such relations ‘would stand in such strange correspondence with degrees of belief’ (notes quoted in Misak 2020, 114). It is no surprise to find Ramsey on the subjectivist side, of course, but one of the interesting things about probability is that we find distinguished metaphysicians there, too. One of them is D H Mellor, who defends a version of the subjectivist option, calling it ‘personalism’. Following Kneale (1949), Mellor insists that personalism is compatible with the view that chances are real and objective – it is just that in saying what they are, we need to begin with rational degrees of belief, or credence.

[C]an we not analyse full belief that the chance of heads on a coin toss is 1/2 without reference to some supposedly corresponding partial belief that the coin will land heads? The reason for denying this is the fact to which Kneale himself draws attention (p. 18) “that knowledge of probability relations is important chiefly for its bearing on action”. It follows as Kneale says (p. 20) that “no analysis of the probability relation can be accepted as adequate ... unless it enables us to understand why it is rational to take as a basis for action a proposition which stands in that relation to the evidence at our disposal”. Similarly with chance. It must follow from our account that the greater the known chance of an event the more reasonable it is to act as if it will occur. This concept of a quantitative tendency to action is just that of partial belief [i.e., credence] as it has been developed by the personalists. It is thus available to provide in our account of chance that necessary connection with action on which Kneale rightly insists. A great difficulty facing other objective accounts of chance, notably the frequency theories, has been to

build such a connection subsequently on their entirely impersonal foundations. (Mellor 1971, 3)

Even more prominently, David Lewis came to similar conclusions, a few years later. Like Mellor, Lewis takes chance to be real and objective, but takes it to be definitive of chance that it plays a distinctive role in guiding credence. He insists on the importance of beginning with the subjectivist's materials. As he says, he is 'led to wonder whether anyone but a subjectivist is in a position to understand objective chance?' (1986 [1980], 84). Returning to this theme in later work, he criticizes rival approaches on the grounds that they pay insufficient attention to this connection between chance and credence: 'Don't call any alleged feature of reality "chance" unless you've already shown that you have something, knowledge of which could constrain rational credence,' he says (1994, 484).

Lewis gives a name to the principle that knowledge of chance properly constrains rational credence. In view of its centrality to an understanding of chance – 'this principle seems to me to capture all we know about chance', as he puts it – he calls it the Principal Principle (PP). His original formulation of PP is as follows:

Let  $C$  be any reasonable initial credence function. Let  $t$  be any time. Let  $x$  be any real number in the unit interval. Let  $X$  be the proposition that the chance, at time  $t$ , of  $A$ 's holding equals  $x$ . Let  $E$  be any proposition compatible with  $X$  that is admissible at time  $t$ . Then  $C(A|XE) = x$ . (1986 [1980], 87)

Not all theorists of objective probability, by any means, agree with Kneale, Mellor, and Lewis about the need to make the connection to credence definitive in this way. As David Papineau notes,

many philosophers in this area now simply take it to be a primitive fact that you ought to weight future possibilities according to known objective probabilities in making rational decisions. ... It is not just that philosophers can't agree on the right justification; many have concluded that there simply isn't one. (Papineau 1996, 238)

Papineau himself is not endorsing this view. On the contrary, he wants to argue that a justification for the Principal Principle can be found in Everettian quantum theory. So he agrees that this is not something we should simply treat as primitive.

As I noted earlier, the problem of justifying the Principal Principle is simply the probabilistic case of the Practical Relevance Constraint. Probability is interesting because the constraint is widely visible here, influential in the views of major figures such as Mellor and Lewis. Does this mean that Lewis and Mellor should be regarded as expressivists about chance, in my terms? That inference would be a little swift, and it would take us too far afield to explore the obstacles to it. But I do think it is fair to say that Mellor and Lewis are pragmatists, in one important sense of the term. The link with credence, and hence with the use we make of probabilistic judgements, is central in their views, in a way that isn't true of many accounts of objective probability. (This difference is reflected, of course, in their advantage concerning the Practical Relevance Constraint.)

In recent work I have called attention to a close analogy between the case of chance and that of causation (Price 2012; Price and Liu 2018). I have proposed that Causal Decision Theory be seen as a close analogue of Lewis's Principal Principle; as a principle connecting beliefs about causation to rational decision, in the way that the Principal Principle connects beliefs about

chance to rational decision. I have also argued that the lesson Lewis rightly draws for chance is needed for causation, too. An adequate account of causation needs to be subjectivist, in the same way that Lewis's account of chance is subjectivist. Failing that, an account of causation cannot meet the Practical Relevance Constraint. (In particular, it cannot explain why it is rational to follow CDT rather than EDT, in cases in which they conflict.)

This view has the consequence that CDT cannot be seen, as it is usually intended, as an explication of rational decision in terms of causal dependency. On the contrary, we need a prior notion of rational decision, for the kinds of decision problems in question (roughly, those allowing dependence of outcomes on acts). Hence, this is in an important sense a vindication of EDT, too, so long as EDT is construed in terms of agentive evidential probabilities. It is Cartwright's programme in reverse, as I put it in (Price 1991a): we build causal judgements on effective strategies, the latter grounded in evidential terms. As Yang Liu and I put it more recently, '[c]ausal relations become objectified relations of conditional credal dependence, the latter understood from the agent's point of view' (Price and Liu 2018, 178).

### 12.3 Is causation really perspectival?

Another recent writer who links the cases of causation and chance – though from a somewhat different orientation from mine, in both cases – is Jenann Ismael. Ismael does this in a (2016) paper criticising my 'perspectival' account of causality. She argues that there is more difference than I allow between the indexical and perspectival concepts, on the one hand, and causation, on the other. If she is right, then the effect is to call into question the very broad conception of expressivism offered in Chapter 11.

As we'll see, Ismael's criticisms link to an issue we discussed in §11.12. We saw that in cases of *divergent* NFD (real or imagined), it seems open to us to say that the parties concerned are simply talking about different things, in the way that seems so natural in the indexical case. I noted, but didn't recommend, the suggestion that the boundary between the convergent and divergent cases might therefore be taken to mark a version of a factual/non-factual distinction (resuscitating the bifurcation thesis and confining expressivism to the convergent cases).

How does this relate to Ismael's discussion? Her criticisms rely, as she says, on a 'distinction between different ways in which a class of concepts can be said to depend on facts about their users' (Ismael 2016, 245). She identifies '[t]hree importantly different forms of dependence' (245). The distinctions turn on different ways in which truth conditions for beliefs employing the concepts in question ('x-beliefs', as Ismael puts it) relate to the language user or context in question. The first of these – the one that she goes on to argue is plausibly characteristic of causal beliefs – she calls '(1) Pragmatic dependence on us' (245). In this case, 'truth conditions for x-beliefs' can indeed be given by a function that makes 'no explicit reference to human agents'; but *which function it is*, from a space of similar options, does depend on facts about us: '[W]hat explains the distinguished role [that that function] plays in our practical and epistemic lives are facts about us' (245).

Ismael contrasts this kind of dependence with two others:

- (2) Implicit relativization: truth conditions for x-beliefs are relative to agent or context; the context supplies the value of a hidden parameter ('hidden' in the sense that it is not explicitly represented in the surface syntax) that determines the truth of x-beliefs. (3)
- Indexicals: like implicit relativization except that the surface syntax contains a term whose semantic value is context-dependent. (245)

She goes on to argue that the claims of my approach to causation

are best understood in the first way. This [draws] a crucial disanalogy with [Price's] central examples of perspectival concepts, but [refines] the thesis in a way that is more faithful to what his arguments show. (245)

As she says:

One can be persuaded that the direction in which we see causal dependence as running depends on facts about us (metaphysical contingencies from a cosmic point of view) without agreeing to the claim that the semantic value of claims about cause and effect vary with those facts. It is a further claim to say that our causal concepts have the extra, implicit argument place and that their truth conditions vary with context of use in a manner analogous to paradigm cases of perspectival concepts. (252)

Ismael takes my use of a claimed analogy with cases such as near and far, local and foreigner, to commit me to such a claim about the truth conditions of causal claims. In place of this, she offers a view of causal claims as 'frame-dependent'.

The idea that causal influence runs from past to future is an artifact of the epistemic and practical lenses through which we view the world. These determine what we regard as open and what we regard as fixed. We can think of the psychological context as a frame of reference because it introduces epistemic and practical distinctions that reflect facts about the viewer's relationship to events in the field rather than among the events themselves, and that structure how viewers see the events that fall within their shared field of vision. (264)

How does this differ from the view she ascribes to me? This is how Ismael puts it:

By leaving reference to humans or human agency out of the truth conditions, such an account makes beliefs about causes beliefs *about* the world rather than about us. It captures a kind of dependence on the human viewpoint, but it puts the reference to the human viewpoint in the pragmatics not the semantics. (262–263)

I hope it will be clear to readers here that this proposal is highly congenial, from my perspective. The general account of expressivism sketched in the previous chapters is all about pragmatics, not semantics.<sup>45</sup> In most respects, then, Ismael and I are on the same page. Still, I think it is worth paying attention to the remaining differences. For one thing, her discussion makes it clear that the analogy I attempted to draw in (Price 2007) – between causation, on the one hand, and perspectival concepts such as near and far, on the other – has left my view of causation open to misinterpretation. This is certainly a concern in the present context, where I have again argued for important continuities between the indexical and perspectival cases and that of causation. So, in order to forestall misinterpretations, I want to explore the points where Ismael and I differ.

These differences are of two kinds. The first is that in the indexical and perspectival cases, I think there is more to be said than Ismael allows on the *pragmatic* side (and indeed on the *metaphysical* side, too, in a sense of the term that Ismael herself uses to characterise the insights of this kind of view of causation). In other words, my point about those cases is not simply, or even

---

<sup>45</sup> At least if the term 'semantics' is understood in terms of truth conditions and the like. In the broader sense of the term noted in §11.2, *use-first* semantics is among my ingredients for expressivism.

primarily, a matter of *semantics*. There, too, I'm interested in the pragmatics, and in the avoidance of unnecessary metaphysics. The second difference is that I think there are cautions needed about claims about truth conditions in the causal case, even in Ismael's first sense (the one she calls 'pragmatic dependence on us'). Taking these two points together, I regard the distinction between the perspectival cases and the causal case as less clear cut than Ismael takes it to be.

Concerning the first difference, recall some of the key ingredients in our expressivist recipe. It was an account based on the identification of *pragmatic grounds* for target vocabularies, offering an *alternative to metaphysics*, and explaining the *practical relevance* of the concepts in question. This is exactly what we get from standard treatments of the indexical and perspectival cases. Thus, as we saw in §11.7, we explain the special *pragmatic* significance of the essential indexical for temporally and spatially located speakers, without populating the world with something that would enable us to answer 'Yes' to Brandom's question, 'Are there indexical and demonstrative *facts*, over and above those expressible in non-indexical terms?' We have explanatory *pragmatics*, without additional and unnecessary *metaphysics*, with spatial and temporal location playing the role of pragmatic grounds.

Similarly in my example of foreignness. For better or worse, the fact that someone else is a foreigner often makes a practical difference to us. Hopefully, most of us realise that that's not to be explained in terms of their having an intrinsic property, that of foreignness, that we lack. My example trades on the fact that we are all too familiar with fellow citizens who fail to appreciate this point. But properly understood, foreignness is a perspectival matter, a view from the frame of our own tribe. (Leave aside the question whether that justifies the practical differences in question.) Again, then, we have explanatory pragmatics, without additional and unnecessary metaphysics, founded on the facts about speakers' tribal 'locations' that provide pragmatic grounds.

These are the lessons that I want to carry over to the causal case and others, and they are lessons to which a narrow focus on truth conditions risks being blind. This is not to say that the standard truth-conditional story is *mistaken* in these indexical and perspectival cases, only that it is not the full picture. Among other things, it is not the full picture about *metaphysical* questions in the vicinity. Ismael represents my disagreement with Woodward as a matter of whether causal claims have truth conditions, suggesting that Woodward wins if I allow that they do. But these indexical and perspectival claims have truth conditions, and yet that's not a victory for the kind of realism that thinks of 'here', 'now' or 'foreign' as labels for distinctive kinds of properties or facts. Remember that my primary target in the causal case is Cartwright's hyper-realism (as Field calls it), not Woodward's quest for truth conditions. As I put it: '[T]he reality of foreigners notwithstanding, there's a sense in which foreignness is a less objective matter than we used to think.' (Price 2007, 251) That's the lesson I want for causation, too, and it doesn't require denying that causal claims have truth conditions.

This brings me to the second difference. It turns on the fact that I am less willing than Ismael to grant that there are determinate *non-deflationary* facts about truth conditions, in cases like that of causation. My hesitation turns in part on reservations about the Lewisian programme for metaphysics, the so-called Canberra Plan, reservations that I mentioned in Chapter 11. Ismael invokes this Lewisian programme, introducing the term 'the Causal Role' in order to do so. As she puts it: 'We can summarize the role that causal beliefs play in practical reasoning as the *Causal Role*.' (250) Note that by 'Causal Role' Ismael does not mean the causal or functional role of *something else* (e.g., a mental state), which is the typical kind of definite description at the heart of the original Lewisian programme for theoretical identification. Rather, she has in mind the conceptual role of *causation* (or our *causal concepts*) themselves. This is a use that has its home in an

application of the Lewisian programme *to causation*, not in a use of causal or functional concepts in an application of the programme to something else.

Thus, in an explicit Canberra Plan treatment of causation (as defended, for example, by Menzies 1999), what Ismael means by the Causal Role is what is summarised in the Ramsey sentence formed from a conjunction of our core platitudes about causation. What ‘plays the Causal Role’ is what satisfies or makes true that Ramsey sentence.

Ismael says that in the kinds of cases of imagined time-reversal that I use to argue for the perspectival character of causal judgements, this account leaves open an important question:

I am going to grant the physical possibility of creatures in whom the Causal Role is played by relations that are the temporal reverse of relations that play that role in us, and focus on whether and in what sense this means that the direction of cause is perspectival. For simplicity, we’ll imagine that their causal notions are a simple temporal reflection of ours so that when we say ‘A causes B’, looking at the same events, they say ‘B causes A’. But the fact that there might be creatures in whom the Causal Role is played by something time-reversed, relative to the relations that play that role in us, does not settle whether the time-reversed relations are *causes*. They can use words in whatever way they like; the question is whether *we* would call them causes. The reason it doesn’t settle the question is that it leaves open the possibility that what we mean by cause and effect is fixed rigidly by their role in our deliberative practices. And in our deliberative practices, causes always precede their effects. One way of putting the question is to ask whether ‘cause’ just means ‘whatever plays the Causal Role’ or whether it refers rigidly to whatever plays the role of causal relations *for us*. If causal relations are defined implicitly in this latter way, then it would not follow that there is contextual variation in the extension of the concepts (or terms) ‘cause’ and ‘effect’ akin to the contextual variation in the concepts ‘us’ and ‘them’, or ‘local’ and ‘foreigner’, or ‘nearby’ and ‘far away’. The beings in Price’s imagined scenario would simply be using words differently. (250–251)

In other words, as she summarizes the point:

The possibility of creatures in whom the Causal Role is played by relations that are temporally reversed with respect to ours doesn’t establish that the concept of cause exhibits the contextual variability of the paradigmatically perspectival concepts that Price offers as analogies. It leaves open both of these positions:

*Explicit definition:* causes for creature *x* are whatever plays the Causal Role for *x*.

*Implicit definition:* causes are whatever plays the Causal Role for us.

Which of these does Price intend? (251)

Ismael is right that I should have been clearer on this point, but my preference is to be sceptical about the terms in which the choice is posed. In my view, in fact, the Lewisian framework is doubtfully coherent in the causal case. As Menzies and I point out in our second joint piece (Menzies and Price 2009), it is far from clear what idioms such as ‘plays the role’ should be taken to mean, when causation itself is in the frame. In that context, it can’t mean Lewis’s original ‘causal-functional role’, on pain of circularity. The natural alternative is semantic role, but this commits the Canberra Plan to inflationary semantics (and the same problem occurs

there when we seek to apply the Canberra Plan to those notions).<sup>46</sup> And in any case, as I noted in Chapter 11, Stich's (1996) arguments should make us wary of the idea that any naturalistically-respectable notion of reference can deliver useful verdicts in metaphysical hard cases. (If we try to put weight on a causal theory of reference, we'll again be in danger of biting our own tail.)

In my view, the upshot is that the issue of the truth conditions of causal claims is a great deal less straightforward than Ismael thinks, at least if interpreted as a metaphysical thesis. To her charge that I leave it ambiguous which of those two readings I mean, I respond that I am not convinced that the question is well posed.

Late in her paper, Ismael picks up the Republican metaphor that Richard Corry and I develop in (Price and Corry 2007a),<sup>47</sup> and offers it as a generalisation of her notion of frame-dependence.

Republicanism is the generalization of frame-dependence that allows a 'form of life' to play the role of a reference frame and the component of the account that gives the invariant content to drop out of the picture. Methodologically, what distinguishes Republicans from perspectivalists, or those looking for the sort of semantic dependence on viewpoint we see with notions like 'us and them', from frame-dependent accounts that make the invariant content explicit, is that *they don't look for non-deflationary truth conditions*. They look for a functional story that describes the job that the target concepts play in our lives, citing facts about us and the world in that account and the human practices in which they play that role. (266, my emphasis)

This enables me to explain why the lesson I draw from the indexical and perspectival cases is more Republican than Ismael realises. First, as a methodological point, I'm a semantic deflationist, so not in the business of looking for non-deflationary truth conditions, in any of the instances under discussion. And second, even more importantly, I've stressed the 'functional story' in the indexical and perspectival cases, too.

As I noted a moment ago, I chose the *foreigner* example because, sadly, we are all familiar with people who don't realise that its standard use is perspectival. We can imagine such people proposing that 'foreigner' is fixed rigidly by our use. Concerning folk on the other side of the Channel, these people protest: "They can use words in whatever way they like; the question is whether *we* would call them [foreigners]". So, even here, the distinction Ismael wants to draw is hardly clear cut. But the more important point, in my view, is that trying to draw this distinction is a matter of asking the wrong question (by my lights). On both sides of the line, the interesting issues are pragmatic, not semantic.

---

<sup>46</sup> See §14.6 for further discussion.

<sup>47</sup> Corry and I describe the metaphor like this.

In the political case, rejecting the view that political authority is vested in our rulers by God leaves us with two choices: we may reject the notion of political authority altogether; or we may regard it, with republicans, as vested in our rulers by us. Arguably, the republican option exists in metaphysics, too. Causal republicanism is thus the view that although the notion of causation is useful, perhaps indispensable, in our dealings with the world, it is a category provided neither by God or by physics, but rather constructed by us. (From this republican standpoint, then, thinking of eliminativism about causality as the sole alternative to full-blown realism is like thinking of anarchy as the sole alternative to the divine right of kings.) (Price and Corry, 2017b, 2)

Still, Ismael is obviously right that truth conditions are unproblematic in (many of?) the perspectival cases, in a way not true of the causal case. She is also right about this feature of perspectival concepts:

Perspectival concepts have a quite specific conversational (and cognitive) function. We have a use for terms with this kind of extensional variation where we talk to people whose perspective differs from ours in the relevant ways. (256)

With FFT in my history, however, I want to insist that this line is neither fixed nor sharp. Such cases occur to different extents in different vocabularies; and, at least arguably, to *some* extent in *all* vocabularies, if we are prepared to allow some epistemically possible imaginary cases. I'm interested in the full spectrum, and for similar reasons: in particular, the light that these possibilities throw on the features of speakers that constitute these differences of perspective (i.e., on pragmatic grounds, as we have called these features). For me, the indexical and perspectival cases are not a neat class sitting by themselves, but the visible tip of an all-encompassing iceberg.

Still, our discussion §11.12 revealed a line describable in my terms that does seem to coincide with Ismael's characterisation of perspectival concepts. We drew a distinction between *divergent* and *convergent* NFDs. The former are those in which it is open to the speakers in question to agree that they were 'talking about different things', in the way for which we are already on the lookout in indexical cases. The latter are those in which the context of conversation requires that we see each other as talking about the same thing, even if the initial difference needn't indicate fault on either side. (In other words, they are cases in which although we talk to people whose *initial* perspective differs from ours, we do not resolve the issue by 'agreeing to differ' in this way.) I noted that the probability cases that first got me into this topic seem to be convergent, and that the distinction seems to align with the distinction I drew between the probability cases and indexicals.

Where does causation lie, with respect to this divergent/convergent boundary? Interestingly, I think it may reveal how the same vocabulary may go one way for one reason and the other for another. The dependence of causation on our temporal viewpoint might safely be held to be divergent. At any rate, we can safely take ourselves to be talking about causation *for us*, for the kind of reasons Ismael notes. We are not likely to find ourselves engaged in common conversational projects with creatures who do things the other way. But for the kind of causal NFDs I seem to have had in mind in FFT (§8.10), it is less obvious that divergence will be a possibility. As we noted, NFDs about causal matters may arise from varying assumptions about what may be treated as 'standard background conditions' in cases under discussion. It is easy to imagine cases in which a conversation requires that we come to agree on such matters, without there having been a pre-existing standard, by the lights of which one party or other may be held to be at fault. If so, a convergent resolution seems the likely course.

At the end of her paper Ismael characterises her disagreement with me like this:

The most substantive disagreement I have with Price is that I make a great deal more room than Price for attempts to give objective truth conditions for beliefs. I think that such accounts are valuable for a wide class of structures, which include some of his own examples—e.g., chance and causation—so long as those accounts are properly understood, and not offered as reductions of *cognitive* content, but as a way of revealing the invariant representational content. He seems to think that these attempts are misguided across the board. (266)

I think this does capture a large part of our differences, although, as I have said, I think there are additional pragmatic reasons for regarding the indexical and perspectival cases as more of a piece with causation than Ismael allows. These disagreements shouldn't be allowed to overshadow the points of agreement. In many passages such as these, Ismael captures what I take to be the basic insights of causal expressivism:

The idea that causal influence runs from past to future is an artifact of the epistemic and practical lenses through which we view the world. These determine what we regard as open and what we regard as fixed. We can think of the psychological context as a frame of reference because it introduces epistemic and practical distinctions that reflect facts about the viewer's relationship to events in the field rather than among the events themselves, and that structure how viewers see the events that fall within their shared field of vision. ... (261)

To discover that the direction of causation is frame-dependent in this sense is to discover that the idea that earlier events bring about later ones is a matter of point of view, an artifact of the epistemic lenses through which we view them, not intrinsic to the field of events but imposed by distinctions that we make because they have practical and epistemic importance to us. That is a quite astounding surprise to pre-theoretic assumptions about the world. (264)

#### 12.4 Expressivism and response-dependent realism (I)

In sum, my apparent disagreements with Ismael are mostly an artifact of trying to fit my views into a frame that I simply want to avoid. This feels like a familiar juncture, from my perspective. At many points, as I sought to identify as what I now call an expressivist, I have found myself trying to mark out a space that does not address the matters in question in terms of truth conditions at all. FFT's distinction between *analysis* and *explanation* is very much in this vein, of course. By way of transition to our remaining topics, I want to describe an early attempt of this kind.

When FFT appeared in 1988, there was growing interest in so-called *response-dependent* concepts. The key ideas were introduced by Crispin Wright and Mark Johnston, in papers that were eventually to appear in (Haldane and Wright 1993). At ANU, where I was based in 1988–89, these ideas were taken up by Peter Menzies and Philip Pettit, amongst others. Pettit characterised the notion like this.

There are many different accounts of the distinction between primary quality and secondary quality concepts. But one thing is generally agreed. Secondary quality concepts *implicate subjects* in a way primary quality concepts do not. Consider the concepts of smoothness, blandness, and redness. They are tailor-made for creatures like us who are capable, as many intelligences may not be, of certain responses: capable of finding things smooth to the touch, bland to the taste, red to the eye. The concepts, as we may say, are response-dependent. They are fashioned for beings with a capacity for certain responses and it is hard to see how creatures which lacked that capacity could get a proper, first-hand grasp of the concepts. (Pettit 1991, 587)

To me, however, it seemed that this starting point took us quickly to a fork in the road, and that Wright, Johnston, Pettit, and others writing about the topic were blind to one of the options. 'Two Paths to Pragmatism' (Price 1991b) was my attempt to call attention to the fork, and to argue the case for the other path.

The sense in which both paths lead to pragmatism is that both incorporate a subjective or ‘human’ element – as Pettit puts it, ‘a capacity for certain responses’ – into their account of the concepts concerned. The fork turns on the fact that there are two different ways of incorporating such elements. The ‘response theorists’, as I called them, put the subjective element into the foreground, into an account of the *content* of the claims concerned. Typically this proceeds by specifying the *truth conditions* for the claims in question, where these truth conditions involve the subjective element. The alternative approach, what I would now call expressivism,<sup>48</sup> puts the subjective element into the background, into an account of *when* or *how* such claims are appropriately or typically made.

I offer several arguments for preferring expressivism to response-dependent realism (RDR). Among other things, I say,

the usage approach gives a more accurate explanation of some of the subtleties of ordinary language. In particular, it makes much better sense than its content-based rival of the peculiar mix of objectivity and subjectivity we encounter in discourses subject to no-fault disagreements [in the sense of FFT]. (1991b, 88)

Later in the paper I illustrate this point with a familiar kind of a case in which speakers’ applications of a term turn out to diverge.

[C]onsider the consequences of [RDR] in perfectly ordinary cases. Canberra is a small city, with a lot of open space. By Sydney standards, it is a quiet place. But if I were to deny that Canberra is a bustling place, would I be speaking to the rather refined bustle sensitivities of Canberra residents themselves, to those of my own Sydney community, or to some other ‘normal’ class? ...

A great advantage of the usage-based approach is that it avoids these ambiguities *at the level of the meaning specification*. Roughly, Canberra residents and Sydney residents learn to use the term ‘bustle’ in accordance with the same usage condition: the assertion ‘This place is bustling’ is *prima facie* appropriate when one experiences the ‘Oh, it’s so busy!’ response. This provides the sense in which the two communities speak the same language in using the term. But it leaves room for divergence later, if it turns out that Canberra residents and Sydney residents are differently sensitive to experiencing this response. The response theorist has to try to account for this divergence in terms of a difference in meaning which—perhaps unknown to the participants—has been present in their linguistic practice all along. (The term ‘bustle’ simply latched onto different concepts in Canberra and Sydney.) By specifying meaning in terms of usage, in contrast, my kind of pragmatist is able to explain the fact that in one important sense, the term does mean the same to the members of the two speech communities—and to explain the divergence in its application by differences between the speakers which lie in the background, as preconditions for the practice in question.

There are two main arguments for leaving these differences in the background, rather than incorporating them into the content of the assertions concerned. First, it makes much better sense of ordinary usage. Naive Canberra and Sydney folk need have no sense that their notion of bustle is not universal, and to insist on such gross indeterminacies in content unreasonably detaches meaning from speakers’

---

<sup>48</sup> Post FFT, as I have said, this was already for me a global expressivism, and some of my arguments against RDR turn on its inability to handle ‘globalisation’.

understanding. ... Second, and ultimately even more telling, not all the preconditions of language can be made explicit, in the sense of figuring in the content of the linguistic expressions for which they are preconditions. This is a familiar point, related to what the Tortoise said to Achilles. Something must stay in the background, crucial and yet unsaid. Whatever it is, it might vary from speaker to speaker, or from community to community. In this way, we'll get precisely what we have in the case of 'bustle': a divergence in application which rests not on any difference in content, and not on any mistake by either party, but simply on a difference in the background. (1991b, 106–107)

The second point here was an argument that RDR cannot be considered a global programme for pragmatism, because the task of making explicit these dependencies on attributes of speakers in content conditions would be never-ending. In part I had in mind, and had mentioned a few paragraphs earlier, the reasons that FFT had offered for thinking that its NFD-based non-factualism is a view with global application – viz., the reasons that turn on rule-following considerations.<sup>49</sup> I now want to turn to those issues, in the light of the account of the ingredients of expressivism offered in Chapter 11. In Chapter 13 we'll return to 'the peculiar mix of objectivity and subjectivity', when I compare and contrast my approach to these issues to that of John MacFarlane's New Relativism, and to recent work by Crispin Wright.

## 12.5 Why go global?

In other words – returning now to the big picture – the remaining task of the chapter is to explain why I take expressivism in the sense of Chapter 11 to be a global view; in particular, why I continue to reject the bifurcation thesis, in the form that sets a limit to the scope of expressivism. I'll do this in two steps. I'll first respond to a challenge to the possibility of a global view that I noted at the end of §10.2, a version of the so-called No Exit problem for expressivism. And I'll then explain why, as in FFT, I take the rule-following considerations to be a key to the necessarily global character of expressivism.

First, then, to the No Exit point. As we saw, Dreier's characterisation of expressivism turns on the expressivist's claim that 'to explain what it is to make a moral judgment, we need not mention any normative properties' (Dreier 2004, 39). But as I said in §10.2, if this is what expressivism amounts to, how could we possibly be global expressivists? Explanations need to mention something, and the criterion doesn't seem to allow us to be expressivists about that.

This challenge shows, I think, why it is important to have all the ingredients of the expressivist recipe in view. Dreier's criterion appeals to the explanatory, and 'alternative to metaphysics', ingredients, but neglects two other crucial ingredients: the use-first approach to meaning, and the identification of pragmatic grounds. Once we have these ingredients in view, there's an obvious path to globalisation. We simply need to argue that any kind of declarative cake needs a handful of pragmatic grounds, blended into a use-first account of some aspect of its meaning.

Where to find such an argument? In effect, FFT claimed to do so in the rule-following considerations, and what they reveal about the way in which meaning depends on what are at base simply contingent dispositions to treat one thing as like another. Communication is possible because, most of the time, we are disposed to 'go on in the same way' *in the same way* – but divergence is always possible, leading, in principle, to NFD. These dispositions are themselves

---

<sup>49</sup> I also had in mind a second point, concerning the inability of an RDR-style analysis to make *everything* explicit – hence my mention of what the Tortoise said to Achilles (Carroll 1895). I returned to this point a few years later in (Price 1999).

pragmatic grounds, in the terminology we have been using, and they are absolutely global. Anything that counts as language depends on them.

These dispositions are thus an essential ingredient, without which no linguistic cake can possibly stand up. In fact, we have more than we need. It would have been enough to show that any assertoric language game needs some sort of use-first component – perhaps a different one in different games. We have shown that there is a *particular kind of use-first component* that is needed in all such games.<sup>50</sup>

I'm not the only writer to have suggested that the rule-following considerations have some sort of global import of this kind. Philip Pettit (1991) takes them to imply that all our concepts are to some degree response-dependent, and interprets this as a kind of pragmatism. As Pettit notes, his views were in part a product of discussions of Kripke's *Wittgenstein on Rules and Private Language* (Kripke 1982) at ANU in the early 1980s, in which he, Peter Menzies, and I were all involved. I mentioned these discussions in §0.1, as an influence that led me in the direction of FFT. So our similar views have a common cause, at least in part (although we disagreed at the time about whether response dependence provided the best framework within which to develop them).<sup>51</sup>

In FFT, as we have seen, I linked my conclusions to the finitistic considerations underlying both the rule-following considerations and Quine's thesis of the indeterminacy of meaning. As I put it, 'no finite level of experience can determine the application of a linguistic term to all possible cases, or exclude the possibility that two speakers – fully competent speakers, by existing community standards – will diverge in their application of the term concerned to some future case.' (FFT, §8.9) I suggested that this undermined a certain conception of factual discourse: 'To recognize ineliminable indeterminacy of meaning seems to be to recognize that our utterances are at best approximations to genuine ... statements of fact.' (FFT, §8.9)

At that point the conception of factual discourse I had in mind amounted to an absence of NFDs, but the same conclusion holds if we put the emphasis on pragmatic grounds. Put like that, it amounts to the view that no part of language escapes the kind of contingency that dependence on shared pragmatic grounds introduces.<sup>52</sup>

I noted that this point had global implications:

The indeterminacy of meaning thus seems to be the basis for a form of universal non-factualism. ... This topic-neutral form of non-factualism should not be confused with the form that Quine himself extracts from the indeterminacy thesis: the view that there are no genuine facts about meaning. ... Ignoring for a moment our scruples about the analytic viewpoint [i.e., ignoring my own insistence that the worthwhile project was explanation of linguistic practice, not metaphysics], we might say that the present view, in contrast, is that in virtue of the nature of meaning there can be no genuine facts about anything. (FFT, §8.9)

---

<sup>50</sup> I am leaving aside here some interesting objections to my use of the rule-following considerations raised by Shapiro (2021). But Shapiro proposes an alternative that does much the same job, in my view.

<sup>51</sup> Crispin Wright (1992, 211) also notes a reading of the rule-following considerations that gives them a global import of this kind.

<sup>52</sup> As I noted in §11.10, Putnam's interpretation of Kant provides a useful point of comparison here, provided that we read it in terms of expressivism rather than RDR.

But this was not by my lights an argument for global scepticism, of course. It was an argument for the bankruptcy of a certain picture of language, within which the implications of the rule-following arguments appear to be sceptical.

These finitistic considerations provide yet another link to Ramsey's 'General propositions and causality' (GPT), whose expressivist views on generalisations and causation I mentioned above. I now want to show that by exploring this link we can strengthen the claim that these considerations have global implications for language – this despite the fact that, at least on the surface, Ramsey himself is *not* defending a global view.

## 12.6 Ramsey points to the future

GPT begins with the claim that 'variable hypotheticals' (i.e., unrestricted universal generalisations) – are not, as Ramsey had previously maintained, infinite conjunctions. And as Ramsey puts it, 'if [a variable hypothetical] is not a conjunction, it is not a proposition at all' (GPT, 134). This looks like what we would now call an expressivist view of generalisations, and as we saw in §12.1, Ramsey immediately makes a start on the quasi-realist project. The sentence continues with the remark that 'the question arises in what way can it [i.e., the variable hypothetical] be right or wrong' (134). But it seems to be a local expressivist view. Ramsey is *contrasting* variable hypotheticals to other sorts of claims, that can be treated as propositions.

Ramsey's reasons for denying that variable hypotheticals are infinite conjunctions turn on finitistic considerations. He glosses the idea at one point by saying that an infinite conjunction 'goes beyond what we know or want':

A belief of the primary sort is a map of neighbouring space by which we steer. It remains such a map however much we complicate it or fill in details. But if we professedly extend it to infinity, it is no longer a map. We cannot take it in or steer by it. Our journey is over before we need its remoter parts. (GPT, 134)

He also raises a concern about how we could have grounds for believing an infinite conjunction.

The relevant degree of certainty is the certainty of the particular case, or of a finite set of particular cases; not of an infinite number which we never use, and of which we couldn't be at all certain. (134)

Richard Holton and I have argued that Ramsey's position is unstable (Holton and Price 2003). As the rule-following considerations reveal, similar 'finitistic' points apply to all of language. Indeed, Holton and I point out that the relevant issue isn't infinity as such, but merely that 'the set of true instances of a generalization normally extends well beyond the set of those instances with which any speaker or group of speakers are or will be acquainted' (2003, 329).

Plausibly, ... grasping any term involves an open-ended skill, an ability to use it in previously unencountered cases ... *[F]* or no normal term at all can a grasp of its meaning amount to grasp of a list. The required list always transcends at least our current acquaintance.

We want to emphasize two points about this conclusion. First, it does seem to amount simply to extending to linguistic terms at large the kind of concerns Ramsey has about universal generalisations. In both cases, the concern is precisely that a set of instances (in one case of conjuncts, in the other of true instances of the application of a term) goes beyond what human language users could use or survey. Second, the issue concerned is

that at the heart of the so-called rule following considerations. ... On the one hand, [the rule following considerations] confront us with the fact that the application of terms to novel cases is essential to language (and hence that in a very strong sense, grasp of meaning necessarily precedes acquaintance with the totality of relevant cases). On the other, they point out that no finite basis of acquaintance logically determines a unique extrapolation to new cases. (2003, 330)

The upshot is that if Ramsey's argument is a good one for variable hypotheticals, it is a good one for everything; and it reveals the untenability of the notion of a proposition with which GPC *contrasts* variable hypotheticals. Either Ramsey retreats to his original view that variable hypotheticals are infinite conjunctions, or he moves forward and embraces global expressivism. Holton and I suggest that Ramsey would soon have recognised this instability in his position of GPC, and would have responded, in effect, by becoming a *global* quasi-realist.

Cheryl Misak (2017) has argued that Ramsey had already taken this step, led there by his pragmatist view of belief. She makes a strong case that by 1929 Ramsey was sceptical of the Tractarian picture, and suggests that his use of the term 'proposition' in GPC should be understood as a kind of placeholder, giving him something with which to contrast the function of variable hypotheticals. Whatever Ramsey's own views, however, I'm not convinced that such a pragmatist account of belief of this kind is necessarily in tension with a bifurcated view of belief *contents*, of a kind that would vindicate local expressivism.

By analogy, consider the suggestion that global expressivism is an immediate consequence of a pragmatist view of the use of assertoric language in general, such as Brandom's or the approach proposed in FFT. While a view of this sort certainly entails a kind of global pragmatism – as we might put it, global pragmatism about *assertoric force* – it is not clear that it excludes a bifurcation at the level of *content*. Could it not be the case that while some contents lie 'downstream', constructed from expressivist raw materials in the way that quasi-realists propose, others lie 'upstream', available off the shelf to the game of giving and asking for reasons?

Responding to Misak (Price 2017a), I have suggested that both Sellars and his student Ruth Millikan might be seen as exemplifying this bifurcated option, defending a view with the following distinct components. On the one hand is what Sellars (1963, 198) calls the 'empirical or matter-of-factual' domain. Here content might be explained, *à la* Millikan (e.g., Millikan 1989, 2009), in terms of pre-linguistic analogues in the biological world; Sellars himself saw these as a route to a core Tractarian picture. On the other hand are the domains we humans build on top of this matter-of-factual core, in the kind of expressive language games that the game of giving and asking for reasons makes possible. A successful programme of this kind would yield a bifurcated view of content, despite its pragmatic account of belief and assertion.

From a different direction, we should also count Mark Schroeder as an exponent of the bifurcated view, recalling our discussion of his views in §10.6. In Schroeder's case the non-expressive domain is the one in which our beliefs have what he calls 'representational content'. Again, there doesn't seem to be an immediate reason to regard this view as incompatible with pragmatism about belief and assertion.<sup>53</sup>

---

<sup>53</sup> Another interesting comparison is with Allan Gibbard (2015). Gibbard says that he, too, is now a global expressivist. As we saw in Chapter 11 (§11.2, fn. 3), what Gibbard means by this that there are no cases in which the illuminating way to specify a use-rule for a sentence is to say that it is used to express a belief with a certain propositional content, thought of as available in advance. In other words, paraphrasing Gibbard (2015, 212), there are no cases in which we can 'just say that there's a general relation of *believing* a proposition [P], and that believing

I noted in Chapter 10 that these proposals are at least structurally similar to my own proposal to distinguish a class of e-representational vocabulary. In §10.6 I deferred the question as to why my own view does not rescue local expressivism, by supporting such a bifurcated view of content. We now have an answer, I think. What these proposals for bifurcated content are missing, and what the rule-following considerations force us to confront, is a ubiquitous feature of language: its open-ended, ‘meet the future’, character. This projection into the future has to be grounded on habits and dispositions that are simply *ours*, and these dispositions count as pragmatic grounds, in our terms. This key part of the expressivist recipe is necessarily universal.

As I said, Misak makes a strong case that by 1929 Ramsey was critical of the Tractarian picture. For the above reasons, however, I am not convinced that he had already put together all the ingredients he needed, to defeat a bifurcated account of content. As an indication of how close he gets, however, it is interesting to consider some remarks of Wittgenstein that Anna Boncompagni (2017) has cited, from a lecture delivered the day after Ramsey’s death. Boncompagni argues, very plausibly in my view, that these remarks reflect Ramsey’s influence on Wittgenstein.

Every sentence we utter in everyday life appears to have the character of a hypothesis. ...

The point of talking of sense data and immediate experience is that we are looking for a non-hypothetical representation.

But now it seems that the representation loses all its value if the hypothetical element is dropped, because then the proposition does not point to the future any more, but it is, as it were, self-satisfied and hence without any value. (Wittgenstein 1930, quoted in Boncompagni 2017, 32)

It is easy to hear this as a criticism not merely of ‘sense data and immediate experience’ but of all Tractarian ambitions for a core factual language. Thought of as bare isomorphisms, pictures just don’t ‘point to the future’ in the way that ‘[e]very sentence we utter in everyday life’ does. It may be that the point about language and the future that Wittgenstein had in mind here is not precisely the one I think we need, but it seems clearly in the right territory.<sup>54</sup>

In particular, these points seem directly applicable to explaining why the ‘new bifurcation thesis’ I mentioned in §10.5–10.6 (and compared there to Sellars’ view) does not rescue local expressivism. At least in any *linguistic* manifestation, e-representations, too, are more than bare copies of elements of our environment. As linguistic items, they must ‘point to the future’, in the

---

that [P] is standing in this relation to the proposition [P].’ In my terminology, content is everywhere downstream of use.

At the same time, however, Gibbard defends the view that in the case of naturalistic discourse, the *semantic* notion of reference, or denotation, can be identified with a *causal* notion of tracking. (In my terms, as Gibbard puts it, he wants to *identify* the i-representational and e-representational notions, in the naturalistic case.) The upshot, Gibbard claims, is that he is a *local* quasi-realist, although a *global* expressivist. In my response (Price 2023a), I argue that the position is unstable. In naturalistic cases a causal notion of reference is likely to open the door to a causal account of content, on Ruth Millikan’s lines, providing precisely the bifurcated notion of content that the global expressivist needs to deny.

<sup>54</sup> In (Price 2023b) I distinguish two points we might associate with these remarks. One is the future-looking or dispositional character of all ordinary concepts, such as *chair*; the other is the future-looking character of language itself, a point about the dispositions on our part on which the use of the *sign* ‘chair’ depends. The latter is more closely related to the rule-following considerations, and is the one I need here. But the former is also very relevant to the question of the scope of expressivism, if we approach dispositions in a pragmatic spirit.

sense of being apt for re-use – that’s the whole point of a sign. But there is nothing to project them into the future but our own habits and dispositions. Once again, then, the open-ended, future-oriented character of language ensures that even e-representational language needs to be expressive. Even if on nothing else, it relies in ways that conform to the general model of dependence on pragmatic grounds on contingent features of the *temporal* nature of its users.<sup>55</sup>

---

<sup>55</sup> I explore the relevance of such temporal factors at greater length in (Price 2023b).

## Chapter 13

### Neighbours and Rivals I: No Fault Disagreement

In the final two chapters, I want to turn to the relation between my FFT-derived expressivism and some of its philosophical neighbours and rivals. Small differences are often the interesting ones, of course, and most of the views I'll be discussing qualify on both counts: neighbours but also rivals. It is often difficult in philosophy to distinguish genuine rivalry from mere differences of emphasis, interest, and orientation. So as well as comparing my view to others, I want to raise some issues about the landscape of options. What's at stake in the choice between use-first and truth-first semantics, for example? Are they rivals, or simply in different lines of work? If rivals, how is the issue to be decided? If different lines of work, what are these different enquiries?

I will discuss three authors in some detail, across these two chapters. The first two, in the present chapter, are John MacFarlane and Crispin Wright. Both MacFarlane and Wright have discussed the linguistic phenomenon that FFT calls No Fault Disagreement, and offered differing accounts of it – differing from each other, as well as from me. I want to explain why I think that the framework of FFT does a better job. But in other respects, as we'll see, MacFarlane's and Wright's views are closer to my kind of expressivism than they might seem at first sight. In the closing section of this chapter, I'll interpret them both as quasi-expressivists, by my lights.

In Wright's case, my sense that we are near-neighbours goes back a long way. I mentioned in Chapter 10 that I regard him as an ally on the significance of Creeping Minimalism. Like me, he seemed to see the main threat as being to the traditional realist. My views on truth, too, are in some ways aligned with his. In Chapter 14 I'll begin by explaining how, when I developed FFT's account of truth in two later pieces, both had connections to Wright. Both defended his stance on the normativity of truth from criticisms by others. But I'll argue that what the stance requires is FFT's expressivist *explanation* of truth, not Wright's quest for a *metaphysics* of truth. In effect, then, I'll be defending FFT's preference for *explanation* rather than *analysis* of truth, by comparison with the views of one of the field's most interesting rivals.

This will lead us to the third author I want to discuss, Bernard Williams. Concerning truth, at any rate, Williams looks like an even closer neighbour to FFT. But there are differences. I'll argue that FFT's project is more ambitious than Williams's, in looking for explanations where he sees Davidsonian primitives.

#### 13.1 Expressivism and New Relativism

Let's begin with the connection between expressivism and the so-called New Relativism of writers such as Kölbel (2004) and MacFarlane (2003, 2007, 2014). As noted in §11.1, Mark Schroeder, reversing a suggestion by MacFarlane himself, has suggested that expressivism might be 'relativism done right' (Schroeder 2015, 25). What does this suggestion look like, from the perspective of FFT, and of my sort of global expressivism?

In the case of FFT, its close relationship to recent work on relativism has been noted by Lionel Shapiro (2014). As Shapiro says,

[t]he last decade has seen an explosion of interest in relativism in the philosophy of language. Relativist accounts have been proposed to explain discourse about knowledge, epistemic possibility, matters of taste, and contingent future events. (2014, 137)

Why have relativists been interested in these particular discourses? Because, as Shapiro explains, they share a distinctive feature.

The goal of advocates of relativist semantics has been to explain the apparently distinctive functioning of a wide-ranging class of expressions. As my examples, I will take epistemic modals such as ‘it’s possible that ...’ and so-called predicates of personal taste such as ‘tasty’ and ‘fun’. At issue is the way their use appears to have a subjective or perspectival aspect. (2014, 141)

Shapiro then points out that this ‘subjective or perspectival aspect’ is closely related to the concerns of FFT.

The key phenomenon is pointed to in Huw Price’s ... *Facts and the Function of Truth* (1988), which deserves to be recognized as a close precursor to the current discussion of relativism. This is the seeming possibility of what Price calls “evaporative disputes,” which he says “populate the margins of factual discourse” (1988: 159ff). Price himself focuses on disputes over whether something is probable, but his observations apply equally to disputes involving epistemic possibility and predicates of personal taste. (2014, 141–142)

Shapiro then gives two examples of evaporative disputes, and explains how they were characterised in FFT. I’ll reproduce Shapiro’s examples in full, because his careful schematisation will be helpful below.

- Two physicians, Dr. Adams and Dr. Brown, are discussing a mutual patient. Adams says: ‘It’s possible the patient has Lyme disease’. Brown protests: ‘no, that’s not possible! You must be misinterpreting the test results.’
- Two friends, Alice and Ben, are discussing which cheeses to buy for a party. Alice says: ‘Limburger is a tasty cheese. My sister told me so.’ Ben replies: ‘Did she really? I’ve tried it, and it’s not tasty at all!’

To all appearances, both examples satisfy the following schematic description:

- i. one party has affirmed, and the other has denied, that *P*,
- ii. both parties recognize (i), and
- iii. each takes (i) to mean that there is something *mistaken* about the other’s speech act.

Suppose the appearances are correct and (i)–(iii) are satisfied. In a familiar sense, this suffices for the parties to be *engaged in a dispute about whether P*. Thus Adams and Brown are engaged in a dispute about whether it’s possible the patient has Lyme disease, and Alice and Ben are engaged in a dispute about whether Limburger cheese is tasty.

Imagine, now, that Adams and Brown discover that only Brown has seen the most recent test results ruling out their patient’s having Lyme disease. according to Price, (ii) will continue to obtain. In particular, Brown will continue to regard Adams as having affirmed that it is possible that the patient has Lyme disease. Yet (iii) now seems to fail: each will cease to regard the other’s assertion about the patient as mistaken in any way. In particular, Brown will not merely excuse Adams’s mistake as blameless given Adams’s uninformed perspective (Price 1988: 162–163). Rather, in view of that perspective, Brown will cease taking Adams to have made any mistake, *even though* Brown takes

Adams to have described something as possible that is not in fact possible. What was formerly a dispute about whether it's possible that the patient has Lyme disease has *evaporated*—or so it has seemed to Price and his relativist heirs. (Shapiro 2014, 142)

Shapiro presents his own proposal for dealing with these cases, to which we'll return later in the chapter (§13.7). For now, I simply want to highlight where Shapiro takes the proposal to sit, with respect to what I called the distinction between use-first and truth-first approaches. Shapiro proposes that relativists are addressing the issues posed by these linguistic phenomena cases in the wrong key – as a matter of *semantics*,<sup>56</sup> where *pragmatics* does a better job.

In this context, relativism is usually taken to be, or to presuppose, a *semantic* thesis. According to relativists, understanding how some discourses function requires recognizing that speakers express propositions whose truth or falsity must be evaluated relative to parameters in addition to a possible world ...

In this paper, I propose a different way to think about how the discourses in question motivate relativism in the philosophy of language. The central thrust of relativism, I argue, can and should be understood independently of any semantic framework of relativized truth. Instead, relativism should be understood in *pragmatic* terms, as corresponding to a particular understanding of assertoric force. (Shapiro 2014, 139–140)

I have called attention to this because from my point of view, the disagreement between Shapiro and New Relativists on these points looks parallel to my own disagreement with RDR (§12.4). Returning briefly to that case will help me to explain some issues I want to discuss in what follows.

### 13.2 Expressivism and response-dependent realism (II)

Early versions of RDR were often presented as rivals to expressivism. RDR was a proposal for defending the 'factual', 'cognitive', or 'realist' character of various discourses, by putting pragmatic factors – e.g., desires, in the moral case – into the content. In effect, RDR proposed to retain factuality and realism by reading a discourse as more subject-involving than initially it seems.

By the standards of contemporary expressivism, however, RDR looks like a solution to a non-existent problem. As we have seen, contemporary expressivists don't *deny* that moral claims have truth conditions, or reject simple speaking-with-the-folk realism about moral facts and properties. On the contrary, they *affirm* these things, in the minimal sense, while continuing to insist on an expressivist functional story in the background.<sup>57</sup> I have stressed that semantic minimalism is a friend not an enemy of expressivism of this sort. And minimal semantics seems to bring minimal assertoric content, in the obvious way: the content of the belief that X is good is that X is good.<sup>58</sup>

This means that RDR needs some non-obvious notion of assertoric content, or truth conditions. And now the expressivist can question the need, or our entitlement *qua* theorists, to any further fact of the required kind, *expressed in semantic vocabulary*. The expressivist is not opposed to further

<sup>56</sup> Here 'semantics' is being used in the narrower of the two senses discussed in §11.3, of course.

<sup>57</sup> As I noted in §0.3.3, Blackburn has long made this point about his own quasi-realist position.

<sup>58</sup> Recall Mark Schroeder's lightweight conception of proposition, discussed in §10.6: '[T]he entities which are the objects of [propositional] attitudes and the bearers of truth and falsity.' (Schroeder 2015, 88)

facts in the neighbourhood, of course. She simply suggests that the right vocabulary in which to express them is one of use conditions, or something similar, not truth or content conditions. In an early comment on RDR, Blackburn himself puts the point like this.

[RDR] goes bullheaded at the issue [of meaning] in terms of finding truth conditions, whereas from the point of view of [expressivism], ... if you want to talk in these terms [i.e., in terms of truth conditions], then the best thing to say about 'X is  $\Phi$ ' in the cases considered is that its truth-condition is that X is  $\Phi$ —but this will not be the way to understand matters [i.e., to say anything interesting about the meaning of the claims in question]. (Blackburn 1993, 10-11)

In my own early criticism of RDR (Price 1991b), as I said in §12.4, I argued that the right place for the pragmatic factors was in use conditions, not truth or content conditions. I argued that the truth condition view is incoherent, *as a global view* – something has to go in the background, pragmatically presupposed but not stated, on pain of vicious regress. So RDR cannot be a model for a global pragmatism.

More locally, I also argued that use conditions make much better sense of actual usage, in real cases exhibiting NFDs. We'll come back to that issue below (§13.6), when I compare my approach to recent proposals from MacFarlane and Wright. Before that, I want to clarify the question as to what is at stake in these debates, by FFT's lights.

### 13.3 What is at stake?

We have just noted two examples of disagreements about the best way to characterise the contribution of subjective factors to the meaning of the certain classes of utterances. MacFarlane and response-dependent realists incorporate them into truth conditions, Shapiro and expressivists like me treat them as use conditions, in a broad sense. Everyone agrees that the factors in question make a contribution to the meaning of the statements concerned, in some sense. They disagree about how we should characterise this contribution. What is at stake here, from FFT's perspective? Are the two sides really disagreeing, or simply engaged in different projects?

FFT is opposed to the idea that there are determinate facts of the matter about content, upstream of facts about use. Facts of the matter about content would imply facts of the matter about statementhood. 'It is probable that P' would be determinately either a regular factual claim or something different, depending on whether there were contents of the right 'shape' (i.e., fit to be the content of a full belief that it is probable that P). The same applies, of course, to determinate facts about truth conditions, or propositions expressed, or any of the other more or less interoperable pieces of the Fregean semantic machinery.<sup>59</sup>

This means that from the standpoint of FFT, the 'semanticists' in the disagreements just mentioned – MacFarlane and the response-dependent realists – must be regarded as mistaken, if they are interpreted as committed to content and truth conditions *upstream of use*. (I don't claim that this is necessarily the correct interpretation; more on this question in MacFarlane's case in §13.8 below, and Price 2022b.) For FFT, the facts of the matter are of two kinds. There will be facts about how the utterances in question are *used* – in particular, about how we structure

---

<sup>59</sup> As I have said, I take this commitment on FFT's part to be entirely of a piece with Gibbard's commitment to global expressivism, mentioned in §12.6 above, and with Brandom's commitment to 'the pragmatist direction of explanation', mentioned above (§0.2, §10.4).

*disagreements* about the claims in question. And there will be various functional and genealogical considerations, to which the expressivist will want to appeal in *explaining* their use.

What does it mean, from FFT's perspective, to say that these are the facts of the matter? Mainly, I think, that these are the considerations that will figure in the kind of theoretical account that FFT envisages, of how we come to talk in the ways that give rise to philosophical puzzlement. They are the *explanandum* and the *explanans* for such an account, respectively.

Where does this leave truth conditions? I want to stress three points. First, to repeat, FFT's expressivists will not be *denying* that the claims in question have truth conditions, any more than he will be *denying* that they are genuine factual claims. FFT wants to *reject* that semantic vocabulary, for theoretical purposes, not to use it in this way to classify utterances one way or the other.<sup>60</sup>

Second, as we've seen, FFT's expressivist will be happy to allow the kind of claims that are simple consequences of everyday, minimalist truth: 'X is  $\Phi$ ' is true just in case X is  $\Phi$ , to borrow Blackburn's example. Such claims are fine, but devoid of any interesting theoretical role.

Third, there's another possibility, that I think the expressivist should acknowledge, indeed welcome. This is that there might be some *other* theoretical purpose for which it is helpful to employ the notion of truth conditions. We should be cautious about assuming that there is only one game in the relevant part of town (that part of town being, roughly, the study of human linguistic behaviour). This point connects to a prominent theme in recent work on expressivism, and deserves a section to itself.

#### 13.4 Pluralism about the goals of linguistic theory

One way to argue for pluralism at this point is to appeal to a distinction drawn by Dummett between two conceptions of the goals of a theory of meaning for a given language – in Dummett's terminology, 'modest' and 'full-blooded' conceptions. A modest theory aims, as Dummett puts it, 'to give the interpretation of the language to someone who already has the concepts required', while a full-blooded theory 'seeks actually to explain the concepts expressed by primitive terms of the language' (Dummett 1975, 102). Davidson's (1967) truth-theoretic approach to a theory of meaning was widely interpreted both as modest, in this sense, and, as I noted in Chapter 2, as requiring only a minimal notion of truth.

In (Price 2004a) I introduce the term 'immodest' for a broader class of theories than Dummett has in mind, when he speaks of *full-blooded* theories of meaning. As I put it, an immodest theory is a theory

with a distinctively linguistic theoretical vocabulary, the need for which is independent of the existence of corresponding linguistic concepts in the object language. A full-blooded theory in Dummett's sense thus counts as immodest. It requires substantial semantic notions—truth, reference, content, and the like—regardless of whether the object language in question is sophisticated enough to contain these notions itself. But ... the converse is not true. Not all immodest linguistic theories count as full-blooded theories of meaning, in Dummett's sense. Why not? Because there are possible immodest theories

---

<sup>60</sup> On the importance of this distinction between denial and silence, see (Price 2004c, §6). It is central to my response to the argument of (Boghossian 1990) that semantic deflation is incoherent, because the deflationist needs to use semantic vocabulary in stating her claim.

which are simply not in the business of ascribing *contents* ... in the sense that Dummett has in mind. (Price 2004a, 214)

What I had in mind here are expressivist theories, of course. I go on to explain the contrast in terms of the difference between specifying *content* and specifying *use*. I conclude:

Expressivism ... thus illustrates the possibility of the following combination of views: (i) An immodest explication of the linguistic role of a particular class of judgements, in non-semantic or non-representational terms. (ii) A modest specification of meaning of a truth-conditional sort ... Stage (i) of such an approach would provide a pragmatic account of how there come to be the kind of judgements whose contents may be specified by the modest theory of stage (ii). (2004a, 217)

A great attraction of the truth-theoretic approach to meaning was held to be that it gave us an account of the compositionality of language, of the way in which the meanings of complex sentences depend on that of their constituents. It is often said to be a failing of expressivist accounts that they don't provide such an account. But this criticism assumes that meaning is a single thing, to be explicated by a single kind of theory, and ignores the possibility that the expressivist and semantic parts of linguistic theory simply belong in different boxes.

Again, Shapiro (2014) makes a point of this kind. After noting that in the work of New Relativists there is 'already agreed to be a loose connection between (a) giving a compositional semantics and (b) explaining how our communicative use of the target discourses is sensitive to perspectives', he continues:

I see no obvious obstacle to undertaking (b) in a way that *doesn't depend on a compositional semantics* ... . An account that takes this route, such as the assertoric force perspectivalism proposed in this paper, may still concede the importance of a compositional semantics ... . Such a semantics would have its proper role in a different explanatory project: the project of explaining or (at least) modeling aspects of our semantic competence with compound sentences, in particular our judgments concerning entailment relations. (2014, 165)

Similar points have been made by several recent expressivist writers. Nate Charlow puts it like this, for example.

Although Expressivism about some kind of language necessarily offers a distinctive *theory of meaning* for that language, there is, I argue, no reason to expect it to offer a distinctive *semantics* for it. (Charlow 2014, 636)

As I noted in §11.3, Michael Ridge (2014) and Matthew Chrisman (2015) draw a useful distinction between semantics and metasemantics, arguing that expressivism deals in the latter. As Chrisman (2015, 12) notes, these points go back a long way. He quotes Davidson (1967), making a similar point about the compatibility between emotivism and a truth-theoretic theory of meaning.

What is special to evaluative words is simply not touched: the mystery is transferred from the word 'good' in the object-language to its translation in the meta-language. (Davidson 1967, 317)

Davidson would allow, presumably, that the task of explaining ‘[w]hat is special to evaluative words’ is a task for *linguistic theory*, and concerns their *meaning*, in some sense of these terms. Thus in distinguishing this task from that of a truth-theoretic account of meaning in his sense, he is already signed up to the kind of pluralism we have been describing.

### 13.5 Yalcin’s bleak narrative

Clarity about the relation between these plural theoretical perspectives would be enormously helpful. A pessimistic assessment of the present situation has been offered by Seth Yalcin (2012), making a case for the kind of distinction we have just been discussing. Yalcin’s central example is the one that got me into this topic in the first place: expressivism about probabilistic operators, or *credal* expressivism, as Yalcin calls it. His recommendation is that ‘credal expressivism ... is best construed as a pragmatic thesis, not as a semantic thesis’ (2012, 125). This is highly congenial, of course, not only from the point of view of the previous section, but also from that of my own early version of the view. As I noted in §0.1, I characterised it in (Price 1983a) as the claim that probabilistic operators are *force* modifiers, not *sense* modifiers.

However, what I want to highlight now is Yalcin’s characterisation of the contemporary expressivist landscape, offered as motivation for his own approach. This ‘bleak narrative’, as Yalcin himself calls it, gives a striking sense of the way in which theoreticians asking different questions can simply talk past one another.

Contemporary metaethical expressivism is widely taken to be, in part, an empirical thesis about some fragment of natural language, a fragment including normative vocabulary. Qua empirical linguistic thesis, the proposal has not gotten very far. Some would say that it has not gotten anywhere. The thesis is not visible in the empirical study of natural language. The debate about expressivism in the literature has overwhelmingly centred, not on whether the thesis is empirically superior to rivals in accommodating various subtle linguistic phenomena, but on elementary questions fantastically prior to this, such as whether expressivism is even in principle compatible with a compositional semantics for a very simple possible language (such as the language of propositional logic). From the perspective of serious, detailed inquiry into natural language, metaethical expressivism is not in the game. (Yalcin 2012, 123)

‘This fact is perhaps not surprising’, Yalcin continues.

For despite being in part a thesis about language, expressivism has not traditionally been motivated in the first place by linguistic considerations. Its canonical motivation stems instead from a package of theses concerning the metaphysics, epistemology and (especially) psychology of the normative. (2012, 124)

Yalcin allows that ‘it is fair enough to theorize about language, or whatever, from the point of view of constraints one sees as having substantial independent motivation.’ ‘What is objectionable’, he says,

is just the lack of results. It remains controversial whether there is a version of metaethical expressivism that delivers anything besides problems for the theorist of language. To many, it has at best seemed to wrinkle linguistic theory, to call for the reinvention of perfectly functional wheels; at worst it has seemed simply incompatible with an explanatory semantics and pragmatics. It has not helped matters that expressivists have not exactly coalesced around anything like a specific programme. But,

being so unmoored from linguistic phenomena in the first place, it is unsurprising that there is little agreement concerning just what the expressivist's linguistic programme actually is, or should be. (2012, 124)

Yalcin is clearly right, in my view, about the need for greater clarity, but wrong if he means that the fault lies solely on the expressivist side. Such a criticism would rest on a lopsided conception of, as Yalcin puts it, 'the empirical study of natural language.' Recall Dreier's characterisation of the key insight of metaethical expressivism – that 'to explain what it is to make a moral judgment, we need not mention any normative properties' (Dreier 2004, 39). Clearly, this is a stance in an empirical study of human linguistic behaviour (amongst other things). It is motivated by questions like these: Why do natural creatures like us speak and think in these terms? What role do these concepts, and this vocabulary, play in our lives? How did they arise, and what difference do they make? The stance that Dreier describes reflects the subject naturalist conception of these questions that motivates so much of expressivism.

Are questions of this kind on the agenda for the kind of 'empirical study of natural language' that Yalcin has in mind – the one that goes on in 'the linguistics room', as he puts it? If the answer were 'Yes', then it would be reasonable for the expressivist to insist that she's not leaving the linguistics room until she's offered some halfway plausible alternatives to her own proposals. (She would be the one to complain about the lack of results.)

A much more plausible answer is 'No', meaning that although expressivists and the linguists Yalcin has in mind are each in their own sense engaged in 'the empirical study of natural language', they are effectively working in different rooms. But this means that the accusation of lack of results on the expressivist side looks a bit unfair – as unfair as it would be to complain that the formal semanticists haven't told us anything about the function and genealogy of any of the vocabularies where expressivism claims to have offered useful insight.

The right move, obviously – and the one actually in play in Yalcin's paper – is to seek greater clarity about the relation between these two programmes. In my view, that requires that we step a long way back, to think about the framing and presuppositions of each. The thesis of FFT was that we need to step far enough back to ask the functional, genealogical and explanatory questions about fact-stating discourse as a whole. My present conclusion is not that this *rules out* semantic theories cast in terms of truth conditions and the like, but simply that we need to be clear what theoretical need they serve. We shouldn't allow the formal semantic tail to wag the subject naturalist's explanatory dog.

### 13.6 Three strategies for explaining NFDs

The issue of explanatory priority is nicely illustrated by recent work on one of the key linguistic phenomena that motivated FFT: in my terminology, No Fault Disagreement (NFD). Under the label 'faultless disagreement', such phenomena have been a major motivation for the New Relativists. Relativism in turn has been criticised by Crispin Wright (2021), who offers his own treatment of such cases. In this section I want to compare my account to those of MacFarlane and Wright.<sup>61</sup>

MacFarlane and Wright are both interested in explaining what I characterised above, quoting from (Price 1991b, 88), as 'the peculiar mix of objectivity and subjectivity we encounter in discourses subject to no-fault disagreements.' Let's begin with MacFarlane, who argues that relativism 'is a good tool for understanding parts of our thought and talk that fall short of being

---

<sup>61</sup> I have explored these comparisons at greater length in (Price 2022b) and (Price 2022d).

fully objective’ (2014, *v*). His primary examples are judgements of taste, such as ‘This apple is tasty.’

MacFarlane notes that such judgements combine two features. On the one hand, we seem to have immediate access to whether things are tasty, in normal circumstances – that’s the *subjective* aspect. On the other hand we often disagree about such matters, as if there were something more than personal preference at stake – that’s the *objective* aspect. MacFarlane contends that relativism makes best sense of this combination.

As MacFarlane notes (2014, 248–249), some of his objections to rival views are similar to mine. This is true particularly of the view he calls *contextualism*, characterised like this.

*Contextualism* I am saying that the apple strikes me in a certain way, or is pleasing to my tastes, or to the tastes of a group with which I identify. The word “tasty” is contextually sensitive, so that my use of it ascribes the property of being pleasing to me or my tastes, while your use of the same word would ascribe a different property: the property of being pleasing to *you* or your tastes. (MacFarlane 2014, 1)

Against contextualism, MacFarlane points out that there is an important issue about the breadth of the reference to tastes. Is it to the speaker’s own tastes, or to some broader standard of the community tastes? Here subjectivity and objectivity pull in opposite directions: subjectivity appears to require ‘that the extra argument place be filled solipsistically, while explaining disagreement requires that it be filled collectively.’ (2014, 13) MacFarlane calls the latter issue the ‘problem of *lost disagreement*’, saying that it is ‘the Achilles’ heel of contextualism’ (2014, 118). He says that a ‘primary selling point of relativist views against contextualist ones is that they purport to capture the subjectivity of claims of taste without losing the disagreement’ (2014, 118).

MacFarlane notes that in (Price 1983a) I raised this kind of objection to what was effectively the probabilistic case of contextualism. If we try to build a reference to evidence into probability claims we get the same subjective–objective dilemma, and no specification gets everything right. I made similar points again a few years later, now as an objection to RDR. This passage gives a flavour of my argument in the latter case, and shows how similar it is to MacFarlane’s objections to contextualism.

[Proponents of RDR] thus face the following dilemma. If they leave out the notion of normality [i.e., in MacFarlane’s terms, choose the *solipsistic* alternative], the resulting account is insufficiently objective to make sense of ordinary prima facie disagreements about the matters in question. If they put in normality [i.e., in MacFarlane’s terms, choose the *collective* alternative], on the other hand, the account cannot make sense of the very real possibility of no-fault disagreements, arising in cases in which there is statistically significant divergence in patterns of response. (1991b, 105–106)

MacFarlane and I differ in what we take the preferable alternative to be, of course. For me, it is what I would now call expressivism; for MacFarlane, it is relativism. By my lights, relativism joins RDR in taking the wrong path at the fork – in trying to incorporate subjective elements into (relativised) *truth conditions*, rather than into an account of how the claims in question are *used*.

Another point of similarity, this time with Wright as well as with MacFarlane, is that both raise the question why judgements such as those of taste get expressed at all in what Wright calls ‘the objectifying idiom’ (2021, 441). Why do we say ‘Sushi is delicious’ rather than ‘I find sushi delicious’, to use Wright’s example? Both Wright and MacFarlane offer answers which are in one

sense like mine. As Wright puts it, ‘we have an interest in having an idiom that enables us, more than merely reporting a response we personally have, to *project* it as a possible point of coordination’ (2021, 443).

Similarly, MacFarlane asks:

[W]hat is the point of fostering controversy in “subjective” domains, if there is no (nonrelative) truth on which both parties can converge? Why shouldn't we just talk about our own tastes, rather than ascribing subjective properties to the objects? (2007, 30)

MacFarlane’s suggestion in response is very much like Wright’s.

Perhaps the point is to bring about agreement by leading our interlocutors into relevantly different contexts of assessment. ... Perhaps ... the point of using controversy inducing assessment-sensitive vocabulary is to foster *coordination* of contexts. We have an interest in sharing standards of taste, senses of humor, and epistemic states with those around us. The reasons are different in each case. In the case of humor, we want people to appreciate our jokes, and we want them to tell jokes we appreciate. In the case of epistemic states, it is manifestly in our interest to share a picture of the world, and to learn from others when they know things that we do not. (2007, 30)

Notice that like Wright, MacFarlane is proposing this for the subjective cases, taking for granted, apparently, that some other account will work elsewhere.<sup>62</sup> My view, in contrast, has been that we should be more ambitious. We should ask this question about the point of ‘fostering controversy’ as a question about assertoric language *in general*, rather than about these special cases. In case it is unclear how the answer can be anything but obvious in the general case, note two things.

First, it is easy to find ways of *cancelling* controversy, for any sort of assertion whatsoever. It isn’t just in these evaluative cases that we can fall back, as Wright notes, to a subjective idiom. On the contrary, we do this kind of thing all the time, when we want to reduce the temperature of a disagreement. Instead of simply saying ‘P’, I say ‘Well, my own view on this is that P.’ It is worth asking why this isn’t the default. In other words, why doesn’t saying ‘P’ have this more gentle, controversy-reducing feel from the start?<sup>63</sup>

Second, any seemingly obvious answer in the general case is likely to appeal to our intuitions about truth and falsity, say by pointing out that where two speakers disagree about a non-subjective case, one of them believes something *false*. But that won’t do if our interest is in getting at the ‘point’ of truth and falsity themselves, in Dummett’s (1959) sense; as I said in §0.1, that was how I saw my project. Hence, while my own proposal is in one sense very much like these suggestions from MacFarlane and Wright, there is one very big difference. I’m applying the idea quite generally, to all ‘states of mind ... with respect to which there is reason to seek agreement between speakers’ (Price 1983c, 356).

In Wright’s case, there is a second important difference. Wright’s proposal for explaining faultless disagreement is that in the case of discourses subject merely to a minimal notion of truth, truth lacks its usual normative dimension. When speakers make conflicting judgements of

<sup>62</sup> I discuss MacFarlane’s proposal for the non-subjective cases in (Price 2022b).

<sup>63</sup> This was a central question in (Price 1998a) and (Price 2003), two papers that build on the proposals of FFT – more on those pieces in §14.1 below.

taste, for example, neither thinks that the other is in violation of some norm; hence, according to Wright, the absence of fault.

According to FFT, however, the normative dimension of truth and falsity – the very thing Wright is proposing to abandon in the case of the minimal truth – is a crucial element. It is what provides friction, or *pressure* for coordination, and hence what distinguishes truth as we have it from a mere device for saying ‘Me too’, or ‘Ditto’.<sup>64</sup>

By way of comparison, think of a group of diners, asked by a waiter for their drink orders. Their responses are more than mere reports of preference, of course. (Imagine the reaction of the waiter, if one of them says, ‘I didn’t order this coffee – when I said that I would like a coffee I was simply describing my state of mind, nothing more.’) So the diners are doing more than ‘merely reporting a response [they] personally have’, as Wright puts it (2021, 443). Moreover, their choices certainly provide ‘a possible point of coordination’, as they may find out if the waiter restricts their options – ‘Sorry, madam, the kitchen can’t do a pot of tea just for one.’ And simple verbal devices such as ‘Ditto’, ‘Same again’ or ‘Not me’, for piggybacking on utterances by other diners, can be useful in these contexts, too. But these devices provide no *pressure* for coordination, in normal circumstances, because they are not normative.<sup>65</sup>

Thus, while Wright is right, in my view, that ‘the objectifying idiom’ (Wright 2021, 441) is a device to facilitate coordination, I think his proposal that the subjective cases are non-normative would discard the very ingredient that makes the idiom work in this way: again, no norm, no pressure.

At one point Wright offers a succinct characterisation of the options, as he sees them.

[Faultless disagreement is] a hopeless idea if the discourse is thought of as answerable to a single norm of truth with which no statement and its negation can simultaneously comply. So if faultless disagreement is to be a possibility, there must be no such single alethic norm. That leaves two options. One is, in one way or another, to—as it were—fracture the norm, multiply the ways of being true, and spread the pieces around, so that

---

<sup>64</sup> Strawson’s early ‘performative’ account of truth is sometimes called the *amen* theory. Here he compares ‘That’s true’ to a possible use of ‘Ditto’:

[I]magine a possible, and perhaps vulgarly current, use of the expression ‘Ditto’. You make an assertion, and I say ‘Ditto’. ... I am agreeing with, endorsing, underwriting what you said; and, unless you had said something, I couldn’t perform *these* activities, though I could make the assertion you made. Now the expression ‘That’s true’ sometimes functions in just the way in which I have suggested that ‘Ditto’ might function. (Strawson 1949, 89–90)

As I noted in Chapter 2, Bernard Williams once said that he didn’t see ‘how on such a theory it could be more than an accident of language that “is true” signified agreement with *assertions* rather than agreement with anything else’ (Williams 1966/1973, 203). On FFT’s view, the fact that it is not used this way is linked to its normative force, which would be inappropriate in other cases. Cf. fn. 11 below.

<sup>65</sup> Both Wright and some of his deflationist opponents seem to miss this important distinction between a mere device for saying ‘me too’ and a device for *normative* endorsement. In the following passage, for example, Wright is setting up the ground for arguing that even a deflationist must recognise the normative notion of endorsement. But the characterisation he gives here of the ‘positive deflationist contention’ (i.e., that in ordinary cases ‘That’s true’ could be replaced by re-assertion of the original claim) picks out merely the ‘me too’ aspect, not the normative aspect, as the applicability of a direct analogue in the restaurant ordering case demonstrates:

Let us focus, for ease of exposition, on “true” as predicable of propositions, and on the positive deflationist contention that, in its most basic use, the word is essentially a device of endorsement which, except in cases where the content of the proposition endorsed is not explicitly given, or where quantification over propositions is involved, may be dispensed with altogether in favour of a simple assertion of the proposition characterized as “true.” (Wright 1998, 40)

conflicting opinions can each alight on a shard. [Relativism] attempts a particular implementation of that option. The other option is to suction out the substance of the alethic norm, leaving only the formal shell ... I have argued that the first option will not deliver what is wanted, and that the second is the way to go. I regard it as a strength of the combination of minimalism about truth aptitude and alethic pluralism defended in *Truth and Objectivity* that it provides a natural setting for the elaboration of the second direction. (2021, 445)

FFT offers a third option, that truth provides a flexible norm, positive-presumptive<sup>66</sup> but cancellable. More on this below, but first to the difficulties facing MacFarlane's and Wright's proposals.

In MacFarlane's case, I think he fails successfully to address an objection he himself describes in the following passage. He is referring here to the challenge he has raised for contextualists, of combining subjectivity and disagreement for judgements of taste, and similar cases.

One might ... despair of ever getting subjectivity and disagreement into the same picture. Perhaps we just have to choose. This is where the relativist comes in with her seductive song. "You can have it both ways," she says, "if you just accept that propositions about the funny, the delicious, and the likely have truth values only relative to a person or perspective. When I say that apples are delicious and you deny this, you are denying the very same proposition that I am asserting. We genuinely disagree. Yet this proposition may be true for you and false for me. That is what the 'subjectivity' of these claims comes to: perspectival dependence of their truth on features of the subjects who assess them.  
...

The question I want to address is whether this helps with the problem of lost disagreement. How can there be disagreement between Abe and Ben, on the relativist's view, if the proposition Abe asserts and Ben denies is true relative to Abe's standard of taste and false relative to Ben's? Aren't they just talking past each other, in some sense? (2007, 21–22)

Later in the same paper, MacFarlane claims that his relativist account does 'capture ... the distinctive phenomenology of disagreement about matters whose truth is relative' (2007, 29). But the example he gives is revealing.

The challenger thinks (rightly) that he has absolutely compelling grounds for thinking that the assertion [e.g., 'Liquorice is tasty'] was not accurate. But the original asserter thinks (also rightly, from her point of view) that the challenger's grounds do nothing to call in question the accuracy of the assertion. The asserter's vindication will seem to the challenger not to show that the assertion was accurate, and the challenger will continue to press his claim. (Until the game gets boring.) Thus we have all the normative trappings of real disagreement, but without the possibility of resolution except by a relevant change in one or both parties' contexts of assessment. (2007, 29)

Is MacFarlane right that his account 'captures the distinctive phenomenology of disagreement' about matters of taste, and that it offers us 'all the normative trappings of real

---

<sup>66</sup> To borrow a Wrightean phrase (Wright 1989, 251).

disagreement'? In my view no, on both points.<sup>67</sup> I think MacFarlane's account makes a mystery of the phenomenology of disagreement, *because* it fails to make sense of the normative trappings. Why, given relativism, should I, or a fellow speaker, *care* that I think I have compelling grounds for thinking that their assertion is inaccurate? As MacFarlane himself puts it, '[a]ren't [we] just talking past each other, in some sense?' (2007, 21)

It is true, as MacFarlane notes, that the need or desire for coordination may provide a reason to care about the disagreement, *but such a reason is not there by default*. We see this easily in the expression of intention cases, like my example above. To vary that case slightly, imagine that the waiter approaches the table, and says, 'Would anyone care for coffee?' We can imagine an alternating sequence of responses from two diners: 'Yes, please'; 'No, thank you'; 'As I said, *yes!*', 'And as *I* said, *no!*'; ... This becomes boring, too, and for the same reason. There is nothing more than a trick of the language to make the two diners think there is any issue between them<sup>68</sup> – even though here, too, as I noted, there may sometimes be reasons for coordination.

The trick of the language may be more convincing in the tasty case, because there the labels 'true' and 'false' are permitted, and the two parties apply them differently. This would explain the phenomenology of disagreement if the speakers don't realise that the case is subject to context-relative standards. But this is no help to MacFarlane, who wants to get the phenomenology right for speakers who do know the story. (He wants the relativist's account of truth and accuracy to *explain* the phenomenology, presumably.) Whereas it seems to me we only 'get the normative trappings' if we *ignore* what we know about the perspectival character of the truth conditions – if we pretend that there is a single standard that we are both trying to meet. By 'fracturing the norm', as Wright puts it (2021, 445), relativism throws away the friction-generating character of truth. As a result, relativism, too, falls victim to the problem of lost disagreement.

What about Wright's alternative? In my view, as I said above, these objections apply equally to Wright's account. No fault means no friction, and hence no pressure for coordination. Again, ill-informed speakers might be confused about this, but it seems fair to apply the same test to Wright that we applied to MacFarlane. His account should work for folk who are assumed to know the full story.

This challenge becomes particularly pressing once we realise that no fault disputes can be *dynamic*. By this I mean our sense that we do have a substantial disagreement *at the beginning of the exchange* – that is, when (i)–(iii) are satisfied, in Shapiro's characterisation in §13.1. This must be so if we want to allow, as seems obviously true in many such cases, that evaporation *may never occur*. Even about matters of taste, many disputes take a normal course, with no relaxation of fault. We need to explain why properly informed speakers do take such disputes to involve fault – Shapiro's condition (iii) – before the point, never reached, in many cases, at which the disputes evaporate. MacFarlane perhaps has some resources to account for this, in terms of shifting contexts of assessment. But for Wright, without these resources, it is hard to see how the dynamic character of NFDs could be anything other than a mystery.

---

<sup>67</sup> To be clear, I am happy to allow that what MacFarlane describes here is, indeed, 'the distinctive phenomenology of disagreement *about matters whose truth is relative*', in his sense. What I deny is that such cases 'have all the normative trappings of real disagreement'.

<sup>68</sup> Here is a real example, illustrating that this sort of linguistic trick can briefly confuse an intelligent four year old: "I love you, Mummy"; "No, I love YOU, Ruben"; "No, I love YOU!"; "No, I LOVE YOU!"; "Mummy, how about we love each other?"; "Perfect!" (Rozea, Marnie and Rozea, Ruben, 2021).

In different ways, Wright and MacFarlane both think of the subjective cases as fundamentally distinct from the objective cases – the cases in which they take there to be no option for no fault disagreement. My view, in contrast, maintains that the subjective cases involve exactly the same normatively-loaded game of challenge as disputes about, say, the age of the Earth (one of MacFarlane’s examples). I take ordinary assertoric conversation *in general* to be the kind of coordination device that Wright and MacFarlane both propose for the subjective cases.

Crucially for our current concerns, this game of challenge is well adapted to variability with respect to the ‘need’ for and value of coordination. It is part of the normal functioning of the game that there are exit moves – ways of cancelling the norms, in particular disagreements.<sup>69</sup> The so-called subjective cases are those in which there is often and obviously a rationale for using these in-built escape hatches. The phenomenology of subjectivity arises from the fact that use of the escape hatches cancels the normal presumption that there is a common goal, though we only see this for what it is when we understand the source of the latter. Proper play of the game, required for it to serve its coordinative function, requires that the escape hatches initially be ignored. Fault is positive-presumptive, as I put it earlier.

By analogy, imagine a play-fighting game, say paintball, in which players compete to shoot each other with small paint-filled capsules. Imagine there is a designated escape phrase, say ‘Time out!’, that immediately pauses the game when uttered or displayed. For better or worse, participants who wear ‘Time out!’ on their hats never get to experience the competitive phenomenology of the game. In different ways, MacFarlane and Wright accomplish something similar. MacFarlane puts the players in different rooms, and Wright takes away their weapons. My view puts them in the same room, with their weapons, and relies on their knowledge of the escape options provided in normal play of the game to cancel the conflict, when that seems appropriate.

In my view, as I have said, the escape options are never absent altogether, but I don’t need to insist on that point here. Even if I allowed cases in which there was no possibility of evaporative disputes – no possible role for the escape hatches – the story I would tell about the source of the phenomenology of disagreement would be no different from the one I tell about the so-called subjective cases. In all cases, it takes the source of the phenomenology to be a distinctive norm, at the heart of the assertoric coordination game.

Certainly, there is a good question why expressions of taste, and other obviously subjective cases, should ever invoke the machinery of such a game in the first place. But as we saw, MacFarlane and Wright propose similar answers. It has ‘engineering’ advantages, as MacFarlane puts it (2014, 310), in communities of social creatures like us. In general, it is an advantage of an expressivist conception of the functional role of commitments of various kinds that it stands back far enough back to consider the question of the merits of ‘making those commitments explicit’ in the framework of the game of giving and asking for reasons. As I argued in FFT (§8.11), we should expect that different communities will draw the line in different places, and should not be surprised to find both practices in use in a single community – as we do, of course, with pairs such as ‘Yum!’ and ‘That’s tasty!’

---

<sup>69</sup> There are actually many of these, just some of which are relevant here. The key thing is cancelling the presumption of fault, or genuine disagreement, though MacFarlane and Shapiro have convinced me that this isn’t such a straightforward thing to characterise accurately as FFT assumed. MacFarlane (2014, 133–136) makes some criticisms of Kölbel’s (2004) notion of ‘faultless disagreement’, and Shapiro (2014) notes some infelicities in my use of ‘no fault disagreement’. I won’t try to sort out those issues here.

Thus, to sum up, I think that FFT's approach makes much better sense than MacFarlane's or Wright's proposals of the phenomenology of faultless disagreement. It also offers an approach that extends, as theirs do not, to what they regard as non-subjective cases. In Wright's case, we'll return to the latter point below.<sup>70</sup> I want to argue that FFT's expressivism has a distinctive advantage in explaining the normativity of truth in general, something on which Wright himself places a great deal of weight.

For the moment, I offer this comparison between FFT, MacFarlane, and Wright as an example relevant to the 'What is at stake?' question. I've argued that the thoroughly pragmatic, use-based perspective of FFT offers the best explanation of the phenomenon of faultless disagreement. (This is at least one reason for thinking that expressivism is relativism done right, rather than the other way around.)

### 13.7 Formal options

Still, explanation is not the only game in town. In other cases (probability and causation, for example), a recognition of the practical origins of a concept is clearly compatible with a great deal of objectification of the subject matter concerned.<sup>71</sup> (In a sense, this is the fundamental point of quasi-realism.) In the present case, we might well be interested in modelling the formal properties of this exit-equipped version of the game of giving and asking for reasons. What are the options, at this point?

We saw that putting a fixed context-relative semantic value in the foreground leads us astray. It loses disagreement, even in MacFarlane's form of the proposal. Provided we were clear about the intended direction of explanation, however, I think we might model things in terms of *changeable* semantic values. The dispute about tasty would begin in ordinary objective terms, and switch to a contextualist or relativist model when the relevant escape hatch is operated. On this view there would in a sense be a change of subject matter, or truth conditions, as the dispute develops. But this need not be objectionable in the way that it would be if the subject matter were being used to *explain* the dispute behaviour. It's now the other way round. What the disputants are said to be *saying* is being determined in the light of their (changeable) dispositions to take themselves to be 'really' disagreeing.<sup>72</sup> This approach might find allies in other dynamic and highly contextual approaches to meaning, such as Charles Travis's radical 'occasion-sensitivity' (see, e.g., Travis 2000).

However, if we want a formalisation that makes better contact with the explanatory basis of the relevant phenomena, we need to look somewhere else. Here Shapiro's (2014) suggestion is attractive. As we saw, Shapiro points out that the linguistic phenomena on which MacFarlane and other relativists have focussed have a substantial overlap with the NFDs of FFT. However, Shapiro feels that relativists have been addressing the issues posed by such cases in the wrong key.

The central thrust of relativism, I argue, can and should be understood independently of any semantic framework of relativized truth. Instead, relativism should be understood in *pragmatic* terms, as corresponding to a particular understanding of assertoric force. The

---

<sup>70</sup> I discuss MacFarlane's treatment of the objective cases in (Price 2022b).

<sup>71</sup> Thus subjectivists about probability, including those who understand subjectivism in the expressivist sense, are fully entitled to build formal models of probability, and to use probability elsewhere, say in models in physics. Their subjectivism, or expressivism, rests on an understanding in the background about how such talk gets its meaning, via its connection to decision under uncertainty.

<sup>72</sup> Thus *saying* downstream of *doing*, in Brandom's terms.

idea starts with Robert Brandom’s analysis of “fact-stating discourse” (1994: 607) as a “game of giving and asking for reasons” whose basic move is *asserting*. Brandom’s account of assertion makes no appeal to truth. My proposed revision of his analysis makes room for a broader conception of fact-stating discourse, by allowing *assertoric force* to depend on speakers’ perspectives. What is distinctive and plausible about relativism, I will argue, is best captured by the resulting liberalized version of Brandom’s game of giving and asking for reasons. (Shapiro 2014, 139–140)

Specifically, Shapiro proposes an amendment to Brandom’s account of the ‘authority’ claimed by someone who makes an assertion. According to Brandom, as Shapiro puts it: ‘the *authority* claimed in asserting a proposition is to license others to assert the same proposition and use it as a premise in their theoretical and practical reasoning.’ (2014, 153) Shapiro suggests that to capture ‘what is distinctive and plausible about relativism’, we should restrict this claimed authority.

Stated generally, the idea is that *assertoric force* has a perspectival aspect. On this view, an assertion may carry its licensing potential not vis-à-vis its entire potential audience, but rather vis-à-vis a target audience restricted to those who share a particular perspective. (2014, 154)

Shapiro himself applies the proposal to two kinds of case: epistemic modals, such as ‘it’s possible that ...’, and predicates of personal taste, such as ‘tasty’ and ‘fun’. But he notes that it is likely to have broader application. He points out that MacFarlane’s framework, to which he is proposing an alternative, has been applied ‘to explain discourse about knowledge, epistemic possibility, matters of taste, and contingent future events’ (2014, 139); and that in FFT I argue ‘that there may be room for evaporative disputes concerning indicative and subjunctive conditionals, aesthetic and moral matters, modality, causation and secondary qualities’ (2014, 142).

This is a very promising framework for exploring these issues, in my view. As we have seen, the language of perspective and subjectivity is at the core of my own version of expressivism. True, it isn’t immediately clear that my use of these notions is compatible with Shapiro’s use of them – see (Shapiro 2021b) for some dissent from Shapiro’s side – but Shapiro’s insistence on fitting them into a *pragmatic* rather than a *semantic* framework is clearly congenial, from my point of view, for the reasons we have explored above.<sup>73</sup>

### 13.8 MacFarlane and Wright as quasi-expressivists

Despite our disagreements about how to handle the subjective cases, both MacFarlane and Wright look like near-neighbours, from my point of view. This may not seem surprising in MacFarlane’s case. I have noted his remark that relativism might be expressivism done right. But it may seem more implausible for Wright, who at the time of *Truth and Objectivity* was a leading critic of expressivism.

Let’s start with MacFarlane. We’ve already seen that MacFarlane and I are on much the same page in asking why creatures like us come to speak ‘in objective mode’ about what are, at base, subjective matters. The answers we offer are different in the ways already described, but the explanatory orientation is the same. This is particularly clear in passages in which MacFarlane

---

<sup>73</sup> Shapiro’s framework also handles very easily the distinction we drew in Chapter 11 between *divergent* and *convergent* cases of NFD. In Shapiro’s terms, this amounts to the distinction between cases in which speakers *split* or *amalgamate* their target audiences, when a dispute evaporates.

adopts what he calls the ‘engineering approach’.

Here it is useful to take an engineering approach to the bits of language we have [been considering]. An engineer building a device needs to start from a description of what the device is supposed to do. So, we will start by asking what these putatively assessment-sensitive bits of language are *for*. What role do they play in our lives? What purposes do they serve? ... For concreteness, we will focus on one case—that of knowledge attributions—but the considerations are easily generalized. (2014, 310)

Importantly for the comparison, MacFarlane isn’t just asking this question about what we might call the generic features of the objective mode – the use of the labels ‘true’ and ‘false’, for example. He’s asking it, as my expressivist does, about the discourse-specific concepts themselves. Thus, turning to the knowledge case, he begins like this.

Why do we talk about what people know, and not just what is true and what people believe? Plausibly, our knowledge talk serves our need to keep track of who is authoritative about what: who can be properly relied on as an informant on a given topic (Craig 1990; Chrisman 2007). If we are considering buying a motorcycle and want advice, we will ask for someone who knows a lot about motorcycles. When we say that someone’s beliefs do not amount to knowledge, we are saying not to trust or rely on them in this matter. When we criticize people who act on mere belief (not knowledge), it is for acting rashly, relying on sources that are not authoritative. (2014, 310)

The questions MacFarlane asks here are in what in §11.4 I called *linguistic* rather than *material* mode. In other words, the issue is ‘Why do we talk this way?’, not ‘What is this thing we call knowledge?’ And there is no hint in MacFarlane’s answer of the sort of representational perspective that takes us back to the material level, as there would be if the answer were something like ‘We’re keeping track of a distinctive kind of mental state, and that’s important for these reasons ... .’ We’re keeping track of *something*, in MacFarlane’s view – i.e., of ‘who is authoritative about what’ – but, in the standard expressivist manner, that’s not in any sense the content of what we say.

Of the authors MacFarlane cites here, Chrisman (2007) is avowedly an expressivist, and Craig is widely taken that way. A notable recent example of this reading of Craig is that of Hannon (2019), who calls his own Craig-inspired view ‘epistemic pragmatism’, citing Chrisman (2007) and adding this explanation of his terminology.

In other philosophical contexts it might be important to emphasize the differences between myself and other so-called pragmatists, but for the purposes of this chapter these differences may be ignored. I use this word because it loosely reflects an attitude or approach that I share with pragmatists such as Wittgenstein, Rorty, and Price. These authors seek to dismiss or demote philosophical puzzles in favor of more practical questions about the roles and functions of the words and concepts in terms of which we talk and think about the things that lead to philosophical puzzles. This type of pragmatism has a second-order, or linguistic, focus. An epistemic pragmatist will investigate the role of the term or concept “knowledge.” (Hannon 2019, 180)

In other words, it seems uncontroversial that MacFarlane and I are on the same page with respect to the kind of questions we think worth asking, in such cases. This was the crucial third

ingredient in my expressivist recipe in Chapter 11.<sup>74</sup> If we differ, it is on the question whether this engineering stance should be a global one.

Now to Wright. As I noted in Chapter 10, Wright was one of those who argued that semantic minimalism defeats expressivism. The argument occurs in the last section of Chapter 1 of *Truth and Objectivity*. As Wright says, his object there is

to commend ... a kind of minimalism about truth—a species of deflationism, if you will, but unencumbered by the classical deflationist’s claim that truth is not a substantial property. The minimalist view is that when a predicate has been shown to have the relevant features, and to have them for the right reasons, there is no further question about the propriety of regarding it as a truth predicate. Minimalism is thus at least in principle open to the possibility of a pluralist view of truth: there may be a variety of notions, operative within distinct discourses, which pass this test. (Wright 1992, 24–25)

A couple of pages later, Wright asks the question that interests us here.

[N]eed any of this provoke a proponent ... of expressivism, or “quasi-realism”, . . . to disagree? In effect, I have argued that there is a notion of truth aptitude which is carried in train by possession of assertoric content ... But I have so far ventured nothing about possession of assertoric content as such; in particular, nothing at variance with the suggestion, integral to the expressivist tradition, that possession of genuine assertoric content is a relatively *deep* feature of the sentences of a discourse, which its overt syntax can serve to mask, or merely to simulate. The final ingredient in any conception of truth and truth aptitude which can be common ground between realist and anti-realist is that this is not so.

An analogy may help. Elsewhere I have argued that Frege’s platonism about number is best interpreted as based on the view that an expression’s candidacy to refer to an object is a matter of its syntax: that once it has been settled that a class of expressions function as singular terms by syntactic criteria, there can be no further question about whether they succeed in objectual reference which can be raised by someone who is prepared to allow that appropriate contexts in which they do so feature are true. There is, that is to say, no *deep* notion of singular reference such that an expression which has all the surface syntactic features of a Fregean proper name, and features in, say, true contexts of (by surface syntactic criteria) predication and identity, may nevertheless fail to be in the market for genuine—“deep”—reference. So too, in the present context, the claim must be there is no notion of genuine—deep—assertoric content, such that a discourse which exhibits whatever degree of discipline (there are firmly acknowledged standards of proper and improper use of its ingredient sentences) and which has all the overt syntactic trappings of assertoric content (resources for—apparent—conditionalisation, negation, embedding within propositional attitudes, and so on)—no notion of genuine assertion such that a discourse with all this may nevertheless fail to be in the business of expressing genuine assertions. Rather, if things are in all these surface respects as if assertions are being made, then so they are. (1992, 28–29)

In my copy of *Truth and Objectivity*, as I (re)discovered recently, this last paragraph has a marginal note, left there by me on a forgotten previous reading: ‘CW as my kind of global

---

<sup>74</sup> This means that MacFarlane’s affirmation, as Shapiro puts it, ‘that there is such a property as being tasty, *simpliciter*’ (Shapiro 2014, 145), is not the kind of heavy-duty metaphysical claim that troubles contemporary naturalists. MacFarlane avoids heavy-duty metaphysics in exactly the way that the expressivist does. Both bring the case within the scope of subject naturalism rather than object naturalism.

expressivist! The comment is mistaken in one respect, because Wright does not commit himself to *global* minimalism about truth. But it is entirely accurate, in my view, with respect to all those domains that are by Wright's lights subject only to his minimal notion.

The point is easy to see if we keep in mind the discussion of Chapter 10. As I noted there, traditional expressivism tended to combine two claims: a *negative* claim, expressed in semantic vocabulary, and a *positive* claim, expressed in non-semantic vocabulary. Minimalism about the semantic notions deflates the negative claim, but doesn't touch the positive claim, which is where expressivists make their distinctive moves. Indeed, not only does minimalism not challenge the positive claim, it actually mandates it, in the sense that it implies that interesting theoretical work has to be done in non-semantic terms (the semantic terms having turned out to be too thin to bear any weight).

Once these issues are sorted out, it is clear that expressivists should never have been in the business of denying that moral claims (say) lack some deep sort of assertoric content – this should never have been 'integral to the expressivist tradition', as Wright puts it. Instead they should have said, as leading expressivists such as Blackburn and Gibbard were indeed soon saying, that while of course such claims have assertoric contents in the thin everyday sense, they originate as expressions of affective attitudes of some kind, just as non-cognitivists had always maintained.<sup>75</sup>

From this point, Wright's neo-Fregean platonism about number should have looked like friendly ground for expressivists, in my view. If it did not, the explanation may be that many expressivists were still in the grip of the dogma that expressivism is necessarily a view about normative matters. But the moves Wright describes – waving aside concerns about 'deep' reference, for example – are precisely the moves that expressivism needs, to wave aside metaphysics and focus instead on its explanatory project. As MacFarlane expresses the project in the passage we quoted above, 'we ... start by asking what these bits of language are *for*. What role do they play in our lives? What purposes do they serve?' (2014, 310)

To reinforce the sense that Wright and I are on the same page on this point, let me note a short remark from his recent paper discussed above.

[L]et us be mindful that, where truth is deflated and so registers no independent norm operative over the acceptance and assertion of statements, an account of meaning—of correct linguistic practice—has to proceed in other, non-truth-conditional, broadly *use-theoretic* terms. So we need to focus directly on the use of the signature statements in [the] discourses [in question]. (Wright 2021, 441)

It is common ground among expressivists that one of the central distinguishing marks of the view is that it explains the meaning of a target vocabulary in terms of use conditions, rather than truth conditions. While this may not be all it takes to be an expressivist, it is an important and

---

<sup>75</sup> One of the first writers to use the term 'non-cognitivism' was Carnap, who introduces it in his reply to Abraham Kaplan in the Schilpp volume (Kaplan, 1963; Carnap 1963). Carnap wrote this reply in the mid-1950s, though he had endorsed the view in question decades earlier. Responding to Kaplan's characterisation of Carnap's view as the claim that '[w]hat is essential is that "true" and "false" are held to be inapplicable to value judgments, a point usually expressed by denying them "cognitive meaning"' (Kaplan 1963, 827), Carnap seeks a more general term than 'emotivism'. He suggests 'non-cognitivism (with respect to value statements)', and goes on to say that this view 'is simply a special case of the general thesis of logical empiricism that there is no third kind of knowledge besides empirical and logical knowledge' (Carnap 1963, 999–1000). Carnap appears to have failed to notice that there exists a tension between non-cognitivism and this 'general thesis of logical empiricism', in the light of his own voluntarism and semantic minimalism. I discuss these points in (Price 2018, forthcoming).

major step in that direction – one that Wright here takes to be a direct consequence of a discourse’s being merely ‘minimally truth-apt’, as he puts it (2021, 439).

Of course, as in the case of MacFarlane, this conclusion comes with a qualification. Wright rarely, if ever, comes close to global expressivism, because his default view is that truth is not everywhere the minimal notion. In particular, Wright rarely seems to countenance the possibility of expressivism *about truth itself*. If he does so, by my lights, it is where he has on the radar the possibility that truth might be *everywhere* minimal, in his sense. If that were so, then global expressivism would follow from the fact that minimalism implies expressivism, in my sense, in each particular case (including, obviously, the case of truth itself). In general, however, truth itself remains for Wright a staunch hold-out against global expressivism. It seems likely to bring others in its train, in the sense that a realist, non-minimal notion of truth would ground a bifurcation between expressive and non-expressive discourses, the latter employing the non-minimal notion.

In the final chapter I want to explore this difference between my views and Wright’s about truth itself. As I said at the beginning of the chapter, the interest of the comparison lies especially in the fact that Wright’s views and mine are in other ways rather similar. I’ll begin by describing two papers in which I developed FFT’s proposed account of truth, both of which take Wright’s side against other critics.

## Chapter 14

### Neighbours and Rivals II: Explaining Truth

Now, lastly, to some comparisons concerning truth. I'll begin here by explaining how, when I developed FFT's account of truth in two later papers, both had connections to Crispin Wright. Both defended Wright's stance on the normativity of truth against criticisms by others (Horwich and Rorty, respectively). But I'll argue that what this stance on normativity recommends is FFT's expressivist *explanation* of truth, not Wright's quest for a *metaphysics* of truth (albeit a qualified, pluralistic, and in some cases minimalistic one). I'll thus be defending FFT's preference for *explanation* rather than *analysis* of truth, by comparison with Wright.

One element in Wright's interest in a metaphysics of truth is the attraction he feels to a programme advocated in the 1990s by writers such as Michael Smith (1994) and Frank Jackson (1998) – what Hawthorne and I (1996) dubbed the Canberra Plan. As I noted in §11.4, the popularity of Canberra Plan among my geographical neighbours was an important influence on me, in the years after FFT, as I tried to show that expressivism offered a better solution to the problems the Canberra Plan claimed to address. Here I'll explain what I take its limitations to be, both in general and as a route to illumination about truth.

Turning from metaphysics to explanation, the second near-neighbour I'll discuss in this chapter is Bernard Williams. Judging the book by its title, as it were, Williams's *Truth and Truthfulness: An Essay in Genealogy* (Williams 2002) certainly looks congenial, by FFT's lights. Elsewhere, on the strength of this work but without discussion, I've claimed Williams as a Cambridge pragmatist (Price 2017a, 149). Matthieu Queloz makes a well-argued version of the same claim in his helpful recent account of Williams's project (Queloz 2021b). Here, drawing on Queloz's discussion (and adopting his useful term 'reverse-engineering' for the explanatory, genealogical stance) I'll qualify this assessment of Williams in one respect. I'll argue that FFT's pragmatism is more thoroughgoing than Williams's, in the following sense. Concerning truth itself, FFT proposes to reverse-engineer what Williams, apparently following Davidson, takes finished off the shelf.

#### 14.1 Truth as convenient friction

In the decade after FFT, I defended its approach to truth in two main pieces. The first, 'Three norms of assertibility, or how the MOA became extinct' (Price 1998a), stemmed from my interest in a disagreement between Wright and Paul Horwich. In *Truth and Objectivity* (Wright 1992), Wright had argued that Horwich's view cannot account for the normativity of truth – normativity, Wright claims, that can be shown to be a consequence of Horwich's disquotational schema, via what Wright calls the *Inflationary Argument*. In (Price 1998a) I argue that the Inflationary Argument is unsuccessful, so that Horwich has the better of the immediate battle. Nevertheless, Wright wins the war. Truth is normative, in a way that Horwich's particular version of deflationism does not explain, albeit for reasons that Wright does not correctly identify.

The paper distinguishes the conversational norm I take to be associated with truth from two other norms, those of *sincerity* and *warranted assertibility*. Truth is thus my 'third norm', and the focus of the paper is on what this third norm adds to our conversational practice – on the question *how things would be different if we didn't have it*. As an aid to answering this question, I imagine a community who don't have the third norm, but who nevertheless use speech acts to give voice to their opinions. I call such speech acts 'merely opinionated assertions', or *MOAs*, and the speakers in question *Mo'ans*.

As I noted in §13.6, it is easy to find something like MOAs in our own linguistic practices, using devices to cancel the third norm: ‘My own opinion is that P’, ‘Mine is that not-P’, and the like. In Wright’s (2021) terms, then, the Mo’ans are speakers who don’t, or no longer, have the device of putting their opinions into a public realm. Their assertions are not what Wright calls ‘objectifying statement[s]’, but merely ‘subjective-relational (explicitly autocentric) report[s]’ (2021, 441).

I cover similar ground in ‘Truth as convenient friction’ (Price 2003), a piece first written for a conference with Richard Rorty at ANU in 1999. This paper, too, has a connection to Wright. It began as a response to a piece in which Rorty criticises Wright’s conclusions in *Truth and Objectivity*. Rorty’s main target is Wright’s defence of truth as a norm distinct from warranted assertibility. Once again, I argue that Wright has the better of this contest, though for reasons he himself does not correctly identify. The paper begins like this.

Rorty begins by telling us why pragmatists such as himself are inclined to identify truth with justification:

Pragmatists think that if something makes no difference to practice, it should make no difference to philosophy. This conviction makes them suspicious of the distinction between justification and truth, for that distinction makes no difference to my decisions about what to do. (Rorty 1995, 19)

Rorty goes on to discuss the claim, defended by Wright, that truth is a normative constraint on assertion. He argues that this claim runs foul of this principle of no difference without a practical difference:

The need to justify our beliefs to ourselves and our fellow agents subjects us to norms, and obedience to these norms produces a behavioural pattern that we must detect in others before confidently attributing beliefs to them. But there seems to be no occasion to look for obedience to an additional norm – the commandment to seek the truth. For – to return to the pragmatist doubt with which I began – obedience to that commandment will produce no behaviour not produced by the need to offer justification. (1995, 26)

Again, then, Rorty appeals to the claim that a commitment to a norm of truth rather than a norm of justification makes no behavioural difference. (Price 2003, 163)

I argue, on the contrary, that the third norm makes a huge behavioural difference:

I want to maintain that in order to account for a core part of ordinary conversational practice, we must allow that speakers take themselves and their fellows to be governed by a norm stronger than that of justification. ... [I]t is a norm which speakers immediately assume to be breached by someone with whom they disagree, independently of any diagnosis of the source of the disagreement. Indeed, this is the very essence of the norm of truth, in my view. It gives disagreement its immediate normative character, a character on which dialogue depends, and a character which no lesser norm could provide.

This fact about truth has been overlooked, I think, because the norm in question is so familiar, so much a given of ordinary linguistic practice, that it is very hard to see. Ordinarily we look through it, rather than at it. In order to make it visible, we need a sense of how things would be different without it. Hence, in part, my reason for beginning with Rorty. Though I disagree with Rorty about the behavioural consequences

of a commitment to ‘a distinction between justification and truth’, I think that the issue of the behavioural consequences of such a commitment embodies precisely the perspective we need, in order to bring into focus this fundamental aspect of the normative structure of dialogue. (2003, 164)

Invoking the Mo’ans again, I go on to argue that ‘some of the basic functions of assertoric discourse could be fulfilled in an analogous way, by a practice which lacked the third norm.’

But it will be clear, I hope, that that practice would not support dialogue as we know it. What is missing – what the third norm provides – is the automatic and quite unconscious sense of engagement in common purpose that distinguishes assertoric dialogue from a mere roll call of individual opinion. Truth is the grit that makes our individual opinions engage with one another. Truth puts the cogs in cognition, at least in its public manifestations. (2003, 165)

Again, let me note the similarity between this view and what we saw in §13.6 that Wright (2021) and MacFarlane (2007) propose for the subjective cases. The similarity is even more apparent in this passage.

The third norm doesn’t just hold open the conceptual space for the idea of improvement. It positively encourages such improvement, by motivating speakers who disagree to try to resolve their disagreement. Without the third norm, differences of opinion would simply slide past one another. Differences of opinion would seem as inconsequential as differences of preference. With the third norm, however, disagreement automatically becomes normatively loaded. The third norm makes what would otherwise be no-fault disagreements into unstable social situations, whose instability is only resolved by argument and consequent agreement – and it provides an immediate incentive for argument, in that it holds out to the successful arguer the reward consisting in her community’s positive evaluation of her dialectical position. If reasoned argument is generally beneficial – beneficial in some long-run sense – then a community of Mo’ans who adopt this practice will tend to prosper, compared to a community who do not. (2003, 174–175)

As I said in Chapter 13, what Wright and MacFarlane both miss, by my lights, is the need for an explanation of this kind in what they think of as the non-subjective cases.

## 14.2 Blackburn’s bull’s-eye

To be fair, Wright and MacFarlane are in excellent company. This deeper issue about disagreement is on almost nobody’s radar. In (Price 2006) I make a similar point against Blackburn’s Ramseyan version of minimalism about truth. I use a different analogy to highlight the question posed by the Mo’ans. Since the point is crucial to the senses in which I want to disagree with Wright below, I want to explain this alternative analogy, as well.

I begin by distinguishing two kinds of examination. In the first (‘autological’) case, the aim is simply to be sincere, to say what you actually believe. If the question is ‘Is Aristotle a Belgian?’, then the right answer is ‘Yes’ if you believe that he is, and ‘No’ if you believe that he isn’t. In the second (‘heterological’) case, sincerity isn’t sufficient, or indeed necessary. You simply get marked right or wrong, depending on whether the examiners agree or disagree with what you say.

I then use this distinction to make a parallel point about assertion.

We can draw a parallel distinction between two kinds of assertion, or conversation. In one kind ('autological conversation'), the aim is simply to give voice to what one actually believes, as accurately as possible. In the other kind ('heterological conversation'), the aim is as if one's assertions were answers in a heterological examination, with one's interlocutors taking the role of the examiners. In the second case but not the first, in other words, it is as if one's utterances are being held to be accountable to an objective standard—a standard that mere sincerity doesn't guarantee. (Price 2006, 608)

Our actual assertoric practices are substantially those of heterological conversation, of course. When we disagree with fellow speakers we regard them as *incorrect*, or *at fault*, even if we are convinced that they have succeeded in giving voice to their own beliefs.

I argue that like most minimalists, Blackburn misses this point. But what he misses is not the fact that conversation has such a normative standard. On the contrary, he calls sharp attention to it, by way of pointing out where some forms of relativism go wrong.

[T]here is no one place from which it is right to look at the Eiffel tower, and indeed no place that is better than another, except for one purpose or another. But when it comes to our commitments, we cannot think this. If I believe that O. J. Simpson murdered his wife, then I cannot at the same time hold that the point of view that he did not is equally good. It follows from my belief that anyone who holds he did not murder his wife is wrong. They may be excusable, but they are out of touch or misled or thinking wishfully or badly placed to judge. I have hit a bull's-eye, which they have missed. (Blackburn 2005, 65–66)

Rather, I say, what Blackburn misses

is what I've highlighted by distinguishing two kinds of conversational games: the fact that there's something important that needs explaining here, viz., that in the game as we actually play it, there is a norm, or a bull's-eye, of precisely this kind. At another point he says that '[t]o make an assertion at all is to put a view into the public space, up for acceptance and rejection' (68). Again, he's right, but he misses the question: why are our assertions treated like this? Why aren't a chap's beliefs treated as entirely his own affair, as it were—as they are, by default, in the game I've called autological conversation? (Price 2006, 609)

My point is that Blackburn's bull's-eye isn't an automatic consequence of a Ramseyan deflationism about truth, because, as I put it, 'the truth predicate it offers us is so thin that it works equally well in either game' (2006, 609).

### 14.3 Truth, Euthyphro, and the Practical Relevance Constraint

Thus I have agreed with Wright, against Horwich, Rorty, and Blackburn, that there is something essentially normative to truth – something missed both by the identification of truth with warranted assertibility, and by what Rorty calls 'Tarski's breezy disquotationalism' (Rorty 1995, 21). But I want to stress two ways in which Wright and I differ. One concerns the no-fault or subjective cases, which we discussed in Chapter 13. The second concerns the appropriate *order of explanation* for the normativity on which Wright and I agree.

Wright appears to say that in the objective or non-minimal cases, the fault we attribute to a speaker with whom we disagree is to be explained in terms of a (perceived) failure of a norm of ‘accurate substantial representation.’ In (Wright 2021) he explains the space he takes his view to leave for the no-fault cases in terms of a contrast with some such substantial norm.

A central contention of *Truth and Objectivity* was that—at least over merely minimally truth-apt discourses—truth need carry no payload of accurate substantial representation. When merely minimally truth-apt claims are at stake, to regard a statement as false need not be to attribute any *representational* fault to someone’s acceptance of it. ... There would be an imputation of fault ... only when “true” demands some kind of richer interpretation. (2021, 439)

Focussing on what Wright regards as the non-minimal cases, let us ask the following questions. Can we explain the normative friction of disagreement in terms of some such ‘payload of accurate substantial representation’? Can the friction be a matter of the two parties each thinking that the other’s claim is *inaccurate*, in some such sense?

In my view, no. For suppose the two parties in question are Mo’ans. Each thinks that they have achieved an accurate substantial representation and that the other has not. Why should that feel any more friction, any more need to engage, about that matter than about their original disagreement? A claim about accurate substantial representation becomes simply one more expression of a purely subjective opinion. That’s the point about the Mo’ans. All their claims are like that, autological rather than heterological.

I think that FFT was insufficiently clear about this point (e.g., in §8.4), but I stressed it in (Price 2003).

If the Mo’ans don’t already care about disagreements, why should they care about disagreements about normative matters? Suppose that we two are Mo’ans, that you assert that *p*, and that I assert that not-*p*. If this initial disagreement doesn’t bother me, why should it bother me when – trying to implement the third norm – you go on to assert that I am ‘at fault’, or ‘incorrect’? Again, I simply disagree; and if the former disagreement doesn’t bite then nor will the latter. And if what was needed to motivate me to resolve our disagreement was my acceptance that I am ‘at fault’, then motivation would always come too late. If I accept this at all, it is only after the fact – after the disagreement has been resolved in your favour.

To get the sequence right, then, I must be motivated by your disapproval itself. This is an important point. It shows that if there could be an assertoric practice which lacked the third norm, we couldn’t add that norm simply by adding a normative predicate. In so far – so very far, in my view – as terms such as *true* and *false* carry this normative force in natural languages, they must be giving voice to something more basic: a fundamental practice of expressions of attitudes of approval and disapproval, in response to perceptions of agreement and disagreement between expressed commitments. I’ll return to this point, for it is the basis of an important objection to certain other accounts of truth. (2003, 173–174)

Later in the same piece, with Rorty’s proposed identification of truth with warranted assertibility in mind, I say this.

I noted above that the same point applies to the normative predicates themselves. If we weren’t already disposed to take disagreement to matter, we couldn’t do so simply by

adding normative predicates, for disagreement about the application of those predicates would be as frictionless as disagreement about anything else. My claim is thus that the notions of truth and falsity give voice to more primitive implicit norms, which themselves underpin the very possibility of ‘giving voice’ at all [in the heterological sense]. In effect, the above argument rests on the observation that this genealogy cannot be reversed: if we start with a predicate – *warrantably assertible* or any other – then we have started too late. (2003, 179, n. 22)

In other words, if we don’t make it constitutive that assertion is a move in a cooperative project – *a game in which players care about normative assessments by other players* – adding a norm of substantial representation doesn’t get us there. It just provides another possible topic for frictionless disagreement. But if we start with the third norm in such a practice, and explain truth and falsity *as ways of giving voice to that norm*, we don’t have to put it in at any later stage.

To sum up, I have called attention to the role in normal conversation of a practice of approving and disapproving of speakers with whom one agrees or disagrees. Without these attitudes, linguistic disagreement simply wouldn’t matter to us in the way that it does. In that case, no addition of normative labels could make it matter, because disagreement about these normative matters would be as frictionless as disagreement about anything else.

Once we have this fundamental practice of treating disagreement in this way, labels that we apply to give voice to these dialogical attitudes of approval and disapproval are effectively ‘true’ and ‘false’, or the special senses of ‘correct’ and ‘incorrect’, applicable to assertions, to which Dummett’s question about the point of truth directs us. And we have an immediate explanation for the fact (in the sense that it is a fact) that truth is what speakers are aiming for, in making their assertions. What they are actually aiming for, unconsciously as it were, is gaining the approval that comes with having their assertions agreed to by others, and avoiding the disapproval of the opposite case. But this looks from the inside like aiming for truth, once ‘true’ and ‘false’ are understood as expressions of the relevant sort of approval and disapproval.

One way to understand this point is in terms of a Euthyphro question.<sup>76</sup> When I agree with a claim made by a fellow speaker, do I endorse what she said because I take it to be true? Or should we rather understand what it is to take a claim to be true in terms of a practice of endorsing the claims of others? I have interpreted Wright, and elsewhere MacFarlane (Price 2022b), as taking the former option. Whereas my money, as I have said, is on the latter.<sup>77</sup> The former option cannot explain the *practical relevance* of taking the claims of another to be true or false – its role in our highly non-Mo’an game of giving and asking for reasons, and our sense that there is a single common bull’s-eye that all participants in the game strive to hit.

This is another example of what I have been calling a Practical Relevance Constraint. As I have stressed at several points, a great advantage of expressivist accounts of philosophically interesting notions is that by *beginning* with the practical role of the notion in question, they avoid the problem of having to explain it later. As we saw in §12.2, a famous case is that of probability, where many writers have argued that the connection to betting and partial beliefs needs to be built into one’s theory from the beginning. Another well-known case, as I noted in §0.1, involves what Michael Smith (1994) calls the *moral problem* – the challenge of explaining how evaluative *beliefs* can be relevant to action, in the way in which we take them to be. Smith himself calls this

---

<sup>76</sup> If I’m not mistaken, Wright (1993) deserves much of the credit for the revivification of the Euthyphro point in contemporary debates.

<sup>77</sup> Note that the issue is which is the more fundamental direction of explanation. In this case, as in many others, the pragmatist can quite well allow that day-to-day explanation often runs left-to-right, once a practice is in place.

feature of evaluative beliefs the *practicality requirement*.<sup>78</sup> We could think of the point I have just made about truth as a close cousin of the moral case. (Under different labels, both the probabilistic and evaluative cases were central concerns of Chapter 4 of FFT, of course.)

This is what makes FFT's account of truth an expressivist account, in the sense of Chapter 11. It *begins* with the practical relevance of truth and falsity – in other words, in Dummett's terms, that of having interpersonal norms of correctness and incorrectness in this kind of linguistic activity. It also makes it a pragmatist theory of truth, in one natural sense, although it is very different from what is often regarded as *the* pragmatist approach to truth (viz., the identification of truth with success, or justification, or some such).

#### 14.4 Why does Wright miss truth expressivism?

I just invoked the Practical Relevance Constraint to argue that *only* an expressivist view can account for the inter-personal, friction-generating normativity of truth. Because Wright insists on the normativity of truth, at least in non-minimal cases, it is something of a puzzle why he misses the expressivist option. In other normative cases, after all, it is a familiar claim that the expressivist has at least a prima facie advantage, in the challenge of connecting normative thought to its motivational consequences.

I find this especially puzzling in papers such as (Wright 1998) and (Wright 2001), in which Wright presents his Inflationary Argument, and then carefully considers the deflationist's options for responding to it. For these purposes, deflationism amounts to the denial that truth need be regarded as a 'substantial' property – in other words, regarded as something that contrasts with the 'thin' property of truth that, as Wright notes, deflationists such as Horwich are happy to allow.<sup>79</sup> As I have said, I take FFT's view of truth to be deflationist in this sense, even though it disagrees with Horwich about the essential function of truth. It takes 'true' to be first and foremost an expression of the crucial third norm of heterological conversation, rather than a mere device of disquotation.

Let's pick up Wright's challenge to deflationism in (Wright 2001), after he has reviewed his argument that truth is a norm distinct from warranted assertibility. Wright then says this.

Minimalism [i.e., Wright's own view] now claims that these facts about assertoric practices stand in need of explanation. In particular, it maintains that it needs to be explained what this further norm of correctness amounts to in such a way that it becomes clear how it and warranted assertibility, although potentially divergent in extension, coincide in normative force: how it can be that warrant is essentially warrant to think that this other norm is satisfied when there is no guarantee that they are always *ω*-satisfied. And such an explanation, it is contended, while it will have to do much more than this, must at least begin by finding something for the truth of a proposition to consist in, a property that it can intelligibly have although there may currently be no reason to suppose that it has it, or may intelligibly lack even though there is reason to

---

<sup>78</sup> "The practicality requirement, as Smith initially formulates it, is the proposition that a person who believes that she would be right to do something is thereby motivated to do it, other things being equal." (Copp 1997, 33)

<sup>79</sup> Thus Wright says that although 'Horwich is ... unwilling ... to deny that truth is a property, it is not, he contends, a "complex property"- not "an ingredient of reality whose underlying essence will, it is hoped, one day be revealed by philosophical or scientific analysis" (*Truth*, 2)'. Accordingly, 'there is, for Horwich, nothing to say about what truth really consists in, no real question for, e.g., correspondence and coherence accounts to address themselves to.' (Wright 1998, 39, n. 9)

think that it has it. Warrant can then be required to be whatever gives a (defeasible) reason to think that a proposition has that property. (Wright 2001, 757)

Wright thus identifies an explanatory challenge, one that he thinks any deflationist view will be unable to meet.

For my part, I have identified a different explanatory task, that of accounting for the friction-generating role of the third norm. I have argued that Wright's apparent order of explanation – that is, an approach that *begins* by identifying a property ('something for the truth of a proposition to consist in'), and then seeks to show that the property in question does the normative work required – cannot possibly do the full job. In the absence of the third norm, disagreements about the property in question would be as frictionless as anything else.

But what about the explanatory task that Wright has in mind? Here I think he misses the possibility of an expressivist approach to the challenge. Suppose we ask *what usage rules would be required*, to generate an assertoric practice of ascribing norms with the character of truth and falsity – in particular, with the relation to warranted assertibility that Wright here describes. The answer looks clear, at least for the initial and main steps. Speakers simply need to ascribe truth and falsity to claims by others with which they agree or disagree, respectively. Crucially, these ascriptions need to carry a positive or negative normative load. In other words, they need to express (a special sort of) approval or disapproval. This immediately makes the practice heterological. It renders each speaker's normative status hostage to the judgement of her peers, *so that it is not guaranteed by warranted assertibility from her own point of view*. Yet at the same time it is obvious why for any individual speaker, warrant is, as Wright says, a defeasible reason to think that a claim is true. That falls straight out of the usage rule.

One thing that makes it puzzling why this option is not on Wright's list is that it seems by my lights so close to options he takes elsewhere. I noted above that his Fregean neo-platonism looks very much like expressivism, in my sense. In this case, Wright's message is that once a certain pattern of use is in place, there are no further deep questions of metaphysics, or deep questions of the semantic kind that do indirect duty for metaphysics – no further deep question about whether there are numbers, for example, or whether number terms 'really' refer to anything. Why doesn't Wright explore the possibility of saying the same kind of thing about truth?

Elsewhere, Wright puts his challenge to the deflationist like this.

[W]hat the deflationist clearly cannot allow is that "true," when used to endorse, has the function of commending a proposition for its satisfaction of some distinctive norm which contrasts with epistemic justification and which only "true" and equivalents serve to mark. For if there were a distinctive such norm, it could hardly fail to be reckoned a genuine property of a proposition that it did, or did not, comply with it. And if the norm in question were uniquely associated with "true" and its cognates, that would be as much as to allow that there was a special property of truth – at which point the deflationary game would have been given away. (Wright 1998, 42)

These days, differing a little from the usage in FFT (e.g., §7.9), my answer is that there is indeed a genuine property, in the same sense in which there are genuine properties in any of the other domains in which it is expressivism that offers the best path to philosophical illumination. But as in these other cases, it is not a 'substantial' property, if by that we mean (something like) a property whose nature is an interesting matter for theoretical investigation. In that sense, the

deflationary game has not been given away. On the contrary, with the normative use of the predicate now firmly in its sights, the deflationary game has found its feet.

### 14.5 Asking the right questions

In Chapter 11 I noted that the difference between expressivism and its rivals often shows up as a difference in which questions are taken to be *worth asking*, at least as starting points. Expressivism wants to start with linguistic or psychological questions. Why do we have this term or concept? What does it do for us? Metaphysics often starts somewhere else, with questions about the *nature* and *reality* of causation, moral properties, or whatever. (As we saw that Ramsey puts it, in contrast to his own approach, ‘Is causation a reality or a fiction; and, if a fiction, is it useful or misleading, arbitrary or indispensable?’)

Let’s contrast this aspect of expressivism with a methodological perspective outlined by Wright, in one of the papers referred to above – a paper in which he is defending, as I would call it, the project of a *metaphysics* of truth. Arguing that such a project need not depend on conceptual analysis in the traditional sense, Wright notes the generality of what he has in mind:

[S]uch issues arise for any putative characteristic,  $\Phi$ . Should we (ontologists) take  $\Phi$  seriously at all, or is some sort of error-theoretic or deflationary view appropriate? If we do take it seriously, should we think of the situation of an item's being  $\Phi$  as purely a matter of how it is intrinsically with that item, or are we rather dealing with some form of relation? Is an item's being  $\Phi$  an objective matter (and what does it mean to say so)? These are analytic-philosophical issues *par excellence*, but their resolution need not await – and might not be settled by – the provision of a correct conceptual analysis. (Wright 1998, 35)

My point, and the point I take to be illustrated by Ramsey’s remarks about causation, is that there is an alternative to what Wright here calls ontology, as a path to a philosophically illuminating understanding of a notion we find in common use. The expressivist simply doesn’t ask *these* questions (at least as starting points, though some of them may turn out to have expressivist readings). Nevertheless, like Ramsey about causation, she claims to show us something interesting, important, and central about the notion  $\Phi$  in question.

Wright certainly has this expressivist perspective on his radar, in some cases. Indeed, as we’ve seen, he appears to take it to be mandatory for any domain in which truth is sufficiently minimal, in his sense. But he does not seem to countenance it in the case of truth itself.<sup>80</sup>

### 14.6 Wright and the road to Canberra

There is one other aspect of Wright’s work that seems relevant here, namely his attraction to a platitudes-based approach to ontological questions that he attributes to Michael Smith (1994) and Frank Jackson (1998). I suspect that Wright’s sympathies for this approach contributes to his sense that there is more work for a non-trivial metaphysics of truth than my expressivism allows. Wright describes the idea like this.

---

<sup>80</sup> More accurately, he doesn’t countenance it as a global view about truth. His own minimalist account of truth counts as expressivist, by my lights, but he doesn’t countenance an expressivist account of what he sees as the richer, normative uses of truth.

Let us call an account based on the accumulation and theoretical organization of a set of ... platitudes concerning a particular concept an *analytical theory* of the concept in question. Then the provision of an analytical theory of truth in particular opens up possibilities for a principled pluralism in the following specific way: in different regions of thought and discourse, the theory may hold good a priori of – may be satisfied by – different properties. If this is so, then always provided the network of platitudes integrated into the theory is sufficiently comprehensive, we should not scruple to say that truth may consist in different things in different such areas: in the possession of one property in one area, and in that of a different property in another. For there will be nothing in the idea of truth that is not accommodated by the analytical theory, and thus no more to a concept's presenting a truth property than its validating the ingredient platitudes. In brief, the *unity* in the concept of truth will be supplied by the analytical theory, and the *pluralism* will be underwritten by the fact that the principles composing that theory admit of *variable collective realization*. (2001, 760–761)

For my part, I associate this methodology with the Canberra Plan (CP). As I noted in §11.4, CP begins with placement or location problems: Where do normativity, meaning, mentality, and other puzzling domains 'fit' in the kind of world described by science? It proposes to answer questions of this kind with a generalisation of the Ramsey-Carnap-Lewis approach to the meaning of theoretical terms.<sup>81</sup>

The proposed solution comes in two steps. At Step 1 we collect the core platitudes about the target entity or property – the entity or property *Target*, let us say – and conjoin them to form the Ramsey sentence, R(*Target*). At Step 2 we ask what in the world *satisfies* or *makes true* the sentence R(*Target*), or to what the term '*Target*' *refers*. As Haukioja (2009) puts it, Step 1 is a matter of '*a priori* analysis of our philosophically interesting everyday concepts and folk theories'; Step 2 of '*consult[ing]* the best scientific (typically, physical) theories to see whether ... referents [for the terms so analysed] are to be found in reality.' Typically, as here, this is understood to mean *natural* reality, the world described by natural science, but this isn't essential to the method. A non-naturalist could also frame her investigations in these terms.

So far as I can see, an expressivist need have no distinctive objection to Step 1 (except the obvious objection, from her point of view, that this platitude-marshalling does little to address what she sees as the interesting explanatory questions).<sup>82</sup> Concerning Step 2, I think that an expressivist should simply deny that it leads to non-trivial results, in general. Clearly, *Target* satisfies R(*Target*), if anything does. Why should we expect anything else, *in general*? I have argued elsewhere that in practice, this expectation rests heavily on an implicit appeal to non-deflationary readings of semantic terms such as *satisfies*, *makes true*, or *refers* – readings that expressivism rejects, of course (see Menzies & Price, 2009; Price, 2004c, 2009a). I have also noted that Blackburn makes a similar point:

Blackburn [1998b, 78] notes that on Ramsey's view, the move from 'P' to 'It is true that P' – "Ramsey's ladder", as he calls it – doesn't take us to a new theoretical level. He remarks that there are "philosophies that take advantage of the horizontal nature of Ramsey's ladder to climb it, and then announce a better view from the top." (Price 2011, 15)

<sup>81</sup> For details and discussion see (Jackson,1998) and the essays in (Braddon-Mitchell and Nola 2009).

<sup>82</sup> Other objections may be raised to Step 1, e.g., that it pays insufficient attention either to the analytic–synthetic distinction, or to the grey zone that results from taking seriously Quinean objections to such a distinction. But such objections may trouble expressivism less than they do CP, because they threaten Step 2, which is where expressivism and CP really differ.

CP is one of the philosophies that Blackburn has in mind, presumably. His point is that talk of truthmakers, denotations, and the like adds nothing to the repertoire of metaphysics, unless the semantic notions in question are more robust than those of Ramsey, Quine, and later deflationists. If a proponent of CP tries to embrace this conclusion, saying that their own use of semantic notions is similarly ‘thin’ – that in effect, Step 2 simply asks ‘What is the X such that  $R(X)$ ?’ – then the expressivist says again that we already have a trivial answer to that question, but no reason in general to expect a non-trivial one.

The qualification about generality is important. Expressivism needn’t challenge some cases, including those of theoretical identification in science. But it will argue that in these cases the semantic characterisation is inessential – the questions can be phrased without it. Some proponents of CP might agree, and argue that the science model is all we need – CP should simply be seen as generalised functionalism. On this view,  $R(\text{Target})$  encodes the causal and functional role of Target, and Step 2 simply enjoins us to look for whatever it is that plays this causal role – a question for natural science, in principle. But as Peter Menzies and I pointed out (Menzies & Price, 2009), this version of CP doesn’t have the generality to which CP aspires – it cannot handle the metaphysics of the causal relation itself, for example.

Deflationism is not the only threat to the attempt to ground a general version of CP on semantic notions. In its general version, the programme has ambitions to encompass the metaphysics of the semantic notions themselves.<sup>83</sup> But then it becomes worryingly circular, for reasons related to Boghossian’s challenges to irrealism about content and to Putnam’s model-theoretic argument (see Price 1998b, 2004c, 2009a). It is an interesting question whether such a charge could be pressed against Wright’s use of the methodology in the case of truth.

But are these objections – the one based on deflationism, or the one based on circularity – relevant as objections to Wright? After all, won’t an expressivist about moral value, or beauty, allow that it makes perfect sense to ask what kinds of things we take to be morally valuable, or beautiful, in particular domains. What does beauty amount to for herbaceous borders, say? Something different, presumably, than for racing bicycles. So if the expressivist doesn’t challenge the platitudes about moral value, or beauty, isn’t the way open to a pluralism about these notions, analogous to Wright’s pluralism about truth?

In the case of moral value, however, there’s a familiar response. One of the platitudes about moral goodness links it to motivation, and hence to behaviour. It is *a priori* that moral worth has a (positive) motivating character: believing a possible action to be good or right provides a motivation for doing it, all other things being equal.<sup>84</sup> Notoriously, it is a challenge for any non-expressivist account of the content of moral belief to explain this fact. Again, this is Smith’s (1994) *practicality requirement*. As I noted, it is an instance of my Practical Relevance Constraint.<sup>85</sup>

So long as this motivational principle is included among our platitudes for *good*, there is a well-known obstacle to taking any merely naturalistic property P to instantiate those platitudes, even

---

<sup>83</sup>Jackson (1998, 2) is explicit about this.

<sup>84</sup> As Rosati (2016) puts it, ‘When P judges that it would be morally right to  $\varphi$ , she is ordinarily motivated to  $\varphi$ ; should P later become convinced that it would be wrong to  $\varphi$  and right to  $\psi$  instead, she ordinarily ceases to be motivated to  $\varphi$  and comes to be motivated to  $\psi$ .’

<sup>85</sup> Some philosophers, including Smith himself, attempt to solve such problems by analysing value in terms of rationality. As I noted in §0.1, and pointed out in FFT (§4.4), there is a corresponding move in the case of probability. In both cases, however, this move simply shifts the bump in the carpet. We then have a practical relevance problem for rationality itself.

if it is clear (at least within some given domain) that competent speakers apply the term ‘good’ to things with property P. For P won’t satisfy the motivational platitude. It won’t be true *a priori* that thinking that an action has the property P is a motivation for doing it.<sup>86</sup>

Similarly, I suggest, for truth. It may be true that we can find properties that tend to *correlate* with ascriptions of truth by competent speakers, in various domains. In that sense, there may well be something in Wright’s thought that truth amounts to different things in different domains. (Again, think of bicycles and herbaceous borders.) But if one of our platitudes for *true* and *false* links the ascription of these terms to the attitudes of approval and disapproval that generate friction in ordinary heterological conversation, it will be an additional step to take such properties to ‘[validate] the ingredient platitudes’, as Wright puts it. On the face of it, as I argued, no such property can do the work of explaining friction.<sup>87</sup>

## 14.7 Reverse-engineering truth

In §13.8 I mentioned MacFarlane’s ‘engineering stance’.

An engineer building a device needs to start from a description of what the device is supposed to do. So, we will start by asking what these putatively assessment-sensitive bits of language are *for*. What role do they play in our lives? (MacFarlane 2014, 310)

Engineering metaphors have become popular in philosophy recently, with much interest in so-called ‘conceptual engineering’. This term seems to have been coined by Creath (1990, 1992), characterising Carnap’s view of linguistic frameworks, or systems of concepts, as devices that we create for sundry purposes.<sup>88</sup> It is also used by Blackburn (1999), describing his view of the task of philosophy.

I suspect that all philosophers and philosophy students share that moment of silent embarrassment when someone innocently asks us what we do. I would prefer to introduce myself as doing conceptual engineering. For just as the engineer studies the structure of material things, so the philosopher studies the structure of thought. Understanding the structure involves seeing how parts function and how they interconnect. It means knowing what would happen for better or worse if changes were made. (Blackburn 1999, 1)

The last sentence here has something of the revisionary, forward-looking character of Carnap’s stance, and of recent conceptual engineers. But in much of Blackburn’s own work the emphasis is backward-looking. Blackburn, like MacFarlane, is interested in the functional origins of our existing concepts, as we find them in ordinary use.

---

<sup>86</sup> Among other things, this is a version of the so-called open question argument. For an excellent recent exposition of this argument against referential semantics, including in particular the ethical case, see Zalabardo (2023).

<sup>87</sup> This shows, incidentally, how it can easily turn out to be the case that there is no non-trivial Step 2, in a particular application of the Canberra methodology. Hence the importance of the triviality and circularity objections above, which are needed to counter the impression that CP is simply offering an informative and uncontroversial answer to a compulsory question.

<sup>88</sup> Carnap himself talks of engineering: ‘I admit that the choice of a language suitable for the purposes of physics and mathematics involves problems quite different from those involved in the choice of a suitable motor for a freight airplane; but, in a sense, both are engineering problems, and I fail to see why metaphysics should enter into the first any more than into the second’ (Carnap 1956, 43, quoted in French 2015, 29).

This explanatory, backward-looking stance has recently been called *reverse-engineering*, as in these remarks by Matthieu Queloz, for example.<sup>89</sup>

To trace a concept back to the needs it has grown out of, one must reverse-engineer the problem to which the concept forms a solution. This requires reconstructing the kind of practical situation in which that problem would arise, and determining what might drive people lacking the concept to invent it. The task is similar to that facing archaeologists who unearth a perplexing artefact: they need to imaginatively reconstruct the human affairs that gave rise to and revolved around it in order to identify the point of the artefact. (Queloz 2021a, 35)

As Queloz notes, the term may be new, but the method is not.

[T]he method of telling pragmatic genealogies of concepts ... cuts across the analytic/continental divide, having been employed by David Hume as much as by Friedrich Nietzsche. More recently, it has been rediscovered notably by philosophers such as Edward Craig, Bernard Williams, Miranda Fricker, and Philip Pettit. These philosophers offer genealogies in the sense that they trace concepts to their origins, but they are *pragmatic* genealogies because they are primarily concerned with *practical* origins: with the function that the concept emerged to discharge – and perhaps continues to discharge if the needs to which it answers have endured. (2021a, 35)

I am very happy, of course, to locate FFT, and Cambridge expressivism in general, within the tradition that Queloz here describes. For the last topic of this chapter, I want to turn to the Cambridge author on Queloz's list whose work I'm most conscious of having neglected, namely, Bernard Williams.

Williams's most celebrated work of genealogy is his last book, *Truth and Truthfulness*, which is subtitled *An Essay in Genealogy* (Williams 2002). Drawing on Queloz's recent discussion (Queloz 2021b), I want to compare Williams's approach to that of FFT. Williams is certainly a near-neighbour, in some respects. But the differences turn out to be significant. FFT's pragmatism is more thoroughgoing than that of *Truth and Truthfulness*, in a sense I'll explain.

#### 14.8 A genealogy of what?

Queloz notes that there is disagreement about what it is, precisely, of which Williams offers us a genealogy: 'Some commentators present it as a genealogy of truth ... Others urge that it is a genealogy of truthfulness as opposed to truth' (Queloz 2021b, 156). Queloz himself proposes a reading that, as he says, 'splits the difference':

[I]t is a genealogical explanation of why we might have come to value the truth—where this means valuing the various states and activities expressive of truthfulness—which is given in terms of the practical value of valuing the truth. To complain, as some have done, that the book's title is 'ironically misleading' in promising a treatment of truth thus misses Williams's point. It is a treatment of truth, only one that carries with it the claim that we should transpose Socratic Questions about the nature of truth into Pragmatic Questions about the point of valuing the truth: instead of gazing up at truth itself and asking what truth *is* or which *theory* of truth we should adopt, we should face the deniers with our eyes set on *human concern with* the truth, bring out how that concern relates to the

---

<sup>89</sup> See also Thomasson (forthcoming), Chapter 8, for a discussion of reverse engineering in this sense.

rest of human psychology and to social and political concerns, and show what valuing the truth does for us. This is the task shouldered by Williams's genealogy. (2021b, 159)

Here the ‘deniers’ are writers such as Rorty, of course. Famously, Rorty dismisses the view that, as he puts it at one point, ‘Truth — or, more precisely, Reality as it is in itself, the object accurately represented by true sentences — [is] an authority we must respect’ (Rorty 1999, 8). Commenting on the same piece later, he proposes that

we see heartfelt devotion to realism as the Enlightenment’s version of the religious urge to bow down before a non-human power. The term “Reality as it is in itself, apart from human needs and interests,” is, in my view, just another of the obsequious Names of God. (Rorty 2007, 134)

Queloz thus reads Williams as challenging Rorty on his own pragmatist turf – as responding to Rorty’s denial of truth by setting our eyes on ‘*human concern with the truth*’, and showing in pragmatic terms ‘what valuing the truth does for us.’

This reading immediately suggests some similarities between Williams’s project and mine. As we saw above, I, too, took Rorty as an opponent about truth, and challenged him on pragmatic grounds. I argued that Rorty neglects the practical role of truth in human conversation. Hence, I claimed, Rorty simply fails to see how much we would lose in abandoning truth, even if that were possible.

Moreover, it is certainly true of me, as of Williams, that I’m not ‘gazing up at truth itself and asking what truth *is* or which *theory* of truth we should adopt’, as Queloz puts it, and that my eyes are firmly ‘set on human concern with the truth.’ Nevertheless, my approach is in some ways different from Williams. Among other things, I claim to offer a genealogy *of truth itself*, in a way in which Williams does not.

## 14.9 Williams and the State of Nature

Williams’s genealogy begins with an imagined State of Nature, which he introduces like this.

[A]t our entry to the State of Nature, we concentrate on the basic phenomenon of people’s beliefs contributing to there being a shared pool of information: various observers are in different situations, and they then transport to the pool beliefs (in the favourable case, true beliefs ...) which each of them has acquired from being in that situation. (2002, 43–44.

As Williams notes, it is a similar conception of a State of Nature to that of Edward Craig, in his proposed genealogy of knowledge claims (Craig 1990). Williams discusses Craig’s proposal at some length, and uses it among other things to illustrate the sense in which such a genealogy need not claim to be realistic. That point will be helpful below (§14.12), but for now, I want to use Craig’s example to illustrate a different point about Williams’s version.

Williams goes on to argue that in the State of Nature he describes, instrumental values of *Accuracy* and *Sincerity* (more on these in a moment) will lead speakers to value *truthfulness*; to seek to contribute *true* beliefs to the pool. The point I want to stress is that Williams’s State of Nature thus *presupposes* truth, in two ways. First, it presupposes that the items contributed to the pool are truth-evaluable. And second, more importantly for my purposes, it presupposes that speakers in the State of Nature *have the notion of truth in their conceptual repertoire*, in order to come to value it.

I can explain these points by comparison with Craig's version of the State of Nature, which Craig describes like this.

[T]here is a firmly fixed point to start from. Human beings need true beliefs about their environment, beliefs that can serve to guide their actions to a successful outcome. That being so, they need sources of information that will lead them to believe truths. They have 'on-board' sources, eyes and ears, powers of reasoning, which give them a primary stock of beliefs. It will be highly advantageous to them if they can also tap the primary stocks of their fellows—the tiger that Fred can see and I can't may be after me and not Fred—that is to say, if they act as informants for each other. On any issue, some informants will be better than others, more likely to supply a true belief. (Fred, who is up a tree, is more likely to tell me the truth as to the whereabouts of the tiger than Mabel, who is in the cave.) So any community may be presumed to have an interest in evaluating sources of information; and in connection with that interest certain concepts will be in use. The hypothesis I wish to try out is that the concept of knowledge is one of them. To put it briefly and roughly, the concept of knowledge is used to flag approved sources of information. (Craig 1990, 11)

It is easy to imagine two variants of this story, in which it is made explicit that the notion of true belief is (at best) in the background, rather than part of the assumed conceptual repertoire of the creatures concerned. One variant would retain humans, but focus instead on *practical* expertise – how to evade tigers, for example. Some humans are better at that than others, and it will be helpful to a community to be able to identify the experts. So we have the beginnings of genealogy much like Craig's, in which truth simply isn't in the picture, at the first order level.<sup>90</sup>

The second variant would replace humans with honeybees, whose *actual* State of Nature includes a complex and survival-critical procedure for pooling information about the location of potential new nest sites, when a hive divides (Seeley 2010). At a stretch, throwing Davidsonian caution to the wind, we might describe this in terms of the importance of the bees acquiring *true beliefs* – but this would belong to our theoretical perspective, of course, not to any conceptual repertoire we attribute to the bees themselves.

States of Nature of these two kinds wouldn't do for Williams's purposes. As he himself says, his project is to explain 'human concerns with the truth' (2002, 61). Humans can't be concerned about the truth unless (i) there is a truth-evaluable matter on the table, and (ii) they have truth, or something equivalent, in their conceptual toolkit. So the information being pooled in Williams's State of Nature must not only be truth-evaluable, unlike the practical expertise in our first variant of Craig's story; it must also be truth-evaluable *by the poolers themselves*, unlike the information in our second variant.

In effect, then, we now have three different kinds of information pool in play, and it will be helpful to have a label for the one that Williams needs. Since it depends on the participants' possession of a *heterological* notion of truth, as I put it earlier – Blackburn's bull's-eye – let's call it the Heterological Information Pool, or HIP, for short.

As we'll see, Williams doesn't have on his radar the idea that the HIP itself, or the heterological notion of truth on which it depends, might be a target for genealogy. On the contrary, he takes the view that truth needs to be regarded as primitive, for Davidsonian reasons. By my lights, this

---

<sup>90</sup> Unless we insist on some tortuous translation of knowing *how* into knowing *that*, at any rate.

looks under-ambitious. I think there's interesting work to be done in explaining how such creatures get to the point of thinking of themselves as joined at the HIP, so to speak – work that shows us something central about truth. In failing to see the pool itself as a proper target for genealogy, Williams throws his imagined creatures in at the deep end. My account offers a gradualist alternative.

#### 14.10 Accuracy, Sincerity, and the Limits of Genealogy

Williams thus builds in the assumption that the beliefs transported to the pool may be true or false, correct or incorrect, *and regarded as such by the participants themselves*. Williams's project is to explain why they find an instrumental value in contributing true beliefs to the pool.

With this, we have the idea that one person may ... do better in this [i.e., in transporting beliefs to the pool] than another; and that people maybe discouraged or encouraged ... with respect to this. One significant way in which these processes will work is through encouraging individual dispositions to do these things well. Since we are considering people, who have beliefs and desires and intentions, and who may or may not express their beliefs, it is already natural to think of these dispositions as falling into two different kinds.

One kind of disposition [Accuracy] applies to their acquiring a correct belief in the first place, and their transporting that belief in a reliable form to the pool. ... [The second group of dispositions [Sincerity] centrally contains the motivation, if one is purporting to tell someone something and the circumstances are right, to say what one actually believes.

The distinction between these two kinds of disposition is not only natural but also, I think, basic to questions about truthfulness, and I shall register this by treating each of the two groups, a lot of the time, as one generic disposition. I shall label the two kinds of disposition respectively as Accuracy and Sincerity. ... I shall call Accuracy and Sincerity the two basic virtues of truth. (2002, 44)

As I have already noted, Williams emphasises that he is seeking to explain *human concerns with truth*, not truth itself. Indeed, he makes dismissive remarks about the possibility of making truth itself the focus.

The inquiry is, then, into human concerns with the truth. A basic form of that concern lies in the virtues of truth. ... One thing I shall not consider ... is the history of *the concept of truth*, because I do not believe that there is any such history. The concept of truth itself—that is to say, the quite basic role that truth plays in relation to language, meaning, and belief—is not culturally various, but always and everywhere the same. ... There are indeed scholarly books that describe themselves as histories of the concept of truth, but they typically describe conceptions, varying over time, of belief-formation, or of knowledge, or of the metaphysical structure of the world. Often, they are histories of philosophical theories of the truth. ... The present point is that philosophical theories of truth ... quite certainly have a history, whereas the concept of truth itself does not. (Williams 2002, 61–62)

These are surprising claims for a genealogist, in my view. There's a trivial sense in which the concept of truth has a history, because, like all our concepts, our pre-linguistic ancestors did not possess it. In the light of this, the claim that truth is 'always and everywhere the same' seems cavalier. How could we know, if we're ignoring that part of the history?

The claim is also surprising for a different reason. When Williams was writing, Crispin Wright and others had been calling attention to the possibility of *differences* in the way in which truth is used in different domains, even within a single language community. With this work on the table, shouldn't Williams have regarded it as an empirical matter, for linguists and anthropologists as well as philosophers, whether truth was 'everywhere the same'?

Why does Williams take such an uncompromising stance on these issues? As I noted, he says that he is following Davidson's lead.

What about truth itself? If we are going to say that beliefs and assertions ... aim to be true, or, as the State of Nature story has assumed, that it is a good idea ... that they should be true, should we not say something about what it is for them to be true?

We should say something, but not very much. In particular, we should resist any demand for a *definition* of truth, principally because truth belongs to a ramifying set of connected notions, such as meaning, reference, belief, and so on, and we are better employed in exploring the relations between these notions than in trying to treat one or some of them as the basis of the others. It is also true that if any of these notions has a claim to be more basic and perspicuous than the others, it is likely to be truth itself. As Davidson has put it:

[W]e cannot hope to underpin [the concept of truth] with something more transparent or easier to grasp. Truth is, as G.E. Moore, Bertrand Russell and Gottlob Frege maintained, and Alfred Tarski proved, an indefinable concept. This does not mean we can say nothing revealing about it: we can, by relating it to other concepts like belief, desire, cause and action. Nor does the indefinability of truth imply that the concept is mysterious, ambiguous, or untrustworthy [Davidson 1996, 265]. (Williams 2002, 63)

I have two comments here. First, the genealogist is not interested in *definition* in any case. How is it relevant to the possibility of a genealogy of truth that a different project, one that the genealogist has in any case rejected, is impossible? Second, while truth may indeed be part of a package, it would be absurd to suggest that the package as whole does not have a history. Our ancestors didn't always play the game of giving and asking for reasons, of course.<sup>91</sup>

To me, then, Williams's reliance on a Davidsonian *Schlagbaum* seems unconvincing. To back this up, I want to give a positive reason for thinking that we genealogists should be able to find our way around this barrier.<sup>92</sup> It rests on what I'll call *edge cases*. These involve claims and beliefs that don't have a natural entitlement to find themselves in the HIP, but rather, in senses that will become clear, need to be placed there by us. By asking what we need to *do* to achieve this, we gain traction on the genealogy of truth, and of the HIP itself. As we'll see, Williams himself has the relevant considerations in play. By my lights, he simply fails to see their potential.

### 14.11 Edge cases

I'll mention three of these edge cases. For the first, let's begin with these remarks from Quelo on Ramsey on probability.

<sup>91</sup> 'We'll start out by speaking in simple declarative sentences', as the cavemen in a *New Yorker* cartoon describe the first step (Cotham 2007).

<sup>92</sup> To avoid the Davidsonic Boom, as we might put it, with a nod to an Oxford joke of the 1970s.

As Ramsey saw it, trying to match up the concept of probability with perceivable bits of the world was not the way to elucidate it. ... The concept of probability was less like an impression left in our minds by something already out there, and more like a device we had built to navigate the world more successfully.

Hence, Ramsey proposed to look at the function performed by the concept in our thought and talk. What work does the concept do for us? What does it allow us to achieve that we could not achieve without it? We can seek to demystify the concept of probability by relating it to human needs rather than to objective relations. (Queloz 2021a, 34)

Good, but then how do beliefs ‘about probability’ get into the information pool? More precisely, what do we have to *do*, to treat our probability judgements as the kind of items that we can properly offer to the pool? In effect, this is an instance of the quasirealist’s question. Once we answer it, we’ll have the beginnings of an account of the practices that *underlie*, or *enable*, the kind of HIP presupposed by Williams’s State of Nature.

For the second case, I noted in §13.6 that Wright argues that judgements of taste only give rise to faultless disagreements (Wright 2021; see also Kaspers 2023). In these cases, Wright and Kaspers argue, we take there to be no *objective* truth, for which such claims aim. Yet we can make sense of a personal sense of *accuracy*. We know what it is like to *think* that sushi is delicious, and to discover that one is wrong by trying some. And we can certainly make sense of *sincerity*. So what does it take to add the additional, *objective* element – for us to see ourselves as contributing to a common pool of information? Again, if we can answer the question in this case, we’ll learn something about the practices that underlie the HIP in general.

My third case comes from Williams himself. Williams makes much use of a doubly-historical example. He argues that in the history of Western thought, ‘Thucydides introduces a new, “objective” way of understanding history’, as Clancy Martin puts it, in a review of Williams’s book (Martin 2003). Williams’s claim is disputed, but that needn’t concern us here, except in the sense that it serves as an example for the following point. If the claim were true, then there would have to be something that Thucydides *did* with historical claims, that wasn’t done by his predecessors (such as Herodotus). What did he have to do? In effect, again, this is the quasirealist’s question. It is also Wright’s question, if historical claims are not to be a matter of taste.

In one sense, Williams is well aware of this question. He asks it himself, after attributing the view of Thucydides in question to Hume.

David Hume wrote: “The first page of Thucydides is, in my opinion, the commencement of real history. All preceding narrations are so intermixed with fable, that philosophers ought to abandon them, in a great measure, to the embellishment of poets and orators.” [Hume 1985, 422]. It is a familiar judgement, but what exactly does it mean? ... What exactly, on Hume’s view, did Thucydides *do* in order to start real history? (151–152)

Similarly, Williams speaks of ‘the change from the local to the objective view of the past’ (164), and is careful, when he first introduces this terminology, not simply to build the answer by fiat into the State of Nature.

We need to ask what might actually be involved in forming the objective conception, and we cannot consider that question if we assume that the conception must be there from the beginning and that only stupidity or lack of reflection stops people from operating

with it. So we shall put no more into the State of Nature than what is readily implied by the capacities and interests that we have already assumed. This will include a locally restricted or perspectival conception of the past. (51)

Later, he says that ‘the change from the local to the objective view of the past’ did not, of course, ‘announce itself ... as a metaphysical discovery.’

It was expressed, rather, in a change in people’s practice (indeed, unless it were expressed in that way, there would be nothing to possess a metaphysical substance). So what was involved in this new practice? Here it is essential that there is more to it than merely a change in the way people talk. It is not just that they now apply words that can be translated as “true” and “false” to statements about the remoter past, including stories about the gods. They did that before. ... What matters is the force of such words, what turns on saying “true” rather than “false.” (164)

The last sentence here is strikingly reminiscent of the remark from Dummett which, as I noted in §0.1, was influential in my own early efforts to think about these matters.

At one time it was usual to say that we do not call ethical statements ‘true’ or ‘false’, and from this many consequences for ethics were held to flow. But the question is not whether these words are in practice applied to ethical statements, but whether, if they were so applied, the point of doing so would be the same as the point of applying them to statements of other kinds, and, if not, in what ways it would be different. (Dummett 1959, 144)

Williams knew this classic piece of Dummett’s, of course; he cites it in an earlier chapter. So it is tempting to see Dummett’s influence here.

Williams continues like this.

[I]t cannot merely be that after this development there will be two styles of narration about the past: that in some cases people just come out with stories about the past, as they come out with other things that they intend to assert, but in other cases they bracket their narration with some disclaiming formula suitable to myth, such as that natural legacy from the Herodotean world, “once upon a time.” They may well come to do this, but these distinctions among speech-acts are not self-sufficient: both we and they need to know what turns on telling stories in these different modes, how the social consequences differ. What responsibilities does one take on by telling a tale in what, at this stage, we may call the mode of truth rather than in the mode of myth? (2002, 164)

This question leads Williams into a careful discussion of the distinction between these different sorts of speech acts. He notes, among other things, some aspects of the difference already identified by Thucydides. But for present purposes, I want to stress what Williams’s answer is *not*. He does not say, and it would be unhelpful to say, that the difference is that after Thucydides we in the West came to have *beliefs* about the past, the kind of attitudes that can be true or false, with the kind of content that belongs in the information pool. As I’m sure Williams would agree, that gets things the wrong way round. We came to *treat* the past as a matter for (objective) belief, as grist for the information pool, in its HIP form. But what is it to do that? What turns on it? As Williams says, what responsibilities come with it?

My point is that if we can answer these questions in edge cases, then we will have an answer that will work in the core cases, too. We will have learnt something about what it is to *think of oneself* as contributing to a common, heterological information pool – about what we have to *do*, to possess in our conversational practice the concept of truth, which Williams takes to be always and everywhere the same. In other words, we will have worked our way around the Davidsonian barrier, and gained sight of a possible genealogy on the other side.

My version of this post-Davidsonian genealogy is the one I came to by thinking both about Dummett's remark about the point of truth, and, in a sense, about Williams's own early observation (Williams 1966; see §0.1 and §2.2) about the need for what he termed a *substantial* theory of truth – a notion of truth fit for endorsing fact-stating or assertoric claims, but not speech acts of other kinds.

My claim was that assertoric speech acts are taken to be subject to a distinctive norm of correctness, whose primary use-rule is that disagreement is to be interpreted, by default, as a sign of fault. 'True' and 'false' are labels for this kind of correctness and incorrectness, and hence marks of approval or disapproval that we use with those with whom we agree or disagree. Even in its crudest form, this proposal puts social consequences at the core of things, exactly as Williams recommends.

In Williams's terms, then, I have proposed an answer to the question how a community can *create* an explicit heterological information pool, if it doesn't already have one. I have proposed a more basic genealogy – a genealogy *of* the central posit of Williams's State of Nature. With this option in view, we don't need to jump in at the deep end. We can work our way into the HIP from more basic beginnings. We can treat the HIP itself as a social artefact, in effect, by identifying the crucial thing we need to do, to sustain such an idea.

As a bonus, the proposal immediately explains the value of truth. We seek the truth because we seek the approval of our fellow speakers. By my lights, then, Williams makes the same wrong choice as Wright with respect to the Euthyphro question. Assuming truth as a Davidsonian primitive, he then faces the challenge of explaining why we should value it. Once again, the expressivist order of explanation comes to the rescue. By making 'true' an expression of (positive) valuation, we put the value in at the beginning. The important task is to reverse-engineer this particular, highly specific form of social valuation – but that was the role of my proposed genealogy.

#### 14.12 Rorty and the impossible Mo'ans

My more basic State of Nature is personified by my Mo'ans, who lack the third norm. For them, all assertoric claims work as Wright and Kaspers think that judgements of taste work. All their assertions are 'in subjective mode'. I asked what the Mo'ans need to *do*, to add the third norm, and what difference it makes. My answer was that they simply need to treat disagreement as a sign of fault – i.e., by default, to be critical of speakers with whom they disagree, and approving of those with whom they agree. This makes disagreements unstable, in a new way. It provides the group with an incentive for resolving them, and the individual with an incentive for defending her or his own views.

I was aware, of course, that it is doubtful whether the Mo'ans are a realistic possibility. But I felt that these doubts did not invalidate their use as a kind of thought experiment. Interestingly, Rorty focussed on this point, in comments on 'Truth as convenient friction' later published in (Rorty and Price 2010). He says this, for example.

Price asks us to imagine a community in which there are no attempts at intersubjective justification, but in which its members nevertheless express “the kind of behavioral dispositions which we would characterize as beliefs ... by means of a speech act we might call the *merely-opinionated assertion* (MOA, for short)” [Price 2003, 172]. He admits that one might doubt the possibility of such a community: perhaps, he says, “a truth-like norm is essential to any practice which deserves to be called linguistic” [Price 2003, 168, n. 8]. But he thinks this possibility irrelevant to his thought experiment.

It seems relevant to me. I doubt that we can tell a plausible story about a Mo’an community. In particular, I do not see why a radical interpreter would construe as *assertions* the noises made by organisms that never attempt to correct one another’s behavioral dispositions—never try to get others to make the same noise they do. I would advance arguments familiar from Wittgenstein, Davidson, and Brandom to urge that there must be social cooperation on projects of shared interest before language can get very far off the ground. One cannot justify by own lights if one does not know what it is to justify by the lights of others. Price’s “chatter of disengaged monologues” [Price 2003, 166] is possible only as an enclave within a culture in which there is lots of engaged dialogue. (Rorty and Price 2010, 255)

To me, this response seemed to count in my favour. In (Price 2003) I had argued, against Rorty’s objection to Wright, that the third norm makes a very big behavioural difference. As I put it: ‘It gives disagreement its immediate normative character, a character on which dialogue depends, and a character which no lesser norm could provide’ (2003, 164). Here, Rorty seemed to be conceding that point, in objecting that the Mo’an’s washed-out substitute for dialogue was simply impossible. I put the point like this.

For my purposes, what mattered about the Mo’ans was that by seeing what their linguistic practice would lack, we see what truth adds to our own. What’s missing for the Mo’ans—what the third norm provides for us—is (as I put it) “the automatic and quite unconscious sense of engagement in common purpose that distinguishes assertoric dialogue from a mere roll call of individual opinion.” Let’s agree, with Rorty, that when we consider the Mo’ans in the light of “arguments familiar from Wittgenstein, Davidson, and Brandom,” we realize that there can be no such community. Removing that sense of engagement amounts to removing anything that might count as an assertion, or indeed as an expression of opinion, in the full-blown sense of the term. This is no reason to forget the lesson we learnt by trying to imagine the Mo’ans—on the contrary, as in many cases, the point of the thought experiment lies *precisely* in the fact that it leads us, in thought, to an impossible destination. (The lesson lies in the nature of the impossibility.) (Rorty and Price 2010, 257)

This reply to Rorty was written around 2008, in response to comments that Rorty had given to the editors of the volume concerned in 2005. I now realise that if I had read Williams’s discussion of the functions of genealogy at that point, I could have put my point in his terms. This is what he says about Craig’s project, for example, in a subsection headed ‘How Can Fictions Help?’ (2002, 31) – I could have cited this in my response to Rorty.

Craig’s example, like my own State of Nature story, is an example of what I shall call an “imaginary genealogy”—“imaginary,” because ... there are also historically true genealogies. Imaginary genealogies typically suggest that a phenomenon can usefully be treated as functional which is not obviously so. Moreover, they resemble a larger class of explanations (including those given by natural selection theory) in explaining the

functional in terms of the non-functional or, perhaps, in terms of the more primitively functional. The power of imaginary genealogies lies in introducing the idea of function where you would not necessarily expect it, and explaining in more primitive terms what the function is. (2002, 32)

Finally, what does Williams mean by ‘function’ here? He introduces the term for the first time in the preceding paragraph, at the beginning of his discussion of Craig’s example.

[Craig’s] State of Nature gives an explanation, but what it explains, the concept of knowledge, does not look as though it had been designed. On the contrary: before the story, one may well never have asked what the function of the concept is, and that is part of the point. Craig’s story answers, in its fictional terms, the question “Why should we have a concept such as the concept of knowledge?” and, by answering it, suggests the question itself. What the question introduces is the notion of function, and that step itself does some of the work. If one sees the concept of knowledge as having a function—in particular, a function in relation to very basic needs—this in itself helps one to see why it has the features it has, and can discourage one from less fruitful approaches. (2002, 31–32)

Williams doesn’t himself use the term ‘engineering’ in *Truth and Truthfulness*, but I think that it is clear that that metaphor, in its current usage, is a very good fit for what he has in mind. And Williams’s own formulation of the lesson here – that this approach ‘can discourage one from less fruitful approaches’ – is a very good fit for the central message of FFT concerning truth, and for Cambridge expressivism with respect to many other topics.

## Bibliography

- Adams, E. 1965. *The Logic of Conditionals*.
- Adams, E. 1975. *The Logic of Conditionals*. Dordrecht: Reidel.
- Ahmed, A. 2014. *Evidence, Decision and Causality*. Cambridge: Cambridge University Press.
- Albury, R., and P. Slezak, eds. 1988. *Computers Brains and Minds: Papers in Cognitive Science*. Dordrecht: Kluwer Academic Publishers.
- Anscombe, E. 1975. 'The First Person'. In *Mind and Language: Wolfson College Lectures 1974*, edited by S. Guttenplan, pp. 45–64. Oxford: Clarendon Press.
- Beebe, H., and A. Fisher, eds. 2020. *Philosophical Letters of David K. Lewis, Volume 2: Mind, Language, Epistemology*. Oxford: Oxford University Press.
- Blackburn, S. 1971. 'Moral Realism'. In *Morality and Moral Reasoning*, edited by J. Casey, pp. 101–24. London: Methuen.
- Blackburn, S. 1984. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Oxford University Press.
- Blackburn, S. 1985. 'Supervenience Revisited'. In *Exercises in Analysis: Essays by Students of Casimir Lewy*, edited by I. Hacking, pp. 47–67. Cambridge: Cambridge University Press.
- Blackburn, S. 1987. 'Morals and Modals'. In *Fact, Science and Morality*, edited by G. Macdonald and C. Wright, pp. 119–41. Oxford: Basil Blackwell. Reprinted in Blackburn 1993, pp. 52–74.
- Blackburn, S. 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.
- Blackburn, S. 1994. *The Oxford Dictionary of Philosophy*. Oxford: Oxford University Press.
- Blackburn, S. 1998a. 'Wittgenstein, Wright, Rorty and Minimalism?'. *Mind* 107: pp. 157–82.
- Blackburn, S. 1998b. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.
- Blackburn, S. 1999. *Think: A Compelling Introduction to Philosophy*. New York: Oxford University Press.
- Blackburn, S. 2013. 'Pragmatism: All or Some?'. In *Expressivism, Pragmatism and Representationalism*, edited by H. Price et al., pp. 67–84. Cambridge: Cambridge University Press.
- Boghossian, P. 1990. 'The Status of Content'. *Philosophical Review* 99: pp. 157–99.
- Böddeling, A. 2019. 'Cognitivism and Metaphysical Weight: A Dilemma for Relaxed Realism'. *Australasian Journal of Philosophy*, published online 18 Aug 2019. <https://doi.org/10.1080/00048402.2019.1651355>
- Boncompagni, A. 2017. 'The "Middle" Wittgenstein (and the "Later" Ramsey) on the Pragmatist Conception of Truth'. In *The Practical Turn: Pragmatism in the British Long Twentieth Century*, edited by C. Misak and H. Price, pp. 29–44. Oxford: Oxford University Press.

- Braithwaite, R. B. 1966. ‘Why Is It Reasonable to Base a Betting Rate upon an Estimate of Chance?’. In *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science*, edited by Y. Bar-Hillel, pp. 263–73. Amsterdam: North Holland.
- Braddon-Mitchell, D. 2004. ‘How Do We Know It Is Now Now?’. *Analysis* 64: pp. 199–203.
- Charlow, N. 2014. ‘The Problem with the Frege-Geach Problem’. *Philosophical Studies* 167: pp. 635–65.
- Chrisman, M. 2007. ‘From Epistemic Contextualism to Epistemic Expressivism’. *Philosophical Studies* 135: pp. 225–54.
- Chrisman, M. 2008. ‘Expressivism, Inferentialism, and Saving the Debate’. *Philosophy and Phenomenological Research* 77: pp. 334–58.
- Copp, D. 1997. ‘Belief, Reason, and Motivation: Michael Smith’s The Moral Problem’. *Ethics* 108: pp. 33–54. <https://doi.org/10.1086/233787>
- Creath, R. 1992. ‘Carnap’s Conventionalism’. *Synthese* 93: pp. 141–65.
- Davidson, D. 1967. ‘Truth and Meaning’. *Synthese* 17: pp. 304–23.
- Davidson, D. 1968. ‘On Saying That’. *Synthese* 19: pp. 130–46. Reprinted in Davidson 1984, pp. 93–108.
- Davidson, D. 1974. ‘Belief and the Basis of Meaning’. *Synthese* 27: pp. 309–24. Reprinted in Davidson 1984, pp. 141–54.
- Davidson, D. 1984. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- Davidson, D. 1996. ‘The Folly of Trying to Define Truth’. *Journal of Philosophy* 93: pp. 263–78.
- Davidson, D., and G. Harman, eds. 1972. *Semantics of Natural Language*. Dordrecht: Reidel.
- De Caro, M., and D. Macarthur, eds. 2010. *Naturalism and Normativity*. New York: Columbia University Press.
- Dennett, D. 1978a. *Brainstorms*. Cambridge, MA: MIT Press.
- Dennett, D. 1978b. ‘Three Kinds of Intentional Psychology’. In *Reduction, Time and Reality*, edited by R. Healey, pp. 37–61. Cambridge: Cambridge University Press.
- Devitt, M. 1984. *Realism and Truth*. Oxford: Basil Blackwell.
- Dreier, J. 1996. ‘Expressivist Embeddings and Minimalist Truth’. *Philosophical Studies* 83: pp. 29–51.
- Dreier, J. 2004. ‘Meta-ethics and the Problem of Creeping Minimalism’. *Philosophical Perspectives* 18: pp. 23–44.
- Dreier, J. 2018. ‘The Real and the Quasi-Real: Problems of Distinction’. *Canadian Journal of Philosophy* 48: pp. 532–47. [10.1080/00455091.2018.1432394](https://doi.org/10.1080/00455091.2018.1432394)
- Dummett, M. 1959. ‘Truth’. *Proceedings of the Aristotelian Society* 59: pp. 141–62. Reprinted in Dummett (1978), pp. 1–24.

- Dummett, M. 1964. 'Bringing about the Past'. *Philosophical Review* 73: pp. 338–59.
- Dummett, M. 1973. *Frege: Philosophy of Language*. London: Duckworth.
- Dummett, M. 1975. 'What Is a Theory of Meaning? (I)'. In *Mind and Language*, edited by S. Guttenplan, pp. 97–138. Oxford: Clarendon Press.
- Dummett, M. 1978. *Truth and Other Enigmas*. London: Duckworth.
- Dworkin, R. 2011. *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press.
- Eddington, A. 1928. *The Nature of the Physical World*. Cambridge: Cambridge University Press.
- Ellis, B. 1980. 'Truth as a Mode of Evaluation'. *Pacific Philosophical Quarterly* 61: pp. 85–99.
- Evans, G., and J. McDowell, eds. 1976. *Truth and Meaning: Essays in Semantics*. Oxford: Oxford University Press.
- Field, H. 1972. 'Tarski's Theory of Truth'. *Journal of Philosophy* 69: pp. 347–75. Reprinted in Platts (1980), pp. 83–110.
- Field, H. 1978. 'Mental Representation'. *Erkenntnis* 13: pp. 9–16.
- Fine, A. 1986. 'Unnatural Attitudes: Realist and Instrumentalist Attachments to Science'. *Mind* 95: pp. 149–79.
- Fine, K. 2001. 'The Question of Realism'. *Philosophers' Imprint* 1: pp. 1–30.
- Fodor, J. 1985. 'Fodor's Guide to Mental Representation'. *Mind* 94: pp. 76–100.
- Fodor, J. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- Frege, G. 1960. 'Negation'. In *Translations from the Philosophical Writings of Gottlob Frege*, edited by P. Geach and M. Black. Oxford: Oxford University Press.
- French, C. 2015. 'Philosophy as Conceptual Engineering: Inductive Logic in Rudolf Carnap's Scientific Philosophy'. PhD diss., University of British Columbia.
- Gert, J. 2018. 'Neo-Pragmatism, Representationalism and the Emotions'. *Philosophy and Phenomenological Research* 97: pp. 454–78.
- Gert, J. 2023a. 'Neo-Pragmatist Semantics'. *Philosophy and Phenomenological Research* 106: pp. 107–35. <https://doi.org/10.1111/phpr.12842>
- Gert, J., ed. 2023b. *Neopragmatism: Interventions in First-Order Philosophy*. New York: Oxford University Press.
- Gibbard, A. 1986. 'An Expressivistic Theory of Normative Discourse'. *Ethics* 96: pp. 472–85.
- Gibbard, A. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.
- Gibbard, A. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

- Gibbard, A. 2015. 'Global Expressivism and the Truth in Representation'. In *Meaning without Representation: Essays on Truth, Expression, Normativity, and Naturalism*, edited by S. Gross, N. Tebben, and M. Williams, pp. 210–23. New York: Oxford University Press.
- Gibbard, A., and W. Harper. 1978 [1981]. 'Counterfactuals and Two Kinds of Expected Utility'. In *Foundations and Applications of Decision Theory*, edited by C. A. Hooker, J. L. Leach, and E. F. McClennan, pp. 125–62. Dordrecht: D. Reidel.
- Grover, D., J. Camp, and N. Belnap. 1975. 'A Prosentential Theory of Truth'. *Philosophical Studies* 27: pp. 73–125.
- Habermas, J. 1970. 'Towards a Theory of Communicative Competence'. *Inquiry* 13: pp. 360–75.
- Hacking, I., ed. 1985. *Exercises in Analysis: Essays by Students of Casimir Levy*. Cambridge: Cambridge University Press.
- Hájek, A. 2016. 'Deliberation Welcomes Prediction'. *Episteme* 13: pp. 507–28.
- Haldane, J., and C. Wright, eds. 1993. *Reality, Representation, and Projection*. New York: Oxford University Press.
- Hale, B., and C. Wright. 2001. *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Oxford University Press.
- Hamblin, C. L. 1987. *Imperatives*. Oxford: Basil Blackwell.
- Hannon, M. 2019. *What's the Point of Knowledge?: A Function-First Epistemology*. Oxford: Oxford University Press.
- Harman, G. 1986. *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press.
- Hesse, M. 1978. 'Habermas' Consensus Theory of Truth'. *Proceeding of the Philosophy of Science Association* 1978: pp. 373–96.
- Horwich, P. 1990. *Truth*. Oxford: Basil Blackwell.
- Hawthorne, J., and H. Price. 1996. 'How to Stand up for Non-Cognitivists'. *Australasian Journal of Philosophy* 74: pp. 275–92. Reprinted in Price (2011c).
- Horwich, P. 1993. 'Gibbard's Theory of Norms'. *Philosophy and Public Affairs* 22: pp. 67–78.
- Horwich, P. 2005. 'The Frege-Geach Point'. *Philosophical Issues* 15: pp. 78–93.
- Humberstone, L. 1991. 'Critical Notice of F. Jackson, Conditionals'. *Philosophy and Phenomenological Research* 50: pp. 227–34.
- Hume, D. 1978. *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed. with revisions by P. H. Nidditch. Oxford: Oxford University Press.
- Hume, D. 1985. 'Of the Populousness of Ancient Nations'. In *Hume's Essays*, edited by E. F. Miller. Indianapolis: Liberty Classics.
- Hume, D. 1998. *An Enquiry Concerning the Principles of Morals*, edited by T. L. Beauchamp. Oxford: Oxford University Press.

- Incurvati, L., and J. Schlöder. 2023. *Reasoning with Attitude: Foundations and Applications of Inferential Expressivism*. Oxford: Oxford University Press.
- Ismael, J. 2007. *The Situated Self*. New York: Oxford University Press.
- Ismael, J. 2016a. ‘How Do Causes Depend on Us? The Many Faces of Perspectivalism’. *Synthese* 193: pp. 245–67. <https://doi.org/10.1007/s11229-015-0757-6>
- Ismael, J. 2016b. *How Physics Makes Us Free*. New York: Oxford University Press.
- Jackson, F. 1984. ‘Two Theories of Indicative Conditionals: Reply to Brian Ellis’. *Australasian Journal of Philosophy* 62: pp. 67–76.
- Jackson, F. 1987. *Conditionals*. Oxford: Basil Blackwell.
- Jackson, F. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Jackson, F., G. Oppy, and M. Smith. 1994. ‘Minimalism and Truth Aptness’. *Mind* 113: pp. 287–312.
- Johnston, M. 1993. ‘Objectivity Refigured: Pragmatism Without Verificationism’. In *Reality, Representation, and Projection*, edited by J. Haldane and C. Wright, pp. 85–130. New York: Oxford University Press.
- Joyce, J. M. 2002. ‘Levi on Causal Decision Theory and the Possibility of Predicting One’s Own Actions’. *Philosophical Studies* 110: pp. 69–102.
- Kaplan, A. 1963. ‘Logical Empiricism and Value Judgments’. In *The Philosophy of Rudolf Carnap*, edited by P. Schilpp, pp. 827–58. LaSalle, IL: Open Court.
- Kaspers, T. 2023. ‘Truth and Its Uses: Deflationism and Alethic Pluralism’. *Synthese* 202: 130.
- Kölbel, M. 2004. ‘Faultless Disagreement’. *Proceedings of the Aristotelian Society* 104: pp. 53–73.
- Kraut, R. 1990. ‘Varieties of Pragmatism’. *Mind* 99: pp. 157–83.
- Kripke, S. 1982. *Wittgenstein: On Rules and Private Language*. Cambridge, MA: Harvard University Press.
- Leibniz, G. W. 1951. *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*, edited by A. Farrer. London: Routledge & Kegan Paul.
- Levi, I. 1986. *Hard Choices*. New York: Cambridge University Press.
- Levi, I. 2000. ‘Review Essay: The Foundations of Causal Decision Theory’. *Journal of Philosophy* 97: pp. 387–402.
- Lewis, D. 1972. ‘General Semantics’. In *Semantics of Natural Language*, edited by D. Davidson and G. Harman, pp. 169–218. Dordrecht: Reidel. Originally published in *Synthese* 22 (1970), pp. 18–67.
- Lewis, D. 1976. ‘Probabilities of Conditionals and Conditional Probabilities’. *Philosophical Review* 85: pp. 297–315.

- Lewis, D. 1979. 'Attitudes De Dicto and De Se'. *The Philosophical Review* 88: pp. 513–43. Reprinted in his *Philosophical Papers I*, pp. 133–55. New York: Oxford University Press.
- Lewis, D. 1986 [1980]. 'A Subjectivist's Guide to Objective Chance'. In *Philosophical Papers, Volume II*, pp. 83–132. Oxford: Oxford University Press. Originally published in *Studies in Inductive Logic and Probability, Vol. II*, edited by R. C. Jeffrey. Berkeley: University of California Press, 1980.
- Lewis, D. 1988. 'Desire as Belief'. *Mind* 97: pp. 323–32.
- Lewis, D. 1994. 'Humean Supervenience Debugged'. *Mind* 103: pp. 473–90.
- Liu, Y., and H. Price. 2019. 'Heart of DARCness'. *Australasian Journal of Philosophy* 97: pp. 136–50.
- Liu, Y., and H. Price. 2020. 'Ramsey and Joyce on Deliberation and Prediction'. *Synthese* 197: pp. 4365–86. <https://doi.org/10.1007/s11229-018-01926-8>
- Macarthur, D., and H. Price. 2007. 'Pragmatism, Quasi-Realism and the Global Challenge'. In *The New Pragmatists*, edited by C. Misak, pp. 91–120. New York: Oxford University Press. Reprinted in Price (2011c).
- Macbeath, M. 1983. 'Communication and Time Reversal'. *Synthese* 56: pp. 27–46.
- MacFarlane, J. 2003. 'Future Contingents and Relative Truth'. *Philosophical Quarterly* 53 (212): pp. 321–36. <http://dx.doi.org/10.1111/1467-9213.00315>
- MacFarlane, J. 2005a. 'The Assessment Sensitivity of Knowledge Attributions'. *Oxford Studies in Epistemology* 1: pp. 197–234.
- MacFarlane, J. 2005b. 'Making Sense of Relative Truth'. *Proceedings of the Aristotelian Society* 105: pp. 321–39.
- MacFarlane, J. 2007. 'Relativism and Disagreement'. *Philosophical Studies* 132: pp. 17–31.
- MacFarlane, J. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.
- Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
- Marion, M. 2012. 'Wittgenstein, Ramsey and British Pragmatism'. *European Journal of Pragmatism and American Philosophy* 4(2): pp. 54–80. <https://doi.org/10.4000/ejap.720>
- Martin, C. 2003. 'Review of Williams, Bernard, *Truth and Truthfulness: An Essay in Genealogy*, Princeton University Press, 2002'. *Notre Dame Philosophical Reviews*, 15.09.2003. <https://ndpr.nd.edu/reviews/truth-and-truthfulness-an-essay-in-genealogy/>
- McDowell, J. 1978. 'Are Moral Requirements Hypothetical Imperatives?' *Proceedings of the Aristotelian Society, Supplementary Volume* 52: pp. 13–29.
- McDowell, J. 1981. 'Anti-realism and the Epistemology of Understanding'. In *Meaning and Understanding*, edited by J. Bouveresse and H. Parret, pp. 225–48. Berlin: de Gruyter.
- McDowell, J. 1994. *Mind and World*. Cambridge, MA: Harvard University Press.
- McDowell, J. 1998. 'Values and Secondary Qualities'. In *Mind, Value, and Reality*, pp. 131–50. Cambridge, MA: Harvard University Press.

- McGinn, C. 1977. 'Semantics for Nonindicative Sentences'. *Philosophical Studies* 32: pp. 301–11.
- McGinn, C. 1982. 'The Structure of Content'. In *Thought and Object*, edited by A. Woodfield, pp. 207–58. Oxford: Oxford University Press.
- McGinn, C. 1984. *Wittgenstein on Meaning*. Oxford: Basil Blackwell.
- McGrath, S. 2014. 'Relax? Don't Do It! Why Moral Realism Won't Come Cheap?'. *Oxford Studies in Metaethics* 9: pp. 186–214.
- Mellor, D. H. 1971. *The Matter of Chance*. Cambridge: Cambridge University Press.
- Mellor, D. H. 1988. 'How Much of the Mind Is a Computer?' In *Computers Brains and Minds: Papers in Cognitive Science*, edited by R. Albury and P. Slezak. Dordrecht: Kluwer Academic Publishers.
- Mellor, D. H. 1991. 'The Warrant of Induction?'. In *Matters of Metaphysics*, Cambridge: Cambridge University Press.  
<https://www.repository.cam.ac.uk/bitstream/handle/1810/3475/InauguralText.html?sequence=5&isAllowed=y>
- Menzies, P. 1999. 'Intrinsic versus Extrinsic Conceptions of Causation.' In *Laws and Causation: Australasian Studies in the History and Philosophy of Science*, edited by H. Sankey, pp. 313–29. Dordrecht: Kluwer.
- Menzies, P., and H. Price. 1993. 'Causation as a Secondary Quality'. *British Journal for the Philosophy of Science* 44: pp. 187–203.
- Menzies, P., and H. Price. 2009. 'Is Semantics in the Plan?' In *Conceptual Analysis and Philosophical Naturalism*, edited by D. Braddon-Mitchell and R. Nola, pp. 183–200. Cambridge, MA: MIT Press.
- Miller, E. F., ed. 1985. *Hume's Essays*. Indianapolis: Liberty Classics.
- Millikan, R. 1989. 'Biosemantics'. *Journal of Philosophy* 86: pp. 281–97.
- Millikan, R. 2009. 'Biosemantics'. In *Oxford Handbook of Philosophy of Mind*, edited by B. McLaughlin, A. Beckermann, and S. Walter, pp. 394–406. Oxford: Oxford University Press.
- Misak, C. 2017. 'Ramsey's 1929 Pragmatism'. In *The Practical Turn: Pragmatism in the British Long Twentieth Century*, edited by C. Misak and H. Price, pp. 11–28. Oxford: Oxford University Press.
- Misak, C. 2020. *Frank Ramsey: A Sheer Excess of Powers*. Oxford: Oxford University Press.
- Misak, C., and H. Price, eds. 2017. *The Practical Turn: Pragmatism in the British Long Twentieth Century*, Proceedings of the British Academy 210. Oxford: Oxford University Press.
- Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Nozick, R. 1969. 'Newcomb's Problem and Two Principles of Choice'. In *Essays in Honor of Carl G. Hempel*, edited by N. Rescher, pp. 114–46. Dordrecht: Reidel.
- O'shea, J. 2007. *Wilfrid Sellars: Naturalism with a Normative Turn*. Cambridge: Polity Press.

- Olszewsky, T. M., ed. 1969. *Problems in the Philosophy of Language*. New York: Holt, Rinehart & Winston.
- Palmer, F. R. 1986. *Mood and Modality*. Cambridge: Cambridge University Press.
- Papineau, D. 1996. 'Many Minds are No Worse than One'. *British Journal for the Philosophy of Science* 4: pp. 233–41.
- Parkinson, G., ed. 1982. *Marx and Marxisms*. Cambridge: Cambridge University Press.
- Parfit, D. 2011. *On What Matters, Volume 1*. Oxford: Oxford University Press.
- Passmore, J. 1957. *A Hundred Years of Philosophy*. London: Duckworth. Page references are to the Penguin Books edition, 1968.
- Pearl, J. 2000. *Causality: Models, Reasoning and Inference*. New York: Cambridge University Press.
- Perry, J. 1979. 'The Problem of the Essential Indexical'. *Noûs* 13: pp. 3–21.
- Pettit, P. 1982. 'Habermas on 'Truth and Justice''. In *Marx and Marxisms*, edited by G. Parkinson, pp. 207–28. Cambridge: Cambridge University Press.
- Pettit, P. 1991. 'Realism and Response-Dependence'. *Mind* 100: pp. 587–626.
- Platts, M., ed. 1980. *Reference, Truth and Reality*. London: Routledge & Kegan Paul.
- Price, H. 1981. 'The Problem of the Single Case'. PhD diss., Cambridge University. Accessible at: <https://philarchive.org/rec/PRITPO-47>
- Price, H. 1983a. 'Does "Probably" Modify Sense?' *Australasian Journal of Philosophy* 61: pp. 396–408.
- Price, H. 1983b. 'Sense, Assertion, Dummett and Denial'. *Mind* 92: pp. 174–88.
- Price, H. 1983c. 'Could a Question be True?: Assent and the Basis of Meaning'. *The Philosophical Quarterly* 33: pp. 354–64.
- Price, H. 1983d. 'The Use of Force in a Theory of Meaning'. Retrieved September 19, 2023 from: <https://philarchive.org/rec/PRITUO-3>.
- Price, H. 1986. 'Conditional Credence'. *Mind* 95: pp. 18–36.
- Price, H. 1987. 'Truth and the Nature of Assertion'. *Mind* 96: pp. 202–20.
- Price, H. 1989. 'Defending Desire-as-Belief'. *Mind* 98: pp. 119–27.
- Price, H. 1990. 'Why "Not"?' *Mind* 99: pp. 221–38.
- Price, H. 1991a. 'Agency and Probabilistic Causality'. *British Journal for the Philosophy of Science* 42: pp. 157–76.
- Price, H. 1991b. 'Two Paths to Pragmatism'. In *Response-Dependent Concepts*, edited by P. Menzies, pp. 46–82. Canberra: Philosophy Program, RSSH, ANU. Reprinted in Price (2011c), pp. 80–111.
- Price, H. 1992a. 'Metaphysical Pluralism'. *Journal of Philosophy* 89: pp. 387–409. Reprinted in Price (2011c), pp. 34–53.

- Price, H. 1992b. ‘Agency and Causal Asymmetry’. *Mind* 101: pp. 501–20.
- Price, H. 1992c. ‘The Direction of Causation: Ramsey’s Ultimate Contingency’. In *PSA 1992, Volume 2*, edited by D. Hull, M. Forbes, and K. Okruhlik, pp. 253–67. East Lansing, MI: Philosophy of Science Association.
- Price, H. 1994. ‘Semantic Minimalism and the Frege Point’. In *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, edited by S. L. Tsohatzidis, pp. 132–55. Routledge. Reprinted with a new postscript in *Themes From Wittgenstein*, edited by B. Garrett and K. Mulligan, pp. 15–44. Canberra: Philosophy Program, RSSS, ANU, 1993. Also reprinted in Price (2011c), pp. 54–79.
- Price, H. 1996. *Time’s Arrow and Archimedes’ Point: New Directions for the Physics of Time*. New York: Oxford University Press.
- Price, H. 1997. ‘Naturalism and the Fate of the M-Worlds’. *Proceedings of the Aristotelian Society, Supp. Vol. LXXI*: pp. 247–67. Reprinted in Price (2011c).
- Price, H. 1998a. ‘Three Norms of Assertibility, or How the MOA Became Extinct’. *Philosophical Perspectives* 12: pp. 41–54.
- Price, H. 1998b. ‘Ramsey, Reference and Reductionism’. Retrieved August 13, 2022 from: <https://philpapers.org/rec/PRIRRA-2>.
- Price, H. 1999. ‘Location, Location, Location’. Presidential Address to the Australasian Association of Philosophy, Melbourne. Available via <https://philpapers.org/go.pl?aid=PRILLL>. Retrieved November 24, 2022.
- Price, H. 2001. ‘Causation in the Special Sciences: The Case for Pragmatism’. In *Stochastic Causality*, edited by D. Costantini, M. C. Galavotti, and P. Suppes, pp. 103–20. CSLI Publications.
- Price, H. 2003. ‘Truth as Convenient Friction’. *Journal of Philosophy* 100: pp. 167–90. Reprinted in Price (2011c).
- Price, H. 2004a. ‘Immodesty Without Mirrors—Making Sense of Wittgenstein’s Linguistic Pluralism.’ In *Wittgenstein’s Lasting Significance*, edited by M. Kölbel and B. Weiss, pp. 179–205. Boston: Routledge & Kegan Paul. Reprinted in Price (2011c), pp. 200–27.
- Price, H. 2004b. ‘Models and Modals’. In *Laws and Models in Science*, edited by D. Gillies, pp. 49–69. London: King’s College Publications.
- Price, H. 2004c. ‘Naturalism Without Representationalism’. In *Naturalism in Question*, edited by D. Macarthur and M. de Caro, pp. 71–88. Cambridge, MA: Harvard University Press. Reprinted in Price (2011c).
- Price, H. 2007. ‘Causal Perspectivalism’. In *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, edited by H. Price and R. Corry, pp. 250–92. Oxford: Oxford University Press.
- Price, H. 2009a. ‘The Semantic Foundations of Metaphysics’. In *Minds, Worlds and Conditionals: Essays in Honour of Frank Jackson*, edited by I. Ravenscroft, pp. 111–40. Oxford: Oxford University Press. Reprinted in Price (2011c).

- Price, H. 2009b. ‘Metaphysics After Carnap: The Ghost Who Walks?’ In *Metametaphysics*, edited by D. Chalmers, R. Wasserman, and D. Manley, pp. 320–46. New York: Oxford University Press. Reprinted in Price (2011c).
- Price, H. 2010. ‘One Cheer for Representationalism?’ In *The Philosophy of Richard Rorty*, edited by R. Auxier, pp. 269–89. Open Court, Library of Living Philosophers XXXII. Reprinted in Price (2011c).
- Price, H. 2011a. ‘Expressivism for Two Voices’. In *Pragmatism, Science and Naturalism*, edited by J. Knowles and H. Rydenfelt, pp. 87–113. Zürich: Peter Lang.
- Price, H. 2011b. ‘The Flow of Time’. In *The Oxford Handbook of Time*, edited by C. Callender, pp. 276–311. Oxford: Oxford University Press.
- Price, H. 2011c. *Naturalism Without Mirrors*. New York: Oxford University Press.
- Price, H. 2012. ‘Causation, Chance, and the Rational Significance of Supernatural Evidence’. *Philosophical Review* 121: pp. 483–538.
- Price, H. 2014. ‘Where Would We Be Without Counterfactuals?’ In *New Directions in the Philosophy of Science*, edited by M. C. Galavotti, D. Dieks, W. Gonzalez, S. Hartmann, T. Uebel, and W. Weber, pp. 589–607. Springer.
- Price, H. 2015a. ‘Idling and Sidling Toward Philosophical Peace’. In *Meaning Without Representation: Essays on Truth, Expression, Normativity, and Naturalism*, edited by S. Gross, N. Tebben, and M. Williams, pp. 307–30. New York: Oxford University Press.
- Price, H. 2015b. ‘From Quasirealism to Global Expressivism — And Back Again?’ In *Passions and Projections: Themes from the Philosophy of Simon Blackburn*, edited by R. Johnson and M. Smith, pp. 134–52. Oxford: Oxford University Press.
- Price, H. 2017a. ‘Epilogue: Ramsey’s Ubiquitous Pragmatism’. In *The Practical Turn: Pragmatism in the British Long Twentieth Century*, edited by C. Misak and H. Price, pp. 149–62. Oxford: Oxford University Press.
- Price, H. 2017b. ‘Wilfrid Sellars Meets Cambridge Pragmatism’. In *Sellars and Contemporary Philosophy*, edited by D. Pereplyotchik and D. R. Barnbaum, pp. 123–40. Routledge.
- Price, H. 2017c. ‘Causation, Intervention and Agency—Woodward on Menzies and Price’. In *Making a Difference*, edited by H. Beebe, C. Hitchcock, and H. Price, pp. 73–98. Oxford: Oxford University Press.
- Price, H. 2018. ‘Carnapian Voluntarism and Global Expressivism: Reply to Carus’. *The Monist* 101: pp. 468–74.
- Price, H. 2019. ‘Global Expressivism by the Method of Differences’. *Royal Institute of Philosophy Supplements* 86(2019): pp. 133–54.
- Price, H. 2022a. ‘Cheryl Misak, Frank Ramsey: A Sheer Excess of Powers’. *Society* 59: pp. 52–55. <https://doi.org/10.1007/s12115-022-00675-2>
- Price, H. 2022b. ‘Family Feuds? Relativism, Expressivism, and Disagreements About Disagreement’. *Philosophical Topics*, Spring 2022.

- Price, H. 2022c. ‘Causation and the Open Future; Russell, Eddington and Ramsey’. Presentation at a Blue Sky Thinking Workshop, University of Surrey, 7 July 2022. Accessed 25 August 2022 at: <https://www.youtube.com/watch?v=eJhvF9MUcis>
- Price, H. 2022d. ‘Global Expressivism and Alethic Pluralism’. *Synthese* 200: 386. <https://doi.org/10.1007/s11229-022-03874-w>
- Price, H. 2023a. ‘Gibbard on Quasi-Realism and Global Expressivism’. *Topoi*. [doi.org/10.1007/s11245-022-09873-3](https://doi.org/10.1007/s11245-022-09873-3)
- Price, H. 2023b. ‘Time for Pragmatism’. In *Neopragmatism: Interventions in First-Order Philosophy*, edited by J. Gert, pp. 23–69. New York: Oxford University Press.
- Price, H. 2024 (forthcoming). ‘Review of José Zalabardo, Pragmatist Semantics’. *Mind*.
- Price, H. Forthcoming. ‘From Non-Cognitivism to Global Expressivism: Carnap’s Unfinished Journey?’ In *Carnap Handbuch*, edited by C. Dambock and G. Schiemer. Stuttgart: Metzler Verlag.
- Price, H., with S. Blackburn, R. Brandom, P. Horwich, and M. Williams. 2013. *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.
- Price, H., and R. Corry, eds. 2007a. *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*. Oxford: Oxford University Press.
- Price, H., and R. Corry. 2007b. ‘A Case for Causal Republicanism?’ In *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, edited by H. Price and R. Corry, pp. 1–10. Oxford: Oxford University Press.
- Price, H., and Y. Liu. 2018. ‘“Click!” Bait for Causalists’. In *Newcomb’s Problem*, edited by A. Ahmed, pp. 160–79. Cambridge: Cambridge University Press.
- Price, H., and B. Weslake. 2010. ‘The Time-Asymmetry of Causation’. In *The Oxford Handbook of Causation*, edited by H. Beebe, C. Hitchcock, and P. Menzies, pp. 414–43. Oxford: Oxford University Press.
- Price, H., and K. Wharton. 2015. ‘Disentangling the Quantum World’. *Entropy* 17: pp. 7752–67.
- Putnam, H. 1978. *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.
- Putnam, H. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Putnam, H. 2004. *Ethics Without Ontology*. Cambridge, MA: Harvard University Press.
- Queloz, M. 2021a. ‘Tracing Concepts to Needs’. *The Philosopher* 109: pp. 34–39.
- Queloz, M. 2021b. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*. Oxford: Oxford University Press.
- Quine, W. V. O. 1960. *Word and Object*. Cambridge, MA: MIT Press.
- Quine, W. V. O. 1970. *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice-Hall.
- Rabinowicz, W. 2002. ‘Does Practical Deliberation Crowd Out Self-Prediction?’ *Erkenntnis* 57: pp. 91–122.

- Ramsey, F. P. 1927. 'Facts and Propositions'. *Proceedings of the Aristotelian Society, Supplementary Volume 7*: pp. 153–70.
- Ramsey, F. P. 1929. 'General Propositions and Causality'. In *Foundations*, edited by D. H. Mellor, pp. 133–51. London: Routledge & Kegan Paul, 1978.
- Ramsey, F. P. 1978. *Foundations*, edited by D. H. Mellor. London: Routledge & Kegan Paul.
- Ridge, M. 2014. *Impassioned Belief*. Oxford: Oxford University Press.
- Ridge, M. 2019. 'Relaxing Realism or Deferring Debate?' *Journal of Philosophy* 116: pp. 149–73.
- Rorty, R. 1982. *Consequences of Pragmatism: Essays 1972–1980*. Brighton: Harvester.
- Rorty, R. 1995. 'Is Truth a Goal of Enquiry? Donald Davidson versus Crispin Wright'. *Philosophical Quarterly* 45: pp. 281–300. Reprinted in *Truth and Progress: Philosophical Papers Vol. 3*, pp. 19–42. Cambridge: Cambridge University Press, 1998.
- Rorty, R. 1999. 'Pragmatism as Anti-Authoritarianism'. *Revue Internationale de Philosophie* 53: pp. 7–20.
- Rorty, R. 2007. *Philosophy as Cultural Politics: Philosophical Papers, Volume 4*. Cambridge: Cambridge University Press.
- Rorty, R., and H. Price. 2010. 'Exchange on "Truth as Convenient Friction"'. In *Naturalism and Normativity*, edited by M. De Caro and D. Macarthur, pp. 253–62. New York: Columbia University Press.
- Rozea, M., and R. Rozea. 2021. Personal communication.  
<https://www.facebook.com/marnie.rozea/posts/10158016775756106>
- Russell, B. 1913. 'On the Notion of Cause'. *Proceedings of the Aristotelian Society, New Series* 13: pp. 1–26.
- Scanlon, T. 2014. *Being Realistic About Reasons*. New York and Oxford: Oxford University Press.
- Schaffer, J. 2015. 'Lewis on Knowledge Ascriptions'. In *A Companion to David Lewis*, edited by B. Loewer and J. Schaffer, pp. 473–90. New York: John Wiley and Sons.
- Schiffer, S. R. 1972. *Meaning*. Oxford: Oxford University Press.
- Schilpp, P., ed. 1963. *The Philosophy of Rudolf Carnap*. LaSalle, IL: Open Court.
- Schroeder, M. 2008. *Being For: Evaluating the Semantic Program of Expressivism*. New York: Oxford University Press.
- Schroeder, M. 2015. *Expressing Our Attitudes: Explanation and Expression in Ethics, Volume 2*. New York: Oxford University Press.
- Seeley, T. D. 2010. *Honeybee Democracy*. Princeton: Princeton University Press.
- Sellars, W. 1957. 'Counterfactuals, Dispositions, and the Causal Modalities'. In *Minnesota Studies in the Philosophy of Science, Vol. 2*, edited by H. Feigl, M. Scriven, and G. Maxwell, pp. 225–308. Minneapolis: University of Minnesota Press.

- Sellars, W. 1962. 'Philosophy and the Scientific Image of Man'. In *Frontiers of Science and Philosophy*, edited by R. Colodny, pp. 35–78. Pittsburgh, PA: University of Pittsburgh Press.
- Sellars, W. 1963. 'Truth and "Correspondence"'. In *Science, Perception, and Reality*, pp. 197–224. London: Routledge and Kegan Paul.
- Sellars, W. 1968. *Science and Metaphysics: Variations on Kantian Themes*. London: Routledge & Kegan Paul.
- Shapiro, L. 2014. 'Assertoric Force Perspectivalism: Relativism Without Relative Truth'. *Ergo* 1(6). DOI: <http://dx.doi.org/10.3998/ergo.12405314.0001.006>.
- Shapiro, L. 2021a. 'Truth's Dialectical Role: From Friction to Tension?'. *Inquiry*. DOI: <https://doi.org/10.1080/0020174X.2021.1972834>
- Shapiro, L. 2021b. 'Global Expressivism as Global Subjectivism'. *Philosophical Studies*. DOI: <https://doi.org/10.1007/s11098-021-01694-1>
- Simpson, M. 2018. 'Solving the Problem of Creeping Minimalism'. *Canadian Journal of Philosophy* 48: pp. 510–31. DOI: 10.1080/00455091.2017.1414525
- Smith, M. 1987. 'The Humean Theory of Motivation'. *Mind* 96: pp. 36–61.
- Smith, M. 1994. *The Moral Problem*. Oxford: Basil Blackwell.
- Snowdon, P. F. 2001. 'P. E. Strawson (1919–)'. In *A Companion to Analytic Philosophy*, edited by A. P. Martinich and D. Sosa, pp. 334–49. Malden, MA: Blackwell.
- Spirtes, P., C. Glymour, and R. Scheines. 2000. *Causation, Prediction and Search*. Cambridge, MA: MIT Press.
- Spohn, W. 1977. 'Where Luce and Krantz Do Really Generalize Savage's Decision Model'. *Erkenntnis* 11: pp. 113–34.
- Stalnaker, R. 1970. 'Probability and Conditionals'. *Philosophy of Science* 37: pp. 64–80.
- Stenius, E. 1967. 'Mood and Language-Game'. *Synthese* 17: pp. 254–74.
- Stich, S. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.
- Stich, S. 1996. *Deconstructing the Mind*. New York and Oxford: Oxford University Press.
- Stoljar, D. 1993. 'Emotivism and Truth Conditions'. *Philosophical Studies* 70: pp. 81–101.
- Strawson, P. F. 1949. 'Truth'. *Analysis* 9: pp. 83–97.
- Strawson, P. F. 1950. 'Truth'. *Proceedings of the Aristotelian Society, Supplementary Volume* 24: pp. 129–56.
- Strawson, P. F. 1985. *Skepticism and Naturalism: Some Varieties*. New York: Columbia University Press.
- Strawson, P. F., and G. Evans. 1973. 'TRUTH – A Conversation between P F Strawson and Gareth Evans'. <https://philpapers.org/rec/STRTA-7>
- Thomasson, A. Forthcoming. *Rethinking Metaphysics*. New York: Oxford University Press.

- Toulmin, S. 1950. 'Probability'. *Proceedings of the Aristotelian Society Supplementary Volume* 24: pp. 27–62.
- Toulmin, S., and K. Baier. 1952. 'On Describing'. *Mind* 61: pp. 13–38.
- Travis, C. 2000. *Unshadowed Thought*. Cambridge, MA: Harvard UP.
- Unwin, N. 1999. 'Quasi-Realism, Negation and the Frege–Geach Problem'. *The Philosophical Quarterly* 49: pp. 337–52.
- van Straaten, Z., ed. 1980. *Philosophical Subjects: Essays Presented to P. F. Strawson*. Oxford: Oxford University Press.
- von Wright, G. H. 1993. 'The Wittgenstein Papers'. In *Ludwig Wittgenstein: Philosophical Occasions, 1912–1951*, edited by J. Klagge and A. Nordmann, pp. 480–510. Indianapolis: Hackett.
- Wiggins, D. 1980. 'What Would Be a Substantial Theory of Truth?' In *Philosophical Subjects: Essays Presented to P. F. Strawson*, edited by Z. van Straaten, pp. 189–221. Oxford: Oxford University Press.
- Williams, B. A. O. 1966. 'Consistency and Realism'. *Aristotelian Society Supplementary Volume* 40: pp. 1–22. Reprinted in *Problems of Self*, pp. 187–206. Cambridge: Cambridge University Press, 1973.
- Williams, B. A. O. 1973. *Problems of the Self*. Cambridge: Cambridge University Press.
- Williams, B. 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press.
- Williams, M. 2010. 'Pragmatism, Minimalism, Expressivism'. *International Journal of Philosophical Studies* 18: pp. 317–30.
- Williams, M. 2013. 'How Pragmatists Can Be Local Expressivists'. In *Expressivism, Pragmatism and Representationalism*, edited by H. Price et al., pp. 128–44. Cambridge: Cambridge University Press.
- Wittgenstein, L. 1930. MS107, as catalogued in G. H. von Wright (1993), 'The Wittgenstein Papers', in J. Klagge and A. Nordmann (eds), *Ludwig Wittgenstein. Philosophical Occasions 1912–1951*, pp. 480–510. Indianapolis: Hackett.
- Wittgenstein, L. 1968. *Philosophical Investigations*, 3rd English edition. Oxford: Basil Blackwell.
- Woodfield, A., ed. 1982. *Thought and Object*. Oxford: Oxford University Press.
- Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Wright, C. 1989. 'Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics'. In *Reflections on Chomsky*, edited by A. George, pp. 233–64. Basil Blackwell.
- Wright, C. 1992. *Truth and Objectivity*. Cambridge, MA: Harvard University Press.
- Wright, C. 1993. 'Realism: The Contemporary Debate—W(h)ither Now?' In *Reality, Representation, and Projection*, edited by J. Haldane and C. Wright, pp. 63–84. New York: Oxford University Press.

- Wright, C. 1998. 'Truth: A Traditional Debate Reviewed'. *Canadian Journal of Philosophy* 28, Supplementary Volume 24: Pragmatism, pp. 31–74. Reprinted in *Truth*, edited by S. Blackburn and K. Simmons, pp. 203–38. Oxford: Oxford University Press.  
<https://doi.org/10.1080/00455091.1998.10717495>
- Wright, C. 2001. 'Minimalism, Deflationism, Pragmatism, Pluralism?'. In *The Nature of Truth: Classical and Contemporary Readings*, edited by M. Lynch, pp. 751–87. MIT Press.
- Wright, C. 2021. 'Alethic Pluralism, Deflationism, and Faultless Disagreement'. *Metaphilosophy* 52: pp. 432–48. <https://doi.org/10.1111/meta.12491>
- Yalcin, S. 2012. 'Bayesian Expressivism'. *Proceedings of the Aristotelian Society, New Series* 112: pp. 123–60.
- Zalabardo, J. 2023. *Pragmatist Semantics: A Use-Based Approach to Linguistic Representation*. Oxford: Oxford University Press.