

From the Agent's Point of View

Edgar Phillips

UCL

A Thesis submitted for the degree of Master of Philosophical Studies

September 2015

I, Edgar Phillips, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This is an essay about the role of reasons in explaining human thought and action. Three plausible-seeming ideas appear to be in tension here: that all reasons for action are facts, that we can often explain an action as rational by ascribing a false belief to its agent, and that we always explain actions as rational by identifying the agent's reasons for acting. One of the aims of this thesis is to show how we can gain a clear understanding of the first two ideas if we are willing to sacrifice the third.

I distinguish two forms of 'rationalising' action-explanation: 'worldly rationalisations', which explain an action by stating the reason for which it was done, and 'psychologised rationalisations', which explain an action by stating something about what the agent thought, or how things seemed to them. I outline how, if we take sufficiently seriously the idea of the agent's point of view, we can make sense of the way psychologised rationalisations explain actions as rational without implying that the agent acted for any reason.

This raises an important question: is psychologised rationalisation *all* we need in order to make sense of agents' behaviour as rational? That is, is the role of worldly rationalisation reducible to that of psychologised rationalisation? Early in the thesis I argue that considerations familiar from the literature on the nature of mental states suggest that this reductive approach to worldly rationalisation is not obligatory: there is no conclusive *a priori* argument against the autonomy of worldly rationalisation. In the final chapter, I present an argument which, if successful, would show that the reduction is positively undesirable. The (tentative) conclusion is that we should recognise a fundamental role for reasons themselves in explaining our thought and behaviour.

Contents

1 Introduction	8
1.1 Two ways of rationalising actions	8
1.2 Facts, explanation, reasons	12
1.3 The connection between the two forms of rationalisation . . .	16
2 Factivism: Dismissed with Prejudice?	19
2.1 Is the belief-first theory ineluctable?	19
2.2 Individualism	20
2.3 Davidson	24
3 Davidson and the Functionalist	29
3.1 Belief-first theories	29
3.2 Functionalism	31
3.3 Davidson	36
4 On Taking the Agent's Point of View	39
4.1 Common ground	39
4.2 Rationality and reasons	42
4.3 Taking the agent's point of view	46
4.4 Perspective-taking, 'simulation' and 'theory'	49
4.5 Desire and re-enactment	53
5 The Explanatory Autonomy of Reasons	56
5.1 What is at issue?	56
5.2 Hornsby	59
5.3 Williamson	60
5.4 Hyman	62
5.5 An argument for factivism	66

Acknowledgements

First, thanks to my supervisors Lucy O'Brien and M. G. F. Martin. Lucy gave invaluable support and guidance while this thesis was finding its legs, and was exceedingly generous with her time and attention. Mike's positive influence on the final shape of the thesis cannot be underestimated. His incisive questions steered me away from many errors, and his suggestions planted seeds from which many arguments grew.

I am grateful to the UCL Old Students' Association for helping to fund the year of my MPhil Stud during which this thesis was written.

Thanks also to the many teachers, friends and colleagues who have commented on parts of this work, or with whom I have discussed issues covered herein: Maria Alvarez, Sam Anderson, Julian Bacharach, Tom Blackburn, Vanessa Carr, Henry Clarke, Kathrine Cuccuru, Cressida Gaukroger, Alex Geddes, Mog Hampson, Alec Hinshelwood, Mark Kalderon, Alex Kilcran, Andrew Knox, Heejin Kwon, Clayton Littlejohn, Rory Madden, Ed Nettel, Bárbara Núñez de Cáceres, David Olbrich, Daniel Rothschild, Léa Salje, Josephine Salverda, Ashley Shaw, Harry Sheehan, Paul Snowdon, Tom Stern, and Tom Williams. I have benefitted so much from the wisdom, cleverness and generosity of the staff and students in the UCL Philosophy Department that I'm sure I've made some omissions. I want to thank especially Laura Silva and Charles Jansen, for making me feel at home in philosophy, and for their friendship and intellectual engagement over the past two years. Charles also gave very helpful written comments on parts of this thesis, as did John Hyman, and I am indebted to both.

Thanks to my parents, Bill and Hilary Phillips, for their encouragement and support. This thesis couldn't have been completed without the support also of my grandparents, Betty and Tony Phillips, who are sadly no longer here for me to thank.

Last but certainly not least, thanks to Karoline Zwierzyńska, who from my first expression of interest in philosophy hasn't wavered in her support and has wavered only occasionally in her patience, and to Alice, for keeping me company during long days at my desk, and whose habit of walking across my keyboard I can try to blame for any typographical errors.

For Karoline

Chapter 1

Introduction

1.1 Two ways of rationalising actions

People's actions can be explained in diverse ways. Sometimes we cite some salient fact about the agent's character, temperament or mood, as when we say that someone performed some generous act because they are kind, or that they lashed out because short-tempered. Sometimes we cite the agent's abilities or lack thereof (for example, '... because she's clever/foolish'), sometimes some kind of causal-historical factor that led, one way or another, to the agent's doing what they did ('... because he was tricked', '... because she hadn't had much sleep'). A particularly interesting class of explanations involve our mentioning one or other of the agent's goals, intentions or desires, as when we say that someone performed a generous act because they wanted to get on the beneficiary's good side, or that they got out the ladder because they were going to trim the hedge. Explanations of this last sort might be said to 'rationalise' the agent's behaviour in that, characteristically, they explain the agent's behaviour *as rational*, as a rational way of pursuing some further aim, and not, something done for no reason, or on a whim, or by accident.¹ But there is another (by no means unconnected), class of action-explanations that we might call 'rationalisations', and which will be of central importance to any developed account of rational or moral psychology. These are those explanations which, at least at a first pass, explain the agent's action by identifying the agent's reasons, or the rational grounds or basis on which the agent acted. It is this class of explanations which are the topic of this essay, and I will use the terms 'rationalisation' and 'rationalising explanation' to refer to explanations of this sort. I am of course following Donald Davidson in using 'rationalisation' in this sense rather than in the everyday sense of rationalising-after-the-fact, which is something we characteristically do to *irrational* behaviour. To rationalise an action after the fact, in this sense, is not to explain it; it is typically to say something that might have explained it, had

¹Michael Thompson (2008) gives an interesting discussion of explanations of this sort, referring to them as 'rationalisations'.

it been done more rationally than it was.

Among the most interesting and philosophically challenging features of rationalising explanations is that they fall into two quite different forms. In the first of these, which I will call ‘worldly rationalisations’, we explain why the agent did what they did by citing a fact that counted, or at least that seemed to the agent to count, in favour of their doing what they did.² Notably, the fact cited needn’t be about the agent. For example, we might say that Karoline bought the fungicidal wash because the wall had mould on it, adding, perhaps, that it’s best to remove any mould before repainting a wall. For this explanation to make Karoline’s action intelligible to us, we will of course need to infer (if we don’t already know) that Karoline plans to repaint the wall. Unless such is the case, the fact cited won’t be a reason for Karoline to buy the wash. It seems the fact’s explaining Karoline’s action is not, in this sort of case, independent of its being a reason for her to do what she did. There seems to be an essential connection between the fact’s normative import and its explanatory power.

We might fruitfully note a connection here with Anscombe’s discussion of intention,³ in which she marks out explanations of action that give reasons for that action as a distinctive class. In one place, this shows up in a discussion of expressions of intention as predictions of future behaviour. I can often predict my future actions by employing empirical generalisations about what I tend to do given certain conditions. For example, I might predict that I’m going to make a fool of myself at the party, since at parties I usually get drunk, and when I get drunk I usually make a fool of myself. But I might be doing something different in saying that I’m going to make a fool of myself: I might instead be expressing my intention to make a fool of myself. Perhaps I have some reason for making a fool of myself—that it will endear me to the other people in the department, say. We can think of my statement as a prediction, but it is not one that I arrive at by applying a theory of my own behaviour; I arrive at the prediction by thinking about what *to do*, and the reasons in favour of taking the various options that are open to me. The same kind of distinction holds as we think about others’ actions. As Anscombe says, it roughly establishes something as a reason if one argues with it. You might argue with my saying that I’m going to the party because it’ll be fun by pointing out that I’ll make a fool of myself (the latter being a reason not to go). You won’t argue in the same way against my claim that I’m going to make a fool of myself because I’ll get drunk (although you might tell me not to go). Roughly, I arrive at one kind of prediction through reasoning from empirical generalisations, and at the other through practical reasoning. I’m suggesting

²I will mostly write about actions in the past tense. This is mostly arbitrary. We can certainly also explain why someone *is acting*, but I don’t think this will fundamentally change the forms of explanation that apply.

³Anscombe (1963).

that a similar sort of distinction demarcates the class of explanations that I'm calling rationalisations.

With the second form of rationalisation, which I'll call 'psychologised rationalisations', we explain what the agent did by saying something about the agent—canonically, something about what the agent thought or believed. For example, we might say that Karoline bought the fungicidal wash because she thought it would kill the algae. Despite their seemingly disparate subject-matters, psychologised rationalisations and worldly rationalisations appear to be very closely connected. Both are being called 'rationalisations' for good reason. Both, it seems, explain actions as rational, as rationally intelligible, in a way that distinguishes these explanations from the other sorts of action-explanation discussed above. Yet the facts cited in worldly and psychologised rationalisations respectively connect to the actions they explain in quite different ways.

We saw that, in the case of the worldly rationalisation, the ability of what we say in explanation of Karoline's action to explain her so acting depends upon its being a reason for her to do what she does. Something similar is the case with respect to the psychologised rationalisation: if she doesn't intend to repaint the wall, then, all else being equal, her thinking that the fungicidal wash will get rid of the algae won't explain her buying it. On the other hand, the fact that Karoline thinks this stuff will get rid of the algae will not, except in quite exotic circumstances, be a reason for her to buy it. Karoline might be mistaken in her belief. Perhaps she only thinks the stuff will kill algae because she doesn't know that algae aren't fungi. If her aim is to get rid of the algae, then Karoline would be wasting her money to buy the wash, regardless of what she thinks it will do. So, whether what Karoline thinks is a reason for her to do what she does depends on whether what she believes is true. Nonetheless, learning about her belief, true or false, enables us to understand her action, and not just to understand it, but to understand it as rational.

The difference in the connection between fact cited and action explained can be brought out by considering cases where a fact about the agent's beliefs explains what the agent does in the way characteristic of worldly rationalisation, in virtue of being the reason for which the agent acted. The canonical form for giving a worldly rationalisation is 'A ϕ -ed because p ', where the fact that p was a reason for A to ϕ , and was the reason for which A ϕ -ed. Obviously, an explanation's being given in the form 'A ϕ -ed because p ' doesn't show that it's a worldly rationalisation. *All* of the examples mentioned above are given in this form, including the psychologised rationalisations. It's just as important to note that an explanation's being of the form 'A ϕ -ed because she thought that p ' doesn't guarantee that it's a psychologised rationalisation. It might for instance be more like the causal-historical explanations

mentioned earlier, as when the agent believes something they find distressing, which affects their mood and leads them to act in some self-destructive manner expressive of that mood.

More significantly for present interests, a statement of the form 'A ϕ -ed because she thought that p ' can sometimes give a *worldly* rationalisation. Compare, for example, 'Ivanushka fled Moscow because he thought he had met the Devil at the Patriarch's Ponds' with 'Ivanushka went to the psychiatric hospital because he thought he met the Devil at the Patriarch's Ponds'.⁴ In the latter case, the fact that Ivanushka thinks that he has met the Devil is a reason for him to do what he does, and it makes his action intelligible as such. In the former, if Ivanushka does indeed flee Moscow for a reason, the reason is *that he has met the Devil at Patriarch's Ponds*, not that he thinks that he has. (But if Ivanushka never actually met the Devil, then there was, contrary to how things seemed to him, no such reason; his belief may still explain his action, but not in virtue of being a reason for him to do it.)

We seem, generally speaking, to be competent at telling when an action is being explained with a worldly rationalisation, when with a psychologised rationalisation, and when in some other way. It's typically obvious from the relationship between the fact cited and the behaviour explained, although sometimes further contextual information is required to make it clear. For instance: I explain that I'm not going to climb the cliff because I think it's crumbly; you respond that it's perfectly safe; I reply that I know it's safe but can't get the thought out of my head and suspect it will make me too nervous to climb safely.⁵ Such cases are relatively unusual, and seem to present no special puzzles, because it's clear what's going on in the example. The initial statement mentions a belief that something is the case which, if it was the case, would be a reason not to climb the cliff. This suggests to us that what's being given is a psychologised rationalisation. It's then explained how the agent's *having* this belief is a reason, a worldly reason, not to climb the cliff, and it becomes clear that it is, after all, a worldly rationalisation. But to reiterate: we shouldn't be tempted to think that the cited fact's being a reason for the agent to perform the action *demonstrates* that the explanation is a worldly rationalisation. For example: Karoline bought the fungicidal wash because the wall was mouldy. Additional information: the effect of the neurotoxic spores on her brain gave her a craving for fungicidal wash.

Although the two kinds of rationalisations are perfectly easy to use and understand in the ordinary everyday practice of explaining people's behaviour, their similarities and differences raise difficult philosophical questions about the relationship between them. Acting such that one's so acting is liable to explanation by worldly rationalisation seems to be a kind of cognitive-cum-practical achievement, and plausibly a manifestation of a kind of skill or com-

⁴I believe this form of example originates in Hyman (1999).

⁵Dancy (2000, 124).

petence. It's possible to do what one has reason to do by accident; a worldly rationalisation tells us that that isn't what happened. Because of this, the preconditions for a correct worldly rationalisation are more demanding than those for psychologised rationalisation. Psychologised rationalisation is still rationalisation. In showing that the agent acted rationally, it shows that the agent manifested competence in acting as they did, if we think of rationality, or 'the faculty of reason', as a sort of competence. However, a psychologised rationalisation does not ascribe to the agent an achievement of the same sort as a worldly rationalisation does, because one's thinking that p can explain one's ϕ -ing, where that p , if it were true, would have been a reason for one to ϕ , even in cases in which it's not the case that p . A psychologised rationalisation does not, whereas a worldly rationalisation does, say that the agent did what they did because there was a reason for them to do it. And it seems natural, to me at least, to think that doing things because there are reasons to do them has something importantly to do with what it is to be a rational agent.

1.2 Facts, explanation, reasons

Certain notions, in particular *fact*, *explanation* and *reason*, have cropped up repeatedly in the discussion already. I shall say something briefly about each of them here before moving on, starting with facts.

Speaking in *very* rough and impressionistic terms, we can contrast two competing approaches to the notion of a fact. On the first, facts are concrete entities, 'truth-makers', the things in virtue of which true propositions or statements are true.⁶ On the second, facts are representational, perhaps just identical with true propositions, which might themselves be understood as 'logical constructions',⁷ or as whatever a true statement states.⁸ In this essay, I shall try for the most part to remain neutral about the nature of facts, though not because I think it irrelevant to the issues at hand. If facts are concrete, for instance, this might perhaps support the idea that there is something significant going on in (factive) worldly rationalisations that is missing from (non-factive) psychologised rationalisations.⁹ I think that it would be preferable, though, to handle the issues with which we are presently concerned on their own merits as far as possible, without relying on any substantive preconceptions about the nature of facts. There is an assumption I *will* rely on, which no reader should find objectionable: it is a fact that p if and only if it is true that p , and it is true that p if and only if p .

⁶See for example Austin (1950).

⁷Prior (1971).

⁸Strawson (1950).

⁹Compare Hossack (2007) on the connection between belief and knowledge— though see Textor (2011) for scepticism about Hossack's approach.

A substantive conception of facts would ramify through conceptions of explanation and of reasons, which are both, as I understand them, factive notions. Explanation is factive, in that it can be true that p because q only if it is true both that p and that q . If we were thinking of facts as concrete, then, we might think of explanation as a relation, and see its factivity as a special case of the principle that a relation's holding between two entities presupposes the existence of both. But we don't obviously need the concrete conception of facts to make sense of the factivity of explanation. Explaining why something is the case or why something happened is a kind of achievement, which involves making sense of things as they actually are or as they actually happened. To say something false cannot be to explain anything, any more than appealing to a reason that wasn't your reason for acting can show that you acted rationally.¹⁰

The idea that reasons are facts has been widely defended elsewhere,¹¹ and I will largely take it for granted. The account that I will give, in the fourth chapter, of how psychologised rationalisation makes sense of agents' actions, might however be taken as a defence of this view against one line of challenge: it is sometimes alleged that the factive conception of reasons creates a problem for understanding the rationalising nature of psychologised rationalising explanations. The suggestion is that we must conceive of *false* reasons in order to make sense of the plain fact that agents are sometimes rationally motivated by falsehoods. If we don't say that when an agent acts on a false belief, the psychologised rationalisation of their action gives their reason for acting, then we will, it seems, be forced to say that an agent can act rationally, manifesting their rational competence, and most likely being blameless in doing the wrong thing (if they do), and yet act for no reason. Simon Blackburn, for example, says that this

sounds harsh, for [the agent] may not have been at all irrational, after all. They certainly had their reasons for what they did, and they may have acted well in the light of them.¹²

As another example, Juan Comesaña and Matthew McGrath take the claim that whenever an agent acts rationally they act for a reason to be so obvious that they make the bold move of using it as a premise, in an argument for the view that some reasons are false.¹³ The account I later outline will, I hope, show why that claim is not nearly as irresistible as Comesaña and McGrath think it is. Often when we say someone did something for no reason we mean that they did it without thinking, or irrationally, or for no purpose. This is not what we mean when we say of the agent whose action can only be

¹⁰So long as 'acted rationally' doesn't just mean 'did what was the rational thing to do'.

¹¹See, for example, Collins (1997); Hyman (1999); Raz (2000); Williamson (2000); Alvarez (2010); Littlejohn (2012).

¹²Blackburn (2010, 8).

¹³Comesaña and McGrath (2014).

given a psychologised rationalisation that they acted for no reason. What we are saying is actually something the agent might well agree with after having discovered their error: if I tell Karoline that the Tube strike has been cancelled after she has already called a cab, she might well be annoyed that she called a cab for no reason.¹⁴

There *is* a genuine question here if we want to understand the relationship between reasons and rational motivation: How do we understand an action as rational in the absence of a reason for which it was done? Yet it is no *solution* to this problem simply to be more liberal with the word ‘reason’ and to say that we can have false reasons. Doing so merely serves to obscure a crucial distinction between those cases in which someone’s doing something can be rationalised by there having been a reason for them to do what they did, and those cases in which we need, in order to rationalise their action, to say something about what they thought was the case. When the agent is mistaken, it is not what they believe that makes their action rational. If asked what made Karoline’s calling a cab rational, ‘that there’s a Tube strike on’ isn’t the right answer if there’s no Tube strike. What makes her action rational is something to do with the role of that proposition in her psychology, and this is signalled by the fact that psychologised rationalisations are, necessarily, psychologised.

I think the significance of the need for psychologised rationalisation is well captured by saying that in these cases, seeing how the action was rational requires us to consider things from the agent’s point of view, and to see how there seemed to them to be a reason to do what they did. This is consonant with Maria Alvarez’s suggestion, that agents whose actions can only be rationalised by psychologised rationalisations act for merely *apparent* reasons.¹⁵ Apparent reasons are not reasons of a certain sort, just as a hoax Twombly is not a Twombly of a certain sort. How to cash out this suggestion will be the main focus of Chapter 4.

There is inevitably a degree of stipulation in the insistence that falsehoods cannot be reasons, and the debate about what we should and shouldn’t call ‘reasons’ is obviously, at least to an extent, terminological. I’m sure that many English-speakers would be happy to say that when Karoline calls a cab because she (falsely) thinks there’s a strike, ‘Karoline’s reason’ for calling a cab is that there’s a strike, even though this is false. I would be inclined to interpret the practice of making such statements as an elliptical way of representing the agent’s point of view, not unlike what one does when one insists that one ‘knew’ that such-and-such was the case upon discovering that it is not. But even if ‘reason’ does have a broader meaning in ordinary language, it helps, in a philosophical investigation of rationalising explanation, to be restrictive. Following the looser usage would collapse a crucial distinction,

¹⁴For a detailed discussion of this point, see Alvarez (2010, 141–7).

¹⁵Alvarez (2010).

that between, on the one hand, cases in which agents do what they do on the basis of things which were reasons *for* them to do what they did, and which thereby become reasons *why* they did those things, and on the other hand, cases in which we rationalise the agent's action in a way that is neutral about whether there was any reason for them to do what they did. This is the distinction between the two forms of rationalisation which we are calling 'worldly' and 'psychologised'. Moreover, it may tend to suggest that the way to understand the two forms of rationalisation is as attributing two different kinds of reasons to the agent, and, to reiterate, I doubt whether that will get us anywhere we want to end up.

Some, such as Mark Kalderon and Craig French, have argued that we should think that not just facts, but also objects or property-instances can be reasons.¹⁶ I think that the connections between reasons and explanation, and between reasons and inference (or 'reasoning') make it preferable to think of reasons as being propositional in form, so I would prefer not to accept objects or property-instances as reasons (assuming property-instances aren't just facts) unless there is compelling reason to do so. Kalderon and French are moved by concerns about the epistemology of perception. I will suggest, in the final chapter, that such concerns can be answered whilst maintaining that reasons are facts, only by adopting a certain approach to rationalisation.¹⁷ If this suggestion is cogent, the motivation for expanding our conception of reasons might thereby be undermined.

A final point about rationalisation concerns its scope. Although I will focus mainly on rationalising explanations of action, this is largely a matter of clarity and convenience, and it is by no means only actions to which the dual scheme of worldly and psychologised rationalisation applies, because acting is by no means the only thing agents do for reasons. Various kinds of psychological phenomena are, or can be, 'reason-responsive', including judging, deciding, desiring, having certain affective responses, and believing. That the dual scheme applies to the last of these will in fact be crucial to an argument I present in the final chapter. We typically expect someone's beliefs to be responsive to reasons, to be held for good reasons and to be revisable in the light of new evidence, and so on. As with action, we can often explain someone's believing something by reference to something other than their reasons for believing it, such as their character,¹⁸ their upbringing, that they were taught it in school, that they haven't really thought about it, and various other sorts of facts about them, their psychology, and their history. But two central, and perhaps the normal, ways of explaining why someone believes such-and-such seem to be worldly rationalisation and psychologised

¹⁶Kalderon (2011); French (ms).

¹⁷I also don't think that accepting this suggestion forces us to reject the 'object view' of perception that Kalderon and French are keen to defend, although I won't discuss this further.

¹⁸See Raz (2000, chap. 1) for interesting discussion of this point.

rationalisation: citing reasons for which they believed it or citing other things they thought were the case.

Similarly, an agent's wanting something or feeling a certain way (sad, happy, angry, ...) can also often be rationalised. Often they can't—for instance, it's doubtful whether saying I want something because it just appeals to me, or because I've taken a liking to it, gives a reason for which I want it. On the other hand, when I say that I want to go home because the crowds are making me feel uncomfortable, or that Karoline was distraught because she thought someone had stolen her phone, these look like rationalisations. I trust that, despite my focus on the action case, what I say in this essay will leave room for, and will where relevant apply equally to, the rationalisation of all of the various kinds of psychological phenomena that are capable of being rationalised.

One reason for keeping a narrow focus is that what makes some fact a reason for someone to believe something or to feel a certain way will presumably be quite different from what makes a fact a reason for them to perform some action. This is a potential complicating factor that it will be easier, for the most part, to avoid. And, since the existing literature on rationalising explanation is predominantly concerned with action, sharing this focus makes it more straightforward to engage with that literature. Finally, the rationalisation of action is a worthy subject of consideration in itself. Understanding others' actions, and understanding them as rational, is a crucial part of understanding others in general. Our reasons for doing what we do are often important to others, given the importance of cooperation and the ways in which our actions stand to affect other people. There is an ethical dimension to rational action that is not as vivid when we think about rational thought or emotion. So I think that there is particular interest in focussing on the action case. Nonetheless, the reader should bear in mind that the general points I make about the nature of rationalisation are intended to have broader significance.

1.3 The connection between the two forms of rationalisation

How, then, should we think about the connection between the two forms of rationalisation? It seems quite natural to think that psychologised rationalisations must be, in psychological–explanatory terms, more fundamental. They seem to make explicit something that is merely implicit in a worldly rationalisation; giving a psychologised rationalisation of someone's action, even when a corresponding worldly rationalisation is available,¹⁹ makes more perspicuous the distinctively *psychological* aspect of that explanation. We might

¹⁹It will hopefully be clear enough what I mean by 'corresponding' here.

be tempted by this into saying that the corresponding psychologised rationalisation, perhaps with suitable embellishment, can *exhaust* the psychological aspect of any worldly rationalisation. I will call views of this general sort *belief-first*. The basic thesis of the belief-first view is that psychologised rationalisation is prior to worldly rationalisation, in roughly the sense that to make any worldly rationalisation fully explicit would involve giving it in the corresponding psychologised form. We will consider motivations for, and versions of, the belief-first approach in later chapters.

What I will call *factivism* is the rejection of the belief-first approach. On a factivist view, there are at least some kinds of worldly rationalisation that cannot be adequately accounted for in terms of any corresponding psychologised rationalisation. The factivist *might* add that we should invert the order of explanation and understand psychologised rationalisation essentially in terms of worldly rationalisation, which is primary, but such an inversion is not compulsory. For the purposes of this essay, a factivist is anyone that rejects the belief-first thesis, and that thesis is rejected by a view on which there is no priority between the two forms of explanation. I take the core of factivism to be the idea that psychologised rationalisation on its own cannot give us the whole story of how we understand agents as rational: at least sometimes, it says, we need an irreducibly factive notion of an agent's doing something because of a reason there was for them to do it.

In the next chapter, I will consider possible arguments against the factivist view. I will argue that the considerations adduced are not compelling, and that it is therefore worth seeing how both factivist and belief-first approaches might be fleshed out.

In the third chapter, I will outline two quite different versions of the belief-first approach. The first is a form of analytic functionalism. On this view, the kind of understanding of we are given of others' behaviour by rationalisations is a kind of proto-scientific causal understanding. Beliefs, on this view, are internal states of the agent, defined in terms of their functional role, including their effects on behaviour. Psychologised rationalisation is primary on this approach because a worldly reason can only explain an agent's behaviour via internal states by which that reason is represented, which internal states are the beliefs mentioned in psychologised rationalisations.

The second version of the belief-first theory is a Davidsonian one, on which the mental is anomalous, and there is no straightforward correspondence between rationalisations and explanations of the agent's behaviour in terms of physical or neurophysiological facts. Rationalisations are strictly at the level of the psychological, and facts at this level cannot be captured in terms of the kind of causal laws that the functionalist wishes to invoke. On this picture, the kind of understanding provided by rationalising action-explanations is *essentially* rational, and doesn't depend on any knowledge of

causal generalisations about the relationships between mental states and behaviour. Rationalisations are still causal in that 'A ϕ -ed because she thought that p ' is true only if A's believing that p caused her to ϕ , but the causal laws or regularities that this particular instance of causation instantiates will not correspond to the principles in terms of which we find A's action intelligible. *These* are principles of rationality, and we understand A's ϕ -ing as rationally intelligible, as something that it makes sense for A to do given that she thinks that p .

In the fourth chapter, I will present an account of rationalisation and the kind of understanding it gives us of agents' doings. The account gives a central role to a (fairly minimal) notion of perspective-taking, or of imagining things from the agent's point of view. I will show that, using this notion, we can capture the kind of rational understanding that rationalisations provide, whilst at the same time maintaining that reasons are facts and that subjective rationality must be understood in terms of objective reasons. The account, I will argue, is consistent with either a factivist or a Davidsonian belief-first approach.

In the fifth and final chapter, I will consider possible arguments for factivism. I will tentatively reject the idea that the alleged role of knowledge in making possible worldly rationalisations in itself gives us a sound route to a factivist view. Finally, I will conclude by sketching an argument for factivism, based on considerations in the epistemology of perception.

Chapter 2

Factivism: Dismissed with Prejudice?

2.1 Is the belief-first theory ineluctable?

We have identified two different ways in which actions can be explained as rational. In worldly rationalisation, we explain someone's doing what they did by stating a reason there was for them to do it, while in psychologised rationalisation, we explain their doing what they did by mentioning something they thought, their thinking of which made their acting as they did rational, at least to some extent, even if there was no reason for them to act in that way. How do these forms of explanation relate to one another? We contrasted two approaches to this question. The belief-first approach holds that psychologised rationalisations are primary, and that the fundamental way of understanding another's action as rational is in terms of how things seem to them, or in terms of the non-factive intentional psychological states they are in. The belief-first theorist claims that the rational-explanatory content of any correct worldly rationalisation can be captured by the corresponding psychologised rationalisation, perhaps together with certain further facts. The factivist approach denies this, and says that at least some rationalisations are irreducibly factive.

Some might think that factivism should be dismissed out of hand. In this chapter, we will consider an argument to this effect, which claims that the factivist theory would unacceptably violate *a priori* restrictions on what can be genuinely psychologically explanatory. I will suggest that there are good reasons to be sceptical of this argument's success. A second line of thought, based on less heavyweight assumptions, takes the possibility of error, and the common explanatory structure across cases of acting for a reason and cases of acting on a mere belief, to favour the belief-first approach. This argument is in a sense weaker, in that it does not purport to conclusively rule out the factivist approach. However it is, as we shall see, a much stronger argument.

2.2 Individualism

A presumptive, *a priori* argument against factivism might come from a version of *individualism*. Because our concern is with forms of psychological explanation, the relevant individualist thesis should be formulated in corresponding terms. Let's say then that individualism is the thesis that the psychologically explanatory facts about an agent and their behaviour depend solely on how that agent is intrinsically or internally.¹ It would follow that a worldly rationalisation could be a psychological explanation of an action (which it clearly is) only if it works by indicating some facts about the agent that depend only on how the agent is intrinsically. The obvious candidate would be the one cited in the corresponding psychologised rationalisation—a fact which our individualist will take to depend only on the agent's intrinsic condition. If successful, this argument would simply rule out factivism as an option.

The argument will only be compelling if we have compelling reason to accept the individualist's conception of psychological explanation, which imposes restrictions that a position that rejects it need not. So we must assess the case for individualism. I will argue that this case is not strong enough to force us to reject factivism out of hand. In particular, I will argue that there is good reason to think that facts depending not just on how things are intrinsically with the agent, but also on the agent's relations to her environment and the ways things are in that environment, can be genuinely psychologically explanatory, and that the considerations tendered by the individualist do not show that the explanatory role of these conditions can be reduced to facts concerning only the agent's intrinsic state.

The individualist whose position I want to consider claims that psychological explanations are causal explanations, and that only intrinsic conditions can be causally explanatory because only intrinsic conditions are causally efficacious. They hold that the aspects of an agent's psychological condition that appear to depend upon her relations to her environment are merely 'conceptual', and not genuine causally explanatory differences. The only properties of an agent that can make a genuine causally explanatory difference to her behaviour are her intrinsic properties, so an adequate conception of psychology will take an agent's psychological condition to depend solely on her intrinsic condition.² Presumably, it's only the agent's psychological condition that will be psychologically explanatory of her behaviour. At work here

¹The connection between this 'individualism' and the perhaps more familiar individualism about mental content may not be entirely straightforward. Where the individualist should draw the line between 'inner' and 'outer' is a matter of dispute. I'll assume for the sake of argument that 'internal' means something like 'within the physical boundaries of the agent's body'. As long as the individualist restricts the psychological in such a way as to rule out factivism, where exactly the line is drawn is not important for present purposes.

²This line of thought is exemplified to varying degrees in Crane (1991); Fodor (1991); Jackson (2009).

is the thought that psychological conditions are typed in terms of the sorts of behaviour they are liable to explain. Jerry Fodor's version of the argument, on which I'll focus, says that, when we consider certain sorts of cases, we can see that any difference that the externalist claims to identify in the behaviour of agents who are intrinsic duplicates of one another will conceptually depend upon a corresponding difference in the relationally-individuated content of their psychological states. This is motivated by appeal to thought-experiments. For example, imagine that Charles is here on Earth. He's thirsty and thinks that there is tea in the next room. Carlos is a perfect intrinsic duplicate of Charles on Kepler 186f, a planet exactly like Earth except for that instead of tea it has twea, a substance indiscriminable from tea, made from a plant indiscriminable from, but biologically quite distinct from, *Camellia sinensis*. Apart from these differences, Carlos is in an environment just like Charles's. Neither Charles nor Carlos could tell tea apart from twea or vice versa.

Both Charles and Carlos go into the next room to get a drink. Generally speaking, the anti-individualist will claim that Charles and Carlos have different beliefs, and that there are therefore different explanations to be given of their actions: Charles goes into the next room because he thinks there is tea in there, while Carlos goes into the next room because he thinks there is twea in there. Fodor thinks that the only relevant difference between Charles's and Carlos's actions is that one acts with the intention of getting tea and the other with the intention of getting twea. But this is a conceptual necessity, Fodor argues, given the contents of the beliefs that produce their actions. Of course, Charles will *get* tea and Carlos will get twea, but Charles would get twea if he were in Carlos's situation and vice versa, and Fodor takes this to show that the difference is not a genuine psychological difference. So, Fodor claims, the only salient difference between the two agents' behaviour, and by extension their beliefs, is conceptual, so non-contingent, so not genuinely causally explanatory.³

Such 'twin' cases are certainly compelling. However, if psychological conditions are meant to be typed in terms of their explanatory connections with behaviour, it will do to consider other examples.⁴ Karoline is in the kitchen and is about to head out. It's raining, so she wants her raincoat. She thinks her raincoat is on the hook. She goes to the hook. If someone asked why Karoline went to the hook, a good explanation might be that she thought that her raincoat was there. 'Going to the hook' is a relational description of Karoline's behaviour: whether a movement of Karoline's counts as going to the hook depends on where the hook is in relation to her body. To paraphrase a point made by Peacocke, it is not a conceptual necessity that people with

³Fodor (1991).

⁴The following argument owes much to Stalnaker (1989, 1990); Peacocke (1993); Yablo (1997); Williamson (2000).

thoughts about hooks produce behaviour that involves relations to hooks.⁵ Karoline might not do so, for instance, if she had no idea where the hook was.

That Karoline thinks her raincoat is on the hook and wants her raincoat supports certain counterfactuals. For instance, it suggests that were the hook in the hall, Karoline would go to the hall, and if the hook were in the bedroom, Karoline would go to the bedroom. Of course, this assumes that Karoline knows where the hook is, but that is an assumption we'd typically be quite happy to make. (Even in Fodor's example we assume that both Charles and Carlos know where the next room is and how to get there. If we don't make reasonable assumptions about such things, psychological explanation can never get started.) Karoline's being in whatever intrinsic condition she is in just before she goes looking for the raincoat does not support these counterfactuals. Suppose Karoline's hook is in the hall. Karolinka is Karoline's perfect intrinsic duplicate, in an environment just like Karoline's except that the hook is in the bedroom. Karolinka will move in a similar way to Karoline, but she will not go to the hook, and her action will not be explicable merely by her thinking that her raincoat is on the hook—we will have to add that she thinks the hook is in the hall (which the individualist will take to follow from her intrinsic similarity to Karoline, who uncontroversially has that belief). There is presumably some *other* intrinsic condition, perhaps like one Karoline would have been in had the hook been in the bedroom, that would have got Karolinka to go to the hook.

There are also many different ways Karoline might get to the hook. She might walk quickly, she might walk slowly, she might limp, or hop. She might have started in the bedroom instead of the kitchen, had she been in there. In each case, the intrinsic properties of Karoline and her movement will be very different. But again, such differences are irrelevant to the applicability of the rationalising explanation: in each case, Karoline goes to the hook because she thinks her raincoat is on it. Similarly, Karoline's belief about the location of her raincoat might have come about in different ways. She might have inferred that it's on the hook because that's where she normally leaves it; she might remember leaving it there; she might have asked me where it is; and so on. Again, in each case, her thinking that her raincoat is on the hook explains her going to the hook to get it, but again, in each case, her intrinsic condition will be quite different.

The individualist response to these points will be that there is, across the relevant sets of cases, some aspect of the agent's intrinsic condition that is common, that explains her action, and that is identical with the relevant explanatory psychological condition. But is the individualist entitled to this claim? The existence of such properties might save the individualist position,

⁵Peacocke (1993, 210).

but are we compelled to save it? In questioning this, we can draw support from an earlier incarnation of Fodor himself.⁶

Here is a moderate position on the connection between rationalisation and the intrinsic which I'll assume all will be happy to accept. If Karoline thinks that her coat is on the hook, her thinking this will depend at least in part on how she is intrinsically; Karoline's thinking the coat is on the hook isn't consistent with her being just any old way intrinsically. Similarly, Karoline's going to the hook depends on events of the right sort occurring within her body. Let's assume, in broadly Davidsonian spirit, that if Karoline went to the hook because she thought her coat was on it, then the intrinsic events relevant to her action must have been caused, at least in part, by the intrinsic conditions relevant to her belief. Saying something like this might well take us a good part of the way to explaining why we think that intrinsic duplicates will, at least in the short term, behave in similar ways, in the way that 'duplicating' thought-experiments presuppose, but it doesn't obviously imply that believing is a purely intrinsic condition.

What our moderate position says is that in any instance of Karoline's going to the hook because she thinks that her raincoat is on it, there is some intrinsic realiser *B* of her believing and some intrinsic realiser *G* of her going, and *B* causes *G*. Presumably *B*'s obtaining also explains *G*'s occurring. Let's assume that it does so in virtue of *B*'s being a member of some kind *K*₁, members of which typically cause events of kind *K*₂, of which *G* is a member. To reiterate: our moderate position holds that whenever Karoline goes to the hook because she thinks that her raincoat is there, there are some intrinsic realisers and kinds of intrinsic realisers which are structured in this sort of way.

Our individualist needs a claim much stronger than anything the moderate position says. The individualist wants to say that what it is for Karoline to think that the coat is on the hook simply is for her to be a specific way intrinsically, and thus that there is *an* intrinsic condition that's common across all cases of Karoline's thinking that the coat is on the hook, and *an* intrinsic condition/event common across all cases of going to the hook, such that Karoline's being the relevant intrinsic way is in general suitable to explain her moving the relevant intrinsic way (just as her having the belief is suitable to explain her performing the action). So what the individualist wants is to say that there are some kinds *K*₁ and *K*₂ of intrinsic condition such that whenever Karoline goes to the hook because she thinks her coat is there, some *B* of *K*₁ causes some *G* of *K*₂. As the Fodor of 'Special Sciences' rightly argued, the moderate position in no way entails this latter claim.⁷ If there is some individualistic predicate that Karoline satisfies across all cases of her believing

⁶Namely the one of Fodor (1974).

⁷Fodor (1974). Davidson pre-empts this point in 'Mental Events' and especially clearly in 'The Material Mind'; see Davidson (1980, 214, 249).

that there's a coat on the hook, we have been given no reason to expect it to be anything other than, in Fodor's phrase, a 'wildly disjunctive' one, where the disjuncts are grouped together only because they pick out the possible realisers of Karoline's believing that there's a coat on the hook. The same goes with respect to her going to the hook. The individualist argument, when construed strongly enough to rule out factivism *a priori*, wants to make the unity of these common-sense psychological explanations hostage to a proposal to which the argument has given no support.⁸

We should not make common-sense psychology hostage to that proposal. The way psychologised rationalisation works shows that there *are* natural psychological predicates that apply across the right cases, namely '... thinks that the coat is on the hook' and '... goes to the hook'—and we exploit this broad applicability in our rationalisations of Karoline's action. Thanks to this broad applicability, we have a way of understanding Karoline's behaviour that does not presuppose any detailed knowledge of the fine details of Karoline's internal state or intrinsic bodily movements. Although we may not be able to understand Karoline's behaviour in terms of its proximate neurophysiological causes, we can certainly understand her actions in terms of the (apparent) reasons for which she performed them. Of course, individualism might be—and, I take it, more usually is—considered something more like an empirical conjecture, to be supported by means other than mere thought-experiments. Despite all that I have said, such an individualism might still create trouble for factivism. But such an individualism would have to be assessed, in part, in the light of a fleshed-out account of psychological explanation.

2.3 Davidson

Individualism of the sort considered above is not the only way to motivate the belief-first approach. A belief-first theorist might instead focus on the simple fact that psychologised rationalisation is more general than worldly rationalisation: the former is applicable whenever the latter is, but not vice versa. This theorist will aim to use this observation to support the belief-first approach by using general principles about explanatory relevance, rather than a controversial conception of psychological explanation in particular.

Such a line of thought seems to be suggested, albeit briefly, by Donald Davidson in 'Actions, Reasons, and Causes':

Straight description of an intended result often explains an action better than stating that the result was intended or desired. "It will soothe your nerves" explains why I pour you a shot as efficiently

⁸Some, I suppose—such as eliminativists like Churchland (1981); Stich (1983)—will be fine with this.

as “I want to do something to soothe your nerves”, since the first in the context of explanation implies the second; but the first does better, because, if it is true, the facts will justify my choice of action. Because justifying and explaining an action so often go hand in hand, we frequently indicate the primary reason for an action by making a claim which, if true, would also verify, vindicate, or support the relevant belief or attitude of the agent. “I knew I ought to return it”, “The paper said it was going to snow”, “You stepped on *my* toes”, all, in appropriate reason-giving contexts, perform this familiar dual function.

The justifying role of a reason, given this interpretation, depends upon the explanatory role, but the converse does not hold. Your stepping on my toes neither explains nor justifies my stepping on your toes unless I believe you stepped on my toes, but the belief alone, true or false, explains my action.⁹

Before we consider whether there is a decisive argument somewhere here, some clarification is in order. Davidson’s ‘primary reason’ does not correspond to our ‘reason’: a primary reason, for Davidson, is a combination of a belief and a desire that together rationalise and cause the agent’s action. Setting aside Davidson’s inclusion of a desire in the primary reason, and putting things in our own terminology, we might take him as saying that we frequently indicate the belief on which an agent acted by making a claim about the reasons there were for them to act as they did. In other words: when we explain actions by citing worldly reasons, we are in effect explaining by indicating a corresponding psychological fact. In other words: worldly rationalisations explain actions only by implying corresponding psychologised rationalisations. However, there needn’t be any assumption here that psychologised rationalisations are, as our individualist supposed they would be, explanations in terms of intrinsic conditions of the agent.

The Davidsonian argument for the belief-first approach appeals to the generality of psychologised rationalisation. There appear to be relevant similarities between cases—in the actions of agents and the psychological precursors of these actions—that psychologised rationalisations capture but that worldly rationalisations do not. We can group together certain cases as having something significant in common, for example: (i) you stepped on my toes and I am aware of this, (ii) you stepped on my toes but my belief is ‘Gettiered’, say because I didn’t feel you stepping on my toes but a very short person who I didn’t notice stepped on my other foot at the same time, or (iii) you didn’t step on my toes and I merely think you did (perhaps because of the short person again). In any of the cases (i)–(iii), we could explain my stepping on your toes by saying that I thought you stepped on my toes. Only in

⁹Davidson (1980, 8).

(i) could we say that I stepped on your toes because you stepped on my toes.

The thought here is the simple one that: first, if the fact that p explains the agent's ϕ -ing by being a reason for which the agent ϕ -ed, then the agent's ϕ -ing in the same case can be explained merely by their believing that p ; second, their believing that p would explain their ϕ -ing even if it were not the case that p , and would do so in the same way, namely by making the action intelligible as an exercise of the agent's rationality. This point doesn't rest on any claims about intrinsic proximal causes, it's just the point that when we can give a worldly rationalisation, the agent's having a corresponding belief is, but the fact's obtaining is not, necessary for making the agent's action rationally intelligible. This is taken as evidence for the explanatory priority of the non-factive psychological facts. To assess the argument, we need to consider why it should be so taken.

Here's one way to sharpen the Davidsonian thought into an argument against factivism. The threat to the factivist account is that the psychological fact is better *proportioned* to the action being explained.¹⁰ The two kinds of rationalisation support different counterfactuals. The worldly rationalisation tells us, to a rough approximation, that, had it not been the case that p , A would not have ϕ -ed. This should be granted even by those who take psychologised rationalisations to be primary. To deny it would be to deny that our behaviour is ever sensitive to there being reasons for or against our acting one way or the other. What the psychologised rationalisation tells us is that (again roughly), had A not thought that p , A would not have ϕ -ed. If the psychologised rationalisation is true, then this holds whether the case is one in which a corresponding worldly rationalisation is available or not. Suppose we're in case (iii). Had I not thought that you stepped on my toes, I wouldn't have stepped on your toes. Equally, suppose we're in case (i). Had I not thought that you stepped on my toes, I wouldn't have stepped on your toes—your actually doing it won't get me to respond unless I take notice.

The significant point here is really that the following counterfactuals will (typically, roughly) hold both in cases where worldly rationalisation is available and in cases where only psychologised rationalisation is available: had it been the case that p but that A didn't think that p , A would not have ϕ -ed; and had it been the case that not- p but A thought that p , A would have ϕ -ed. A 's thinking that p seems to be both closer to being necessary and closer to being sufficient for A 's ϕ -ing—better correlated with A 's ϕ -ing—than does its actually being the case that p . So although it is sometimes more informative to give a worldly rationalisation of an action, if we really want to know why the agent did what they did, the fundamental thing seems to be what they thought and not what reasons there were.

Here we might employ some machinery devised by Stephen Yablo, and

¹⁰See Yablo (1992, 1997, 2003) for the idea of proportionality in explanation and its significance. See also Williamson (2000).

used in his discussion of Williamson's arguments about the explanatory relevance of knowledge. (Yablo's discussion uses the rubric of 'causal relevance', but I think we can usefully apply it to the issue of explanatory relevance, without any commitment as to whether that amounts to causal relevance.) Yablo suggests the following:

- A property P of x is explanatorily relevant to effect y iff: (a) had x lacked P , y would not have occurred; (b) P is not egregiously weak or strong.¹¹
- A property P of x is egregiously *weak* relative to effect y iff some more natural stronger property of x is better proportioned to y than P is.
- A property P of x is egregiously *strong* relative to effect y iff some as natural weaker property of x is better proportioned to y than P is.
- Supposing $Q-$ and $Q+$ are, respectively, weaker and stronger properties of x , $Q-$ is *better* proportioned to y than $Q+$ iff y would still have occurred, had x possessed $Q-$ but not $Q+$.
- $Q-$ is *worse* proportioned to y than $Q+$ iff y would not have occurred, had x possessed $Q-$ but not $Q+$.¹²

Since Yablo's conditions are put in terms of properties of the agent, it will be easier here to compare the explanatory relevance of, first, the agent's believing that p , and second, the agent's being aware that p , where awareness of p is just the kind of state, whatever it is, that puts one in a position to do things because p . Is being aware that p egregiously strong relative to ϕ -ing? Clearly it is stronger than believing that p : being aware that p plausibly entails believing that p , but the converse doesn't hold.¹³ And it seems that A 's believing p is better proportioned to her ϕ -ing than is her being aware that p : had she thought that p but not been aware that p then, presumably, she would still have ϕ -ed.¹⁴ If we accept this form of argument, it seems that to avoid concluding that awareness is explanatorily 'screened off' by belief, the factivist will have to claim that awareness is more natural than belief.

To see the kind of structure the factivist theorist might be positing here, consider the following example. We're in case (i): I step on your toes because you stepped on mine. Suppose also that I hate spiders and try to kill them whenever I see them. Now consider the proposition: either I noticed that you stepped on my toes or there is a noticeable spider on your foot. This disjunction is better correlated with my stepping on your foot than is the fact that you stepped on my foot. Let's use the following assignment:

¹¹'Weak' and 'strong' in the sense of: knowing p is stronger than believing p because the former entails the latter but not vice versa.

¹²Yablo (2003, 324; apparent error corrected).

¹³If being aware that p doesn't entail believing p , there will doubtless be *some* non-factive analogue which we could plug in, such as its seeming that p .

¹⁴Though see Williamson (2000); Yablo (2003) for possible doubt about whether this is true in general.

- p : You stepped on my toes;
- q : I noticed that you stepped on my toes;
- r : there is a noticeable spider on your foot;
- ϕ -ing : stepping on your toes.

Call the case in which r and not- p case (iv). Now, in both cases (i) and (iv), the following hold:

1. q or r
2. Had it been the case that (p and not- $(q$ or $r)$), I wouldn't have ϕ -ed.
3. Had it been the case that $((q$ or $r)$ and not- p), I would have ϕ -ed.

In case (i), (2) is true simply because had I not noticed you stepping on my toes, I wouldn't have stepped on yours—the spider has nothing to do with it. (3) is true because had it not been the case that you stepped on my toes, I can't have noticed that you stepped on my toes, so the protasis implies that there is a very noticeable spider on your foot, so I would have stepped on your foot, but for entirely different reasons. Similarly, in case (iv), (2) is true because had the spider not been there, I wouldn't have stepped on your toes—the possibility of your stepping on my toes has nothing to do with it—and (3) is true simply because it *is* the case that there's a very noticeable spider on your foot and I *did* step on your toes. I take it to be obvious that the fact that (2) and (3) hold in (i) does not show that 'I stepped on your toes because either I noticed that you stepped on my toes or there was a very noticeable spider on your foot' gives a more fundamental explanation of my action than does 'I stepped on your toes because you stepped on mine'. The explanation is that the disjunction q or r is less natural than p , and there are good reasons to think as much, independent of those concerning proportionality to the outcome of my stepping on your toes.

The problem for the factivist is that believing seems a lot more natural than the kind of disjunctive condition just discussed. Short of independent reasons for thinking it less natural than awareness, or for rejecting one or other of the argument's assumptions, the Davidsonian-cum-Yablonian argument seems compelling. However, because it is not a presumptive *a priori* argument like the one the individualist sought to give, this needn't be the end of the story. For we might think that the dispute between the belief-first and factivist theorists is, in part, precisely about whether we can always think of an agent's being aware of a fact as no more natural than her believing that that fact obtains. Any positive argument for the factivist view will, therefore, potentially weigh against the Davidsonian argument. So it will be worth our while seeing whether we can articulate a compelling pro-factivist argument. If we can't, then the Davidsonian argument seems good grounds for taking the belief-first approach as the default position. In the next chapter, we will look at how that approach might be most plausibly developed.

Chapter 3

Davidson and the Functionalist

3.1 Belief-first theories

Belief-first theories of rationalising explanation take psychologised rationalisation to be the fundamental way of explaining actions as rational. To paraphrase Davidson's expression of the belief-first thesis: statement of a worldly reason often explains why an agent does something as efficiently as statement of the corresponding fact about what they think since the first in the context of explanation implies the second; but the first does better because, if it is true, the reason will justify the action. Because of the connection between acting rationally and acting for reasons, we frequently indicate the belief on the basis of which the agent acted by making a claim which, if true, would also verify, vindicate or support the relevant belief.¹ So, when we explain Karoline's repainting the wall by saying that the wall was badly weathered, we are efficiently, with a single statement, doing two jobs: (i) explaining Karoline's action (as rational) by indicating that she thought that the wall was badly weathered, and (ii) suggesting that what she did was actually the thing to do, by endorsing her belief as correct. All the rational-psychological-explanatory import of the worldly rationalisation is encapsulated in the first.

'A thinks that p ', of course, can carry the implicature that it's not, or might not, be the case that p . For this reason it can sound odd, or can even be misleading, to give a psychologised rationalisation when a worldly one is known to be available. But just as seeing a pomelo doesn't exclude seeming to see a pomelo, and knowing that pomelo is a citrus doesn't exclude thinking that pomelo is a citrus, and eating a pomelo doesn't exclude trying to eat a pomelo, buying a pomelo because pomelos are delicious doesn't exclude buying a pomelo because one thinks pomelos are delicious. It's simply that, when one *does* buy a pomelo because pomelos are delicious, it's more informative to say so than to give an explanation that restricts its claims to the facts about what one thinks.

¹Davidson (1980, 8).

The central task for a belief-first theory, then, is to give an account of psychologised rationalisation. Preferably we would also have an account of exactly what extra information is provided in a worldly rationalisation, since it is clearly not *just* that the agent's belief is true,² but the primary task will be to give an account of how ascribing to agents non-factive mental states, such as beliefs, explains their actions, since such ascriptions are claimed by the belief-first theorist to be at the very core of rationalising explanation.

In this chapter, I want to consider two general belief-first approaches, with particular focus on what they say about psychologised rationalisation of action and about how we understand actions in virtue of such explanations. A subjective notion of rationality will play an important role in each. Having some notion of rationality is crucial because rationalisations explain actions as rational, and any account of rationalising explanation must do justice to that fact. It is important that the notion of rationality at play here is a subjective one because, if we are to take psychologised rationalisation as the fundamental case, an action's being rational must depend only on how things seem to its agent, not on how things actually are.

The first form of the belief-first theory that I will outline is a functionalist one. According to this way of thinking, beliefs (along with desires) are inner states with causal potential, individuated in terms of their functional roles. These states interact to cause actions, and their potential to rationalise actions reduces to their potential to cause such actions. Rationalisations of actions are causal explanations—not just in the sense that they identify things that in fact caused the agent to act as they did, but in the sense that they subsume the action under a causal law or generalisation.

The second form of the belief-first theory is a broadly Davidsonian one. Compared with the functionalist, the Davidsonian accords rationality and rationalising explanation a more autonomous role in our common-sense understanding of people's actions. The Davidsonian understands rationalisations as causal explanations only in the weaker sense: a rationalisation identifies something that was a causal factor contributing to the agent's doing what they did, but the rationalisation does not show us how to subsume the action under a causal law. Principles of rationality do not correspond in any neat way with causal laws: the principles, constraints or conditions governing rationalising explanations have 'no echo in physical theory'.³

Each belief-first approach may be seen as a different way of developing the idea that rationalisations make actions intelligible by showing them to be rational in the subjective sense. But only for the Davidsonian is the *rationalising* character of these explanations essential to their explanatory character, and only for the Davidsonian is the understanding we gain in virtue of successful rationalisations constitutively rational. I will suggest that we have

²This point will be discussed in more detail in Chapter 5.

³Davidson (1980, 231).

reason to prefer this Davidsonian account.

3.2 Functionalism

Functionalism is, in the first instance, a theory, or a family of theories, of the nature of mental states. But since a psychologised rationalisation explains an action by ascribing a mental state of a certain sort, and since the functionalist seeks to understand mental states partly in terms of their tendency to cause certain sorts of behaviour, the functionalist approach is quite naturally expanded into a distinctive approach to rationalising explanation. Generally speaking, a functionalist account of a given type of mental state accounts for the nature of that mental state purely in terms of its 'functional role' in a psychological system—its causal/counterfactual relations to sensory 'inputs', behavioural 'outputs', and other mental states. So what it is to believe that you can buy fresh fenugreek in Tooting, for instance, is to be in a state that, among other things, tends, when combined with certain other types of state, to cause behaviour of a certain sort. Functionalism thus promises a way of understanding rationalisation as merely an instance of a more general form of explanation, in which the occurrence of an event is explained by its subsumption under a causal law or generalisation. The challenge for the functionalist will be to show how to actually assimilate rationalisation in this way. Perhaps the most sustained attempt at this is in Brian Loar's *Mind and Meaning*, on which I will focus in this section. Although I will not attempt anything like a refutation of the functionalist approach, I will raise a general, but, I think, quite deep, concern about the ability of that approach to account for the nature of rationalisation, in particular the role of rationality as a kind of constitutive ideal, and the essentially rational/normative character of the understanding that successful rationalisation provides.

There are different varieties of functionalism. On some approaches, giving functional accounts of mental states is essentially an empirical, scientific project. Exactly what import such an approach, if correct, would have for rationalising action-explanation would depend in part on the nature of the relationship between common-sense psychology and a developed functionalist cognitive science. However that turned out, though, empirical functionalism would not do what we want to do here, which is to elucidate our common-sense psychological practices of rationalising explanation, and to explain how rationalisations make sense of actions even to people with no knowledge of cognitive science. Our question is, assuming that common-sense psychology is sound, what does it 'contain'? I take this to be, when construed in the intended sense, a philosophical question. Of course, this is not to suggest that cognitive science will have nothing interesting to tell us about rational action, merely that it won't answer *this* question.

The kind of functionalism of interest to us, then, will be an *analytic* functionalism: one which takes a functionalist account to explicate the (largely implicit) content of common-sense psychology, which the functionalist takes to be a kind of ‘folk’ causal theory, much like so-called folk physics.⁴ As David Lewis puts it, ‘names of mental states derive their meaning from ... platitudes’—‘platitudes’ which are ‘common knowledge’, and which we might think of as having something like the following form:

When someone is in so-and-so combination of mental states and receives sensory stimuli of so-and-so kind, he tends with so-and-so probability to be caused thereby to go into so-and-so mental states and produce so-and-so motor responses.⁵

The idea is that common-sense psychology is essentially composed of such ‘platitudes’ (perhaps implicitly). If so, we should be able to understand rationalisation in such terms.

The functionalist’s reasons for favouring a belief-first account of rationalisation will be closely related to those of the individualist discussed in the previous chapter. Whether or not the functionalist endorses the particular *a priori* argument that we considered, a commitment to the functionalist view of common-sense psychology as explaining behaviour in terms of functionally-defined internal states would strongly favour treating worldly rationalisation as explaining an action by obliquely indicating such an internal state. The most obvious candidate for such a state, when the reason cited in the worldly rationalisation is that *p*, will be the agent’s belief that *p*, not least because of the similarity between the ways in which worldly reasons and beliefs seem to rationalise actions. Whether functionalism necessarily *commits* one to a belief-first view is not entirely straightforward, particularly given the possibility of ‘broad’ or ‘long-arm’ functionalism,⁶ which types mental states partly in environment-involving terms. Nonetheless, it is easy enough to see why a functionalist, particularly given the motivations for functionalism, would tend to favour a belief-first approach, and our task in this chapter is to articulate plausible versions of that approach, not further arguments for it.

A basic condition on the adequacy of a functionalist account of *rationalisation* is that it explain how the causal role of a belief can mirror its rationalising role, and how our understanding of rational relations between propositions and act-types, which we exercise in coming to understand actions as rational, constitutes a grasp of the functional roles of attitudes. Loar claims that this can be achieved—that ‘constraints on rationality [can] correspond to purely physical constraints on physical states.’⁷ Loar provides what he takes to be an

⁴See, for example, Armstrong (1968); Lewis (1972, 1983).

⁵Lewis (1972, 256).

⁶Jackson and Pettit (1988).

⁷Loar (1981, 23).

outline of such constraints, and I will consider his attempt as representative of the functionalist approach.

Part of Loar's account is a list of constraints of 'minimal rationality' on the co-occurrence of beliefs, meant to capture the idea that we will tend not to ascribe to agents combinations of beliefs that are inconsistent in virtue of their logical form.⁸ As Loar puts it,

The rationality constraints generate a vast network of such counterfactual relations among physical states, ultimately with the effect of describing a system of physical state types whose counterfactual interrelations mirror the relevant logical relations among beliefs and desires.⁹

Here is one example:

For physical state-types x and y to be related as the theory says the belief that $p \ \& \ q$ and the belief that $\sim p$ are counterfactually related is, in part, for it to be the case that if x were to occur then y would not occur.¹⁰

These constraints on belief are actually the only principles that Loar presents as capturing rational constraints.¹¹ He does, though, suggest that there will be further principles or generalisations about what agents tend to believe or want in given circumstances. (These differ in that they won't have the status of constraints on attitude-ascription.) On typical functionalist accounts, including Loar's, actions are caused by the interaction of beliefs and desires. The most straightforward approach, which Loar follows, is to treat action-explanation as being captured by a certain kind of instrumental 'practical syllogism'—

A desires that q ;

A believes that if p then A's doing ϕ will lead to q ;

A believes that p .¹²

—the conclusion being that A ϕ . So for instance, if we explain Karoline's going to Tooting by saying that she thinks they sell methi in Tooting, this explains her action by allowing us to infer something like:

- Karoline wants methi;¹³

⁸See Loar (1981, 71–74).

⁹Loar (1981, 23).

¹⁰Loar (1981, 23).

¹¹That is, apart from his 'M-constraints', which give constraints on meaning and are not directly relevant here.

¹²Loar (1981, 7).

¹³Or rather something like 'Karoline desires that Karoline get methi', if we follow the assumption that intentions are propositional attitudes. I don't see why we should.

- Karoline believes that if they sell methi in Tooting then her going to Tooting will lead to her getting methi;
- Karoline believes that they sell methi in Tooting.

And we understand her action as having been caused by the interaction of these states—states which tend to interact in this way, again, in virtue of their logical form. There will also be generalisations about the sorts of things that humans in general, and perhaps humans of particular sorts, typically or often want, some kind of generalisation to the effect that these agents generally have true instrumental beliefs about how to achieve the things they want, and generalisations to the effect that people typically believe things that they are in a position to perceptually ascertain.

Following John McDowell, I would suggest that Loar’s principles and generalisations are inadequate to the task of capturing the ‘structure’ of rationality, and that there is good reason to doubt whether any such functionalist principles could be adequate to that task.¹⁴ First, as McDowell argues, the ‘rationality constraints’ that Loar presents cover ‘at best a fragment of rationality in general’.¹⁵ Showing how the principle that rational agents will tend not to believe combinations of propositions that are inconsistent in virtue of their logical form could be ‘echoed’ in a functional analysis is a very, very long way from anything like a functionalist codification of the principles of rationality at play in rationalising explanation, even if we limit ourselves to the rationalisation of deductively inferred beliefs. As McDowell argues, though, if that is the case, then the functionalist theory

would not have the general normative notion of deductive consequence at its disposal; so its explanations could not exploit that notion, but could draw at most on the idea of certain transitions, and refrainings from transitions, that minds are as a matter of fact prone to.¹⁶

Although McDowell’s point is made specifically with respect to deductive inference, it would seem to apply generally. What the point presses is doubt about whether, on the functionalist picture, we can really think of beliefs and actions as ever genuinely being explained as having been formed or done *because* they were the rational thing to think or do.¹⁷

We can illustrate this concern, I think, by considering cases in which we fail in the attempt to find a rationalising explanation of an action, and so

¹⁴McDowell (1985).

¹⁵McDowell (1985, 387).

¹⁶McDowell (1985, 390).

¹⁷McDowell’s argument is inspired by Davidson, but Davidson’s argument is subtly different, and concerned with the different ways in which the ‘laws’ of rationality and of physical science are confirmed. I think both arguments are interesting and compelling, but a detailed discussion is beyond the scope of this essay. See Davidson (1980, essays 11–14), Davidson (2004, essay 8).

fail to find it intelligible as rational. Such an action *could* nonetheless be explained by being subsumed under a (perhaps causal) generalisation. The functionalist will be hard-pressed to account for the sense that, although we have to hand an explanation of why this event occurred, there is nonetheless a deep failure of understanding of a different sort. Sometimes, for instance, we explain certain actions an agent engages in by saying that the agent has obsessive-compulsive disorder (OCD). We might well think of OCD as being functionally-defined, and we might, in functionalist spirit, think of having OCD as a matter of being in a state that causes certain sorts of behaviour. If we know that Alice is counting the lamp-posts as she walks down the street because she has OCD, we understand her action as having been caused by a certain functionally-defined state; we explain her counting in the way the functionalist says we explain all actions. On the functionalist picture, this should then be an explanation on a par with any other. But, it seems, we do not treat it as such. Alice's counting the lamp-posts is irrational, and fails to make sense to us as rational. It is hard to express precisely, but there is a sense that something we might want is missing from our explanation of Alice's action, that the explanation given is one of a very different kind from the one we might have been looking for. The functionalist can of course recognise a distinction here—we're not explaining Alice's action by ascribing beliefs and desires, but by ascribing a different sort of state—but it's hard to see how this could be, for the functionalist, a deep difference, a difference with any significant import for the character of our understanding.

It might help here to recall Anscombe's discussion of the reason-requesting sense of the question 'Why?', which she says applies distinctively to intentional actions.¹⁸ Anscombe suggests that iterated application of this 'Why?'-question should, if an action is intentional, terminate in a 'desirability characterisation' of the end that the agent was pursuing in performing that action—for instance, 'it's beautiful', 'it would suit my needs', 'I find it pleasant', or even 'I just fancy it'. As I will suggest in the next chapter, we can gain the kind of understanding of an action that is distinctively provided by rationalisation when we come to see how, from the agent's point of view, the action seemed to be worth doing. Alice's compulsive counting doesn't make sense in this way, either to us or to her. It makes sense as the kind of thing that sometimes happens, but not as something for her to choose to do. It is hard to see how to make room for the significance of this thought on a functionalist theory; but if the thought is not significant, then the distinction between rationalising and non-rationalising explanations of action appears arbitrary. Again, I don't take this to constitute anything like a conclusive refutation of the functionalist approach, but it does, I think, suggest that, if we are searching for a satisfactory account of the nature of rationalising explanation, it

¹⁸Anscombe (1963).

might pay to look elsewhere. In particular, it might pay to look at a theory that affords a more fundamental role to the rational, normative character of the concepts operative in these explanations.

3.3 Davidson

The alternative belief-first approach that we will be considering is, broadly speaking, a Davidsonian one, although I won't pretend that it matches Davidson's own views in every respect. On this view, rationalising explanations are *sui generis*: a rationalisation does not explain an action by subsuming it under a causal law of the sort that appears in physical science;¹⁹ it does so simply by showing how it was rational, or how it seemed from the agent's perspective to be worth doing.²⁰ Rationality is a 'constitutive ideal' in our interpretation of one another as rational agents,²¹ and nothing plays a corresponding role in physical law. For Davidson, a rationalisation is a causal explanation in that it picks out something that was the cause of the agent's action—this is the difference between (a) having a belief that would rationalise an action and performing that action, and (b) performing that action because one has that belief²²—but the generality under which the explanation explicitly subsumes the action is not a causal law that the causing of the agent's action instantiates. There will *be* such a causal law 'covering' the action, but one needn't be able to subsume the action under this law in order to understand it as rational; so, in contrast with the functionalist account, our understanding of rationality is not construed as a folk causal theory, at least not in the same way.

The notion of interpretation plays an important role here. Very roughly,²³ its significance is that things like rationality and truth play a constitutive regulating role in our ascriptions of psychological attitudes to others. An agent's beliefs must be for the most part true, and their beliefs and actions for the most part rational, if they are to be intelligible to us as rational agents at all.²⁴ Again the notion of rationality is a subjective one: a rationalisation shows an action to be rationally intelligible by showing how it made sense from the agent's point of view. Davidson acknowledges that worldly rationalisation often gives a better explanation than the corresponding psychologised rationalisation, but he clearly maintains a belief-first approach:

Straight description of an intended result often explains an action better than stating that the result was intended or desired. "It will soothe your nerves" explains why I pour you a shot as efficiently

¹⁹Davidson (1980, essays 11–13); Davidson (2004, essay 8).

²⁰See especially Davidson (1980, essay 1).

²¹Davidson (1980, 223).

²²Davidson (1980, essay 1).

²³I am skipping over a lot of detail here. See Davidson (1991) and the essays in Davidson (2001).

²⁴Davidson (2001, essays 9 and 11).

as “I want to do something to soothe your nerves”, since the first in the context of explanation implies the second; but the first does better, because, if it is true, the facts will justify my choice of action. Because justifying and explaining an action so often go hand in hand, we frequently indicate the primary reason for an action by making a claim which, if true, would also verify, vindicate, or support the relevant belief or attitude of the agent.²⁵

For the factivist, there is a sense in which a worldly rationalisation does not necessarily merely explain the action as well as justifying it, but might rather explain it *by* justifying it, the justification being essential to the nature of the explanation, so that an action of the same type could not have been explained in the same way had the agent been mistaken about there being a reason to act as they did. Davidson’s view is belief-first, however:

The justifying role of a reason ... depends upon the explanatory role, but the converse does not hold. Your stepping on my toes neither explains nor justifies my stepping on your toes unless I believe you stepped on my toes, but the belief alone, true or false, explains my action.²⁶

One aspect of Davidson’s view which I will purposely drop is the idea that the ‘primary reason’ for which an agent acts is a pair of psychological states, namely a rationalising belief and ‘pro-attitude’.²⁷ I don’t think this claim is at all central to Davidson’s account of action-explanation, and is little more than a terminological choice, obviously meant to indicate the—in Davidson’s view—fundamental role of these psychological conditions in explaining actions as rational. They are, for Davidson, the primary reason *why* an agent acts, which is simply to say that Davidson is a belief-first theorist. If we want to reserve ‘reason for action’ for objective reasons, things which objectively speaking favour a course of action, that’s no cause for rejecting the Davidsonian approach. After all, as we can see from the above quote, Davidson clearly takes it to be facts that *justify* actions, and takes the relevant facts to be those that correspond to the contents of the agents’ beliefs, when those beliefs are true.

For the Davidsonian approach to explain the nature of rationalisations and of the kind of understanding they provide, it would ideally say something about the idea of rationality that is being invoked. If we maintain that reasons are factive, then rationalisations do not essentially make sense of actions as having been done for reasons; indeed, it seems that the notion of ‘acting for a reason’ that corresponds to worldly rationalisation will need to be understood, like the worldly rationalisation itself, as somehow constructed

²⁵Davidson (1980, 8).

²⁶Davidson (1980, 8).

²⁷See Davidson (1980, essay 1).

out of the notion of 'being done rationally', which corresponds to psychologised rationalisation. What Davidson says, at least in 'Actions, Reasons, and Causes', is suggestive:

In the light of a primary reason, an action is revealed as coherent with certain traits, long- or short-termed, characteristic or not, of the agent, and the agent is shown in his role of Rational Animal. Corresponding to the belief and attitude of a primary reason for an action, we can always construct (with a little ingenuity) the premises of a syllogism from which it follows that the action has some ... "desirability characteristic". Thus there is a certain irreducible—though somewhat anaemic—sense in which every rationalization justifies: from the agent's point of view there was, when he acted, something to be said for the action.²⁸

In the next chapter, I will present a way of thinking of rationalisations as explaining actions by showing the action to be 'justified' (in a 'somewhat anaemic', specifically a subjective rather than objective, sense) from the agent's point of view. On that view, our understanding of an action as rational from the agent's point of view rests on our grasp of reasons, our conception of what facts are reasons for which actions. Whether or not this is how Davidson would have fleshed out the idea of rationalisation, I think the picture will be consistent—coherent, even—with the Davidsonian approach. However, it will be equally coherent with factivism. My hope is thus that it will provide both a plausible picture of rationalisation and a firm enough ground on which to weigh the two approaches against one another.

²⁸Davidson (1980, 8–9).

Chapter 4

On Taking the Agent's Point of View

4.1 Common ground

In *The Quest for Reality*, Barry Stroud, discussing the fact that the contents of psychological attitudes are 'typically specified in terms which mention only circumstances that do or could hold in the nonpsychological world', presents a line of thought that I think both the factivist and the Davidsonian belief-first theorist could endorse, at least for the most part:

[W]e who inhabit the world can understand someone in that world as believing something or as perceiving something only if we can somehow connect the possession of the psychological states we attribute to the person with facts and events in the surrounding world that we take the beliefs and perceptions to be about. We understand one another to be parts of, and engaged in, a common world we all share. If we ourselves had no beliefs at all about what is happening in the environment or what another person is likely to be paying attention to, we would be in no position to attribute any beliefs or perceptions to that person at all. So it looks as if we interpreters and ascribers of beliefs and other psychological states must be engaged in the world, in the sense of taking certain nonpsychological things to be true of it, if we are ever going to attribute beliefs or perceptions to anyone.

In identifying the contents of the attitudes we ascribe, we must inevitably start with what we already know or believe, or can find out, so we have no choice but to attribute to others, at least in general, beliefs in and perceptions of the very things we ourselves take to be true or to exist in the world. We cannot make sense of someone as believing something we know to be false unless we can identify what he believes and can offer some explanation of

how he comes to get it wrong. That involves attributing to the person many other beliefs, the possession of which helps make his particular divergence intelligible. And those further beliefs will typically include many that we share. Those we do not share will, in turn, be attributed only if we can understand how a person inhabiting and reacting to the world we all live in nonetheless came to have them.¹

Rationalisations are explanations of the actions of 'a person inhabiting and reacting to the world we all live in'. Both the factivist and the Davidsonian hold that, because of this, truth is privileged in our understanding of others as rational agents. They disagree, however, on the details of that privilege, and on exactly what it means to be a rational agent reacting to the world. For the Davidsonian, truth is privileged in our interpretation of an agent as rational, because we cannot understand them as rational without taking their beliefs to be largely true.² For the factivist, on the other hand, the crucial thought is that what it is for an agent to be rational is for them to be capable of responding to normative aspects of the world, and we can understand such agents' behaviour as rational only in so far as we can understand it as a manifestation of that capacity.³ This is why explanations of deviance from the truth must somehow trace back to the agent's responding rationally to the world as it is.

This difference of view is reflected in the somewhat different positions that the factivist and the Davidsonian will afford to rationality in the explanation of action. For both, rationalisations explain actions by showing how actions are rational. For the Davidsonian, this is just a matter of showing that the agent acted on a belief the content of which cohered in a certain sort of way with the intention with which the action was carried out. Questions may arise about why the agent had this belief, and if we cannot understand the belief itself as rational then we may fail to find the action fully intelligible, for, as it were derivative reasons; but the rational light by which the belief illuminates the action is one on which the truth or falsity of that particular belief has no direct bearing. For the factivist, on the other hand, to explain an action as rational is sometimes essentially to show how it was done because the agent took there to be something counting in its favour. If the action is explained as having been done on the basis of a mere belief, it may be made intelligible as essentially a certain kind of competent failure: competent because a failure of a sort that an agent who had no competence to respond appropriately to the relevant kind of reasons could not make, and because the failure manifested this competence; a failure because it was not a manifestation of the

¹Stroud (2000, 150f.).

²See especially Davidson (2001, essays 9–11).

³Such an idea is expressed in Raz (2000, 2011) Cf. the conception of evidence and justification in Williamson (2000).

agent's doing what the competence in question is a competence to do, that is to respond appropriately to reasons.⁴ It is because of this structure that the factivist position may be identified as a 'disjunctivist' position⁵—though we will complicate this somewhat in the final chapter.

One difference between the factivist and Davidsonian positions, then, is their view of what is conveyed by a worldly rationalisation—that is, an explanation that explains an action as rational simply by stating the (objective, worldly) reason for which the action was performed, for example an explanation like 'Karoline ordered a cab because there was a Tube strike'. For the Davidsonian, giving such an explanation is always a matter of 'indicating' the belief on which the action was performed by 'making a claim which, if true, would also verify, vindicate, or support the relevant belief'.⁶ So coming to understand the action as rational involves making an inference from the explanation given to the relevant psychologised rationalisation—the explanation that shows how the action was rational by citing a belief of the agent, as in 'Karoline ordered a cab because she thought there was a Tube strike'—and it is in inferring the psychologised rationalisation that we come to see the action as rational. For the factivist, on the other hand, such an inference is not, or not always, necessary. A worldly rationalisation says that the agent acted because some fact obtained, where the fact was a reason for them to do what they did. A reason cannot explain someone's action in this way without one's being rational, in the sense of having the capacity to do things for reasons, but this is more or less trivial: one cannot do something for a reason unless one is capable of doing things for reasons.

Although the two approaches construe the structure of the relationship between worldly and psychologised rationalisations quite differently, and although only the belief-first theory privileges the latter, it is clear that an account of how we understand actions as rational through *psychologised* rationalisations will need to be central to both. On both ways of thinking, we are striving, in giving rationalisations of people's actions, to understand how their behaviour makes sense in the circumstances of the world that we share with them. When the agent's perspective on that world involves a significant enough element of idiosyncrasy, we sometimes have to explain their behaviour with reference to that idiosyncrasy. Both approaches to understanding rationalisation need to give an account of how considering that idiosyncrasy enables us to find the agent's action intelligible. There will be some important respect in which, for the Davidsonian but not the factivist, this way of understanding an action is just the same as understanding an action that is explained by a worldly rationalisation. For the Davidsonian, the

⁴So, the factivist theorist will need to appeal to some notion of fallible competence. See McDowell (2010); Miracchi (2014) for defence and application of such a notion to perception and knowledge.

⁵See Hornsby (2008); McDowell (2013); Roessler (2014).

⁶Davidson (1980, 8).

question is simply one of what it is for an action to be rational, and of how it is that we understand an action *as* rational by being told what they believe. For the factivist, it is a question of how we come to understand an action as rational when that action was *not* done for a reason, and what it is about considering what the agent believed that enables us to do this. In this chapter, I will sketch an answer that is, I think, compatible with both approaches.

4.2 Rationality and reasons

Let's begin with a well-worn but well-loved case, the generous bequest of Bernard Williams. Someone wants a gin and tonic, and believes the stuff in the bottle before him is gin, when it is in fact petrol. Williams outlines our puzzle about rationality and explanation nicely:

On the one hand, it is just very odd to say that he has a reason to drink this stuff, and natural to say that he has no reason to drink it, although he thinks that he has. On the other hand, if he does drink it, we not only have an explanation of his doing so (a reason why he did it), but we have such an explanation which is of the reason-for-action form.⁷

It seems pertinent here to distinguish two possible uses of 'rational', one in which the agent would be rational in drinking the stuff before him, and one in which he would not. In an objective sense, it would be irrational for the agent to drink the stuff before him; there is no reason for him to do it and a very good reason not to, namely that it will make him ill. In the subjective sense, on the other hand, it would be quite rational for the agent to drink the stuff, since he wants a gin and tonic and believes the stuff to be gin. From his point of view, it's the most reasonable course of action.⁸ There are corresponding objective and subjective senses of 'ought'. The agent in Williams's example objectively ought (if he wants to stay in good health) not to drink the stuff, and he subjectively ought (if he wants a gin and tonic) to drink it. When an agent acts for a good reason—when they act in such a way that their so acting can be explained with a worldly rationalisation—objective and subjective rationality coincide.⁹ The explanation shows that what the agent did

⁷Williams (1979, 18).

⁸Kolodny (2005) uses the same terminology of 'subjective' and 'objective' rationality to make the same distinction. Cf. Scanlon (2003, 12f.)'s distinction between 'substantive' and 'structural' normative claims.

It's also worth noting here, as something of an aside, that although many (such as Come-sa-Ása and McGrath) use Williams's example to illustrate subjective rationality, Williams himself appears to favour the objective sense, apparently taking rationality to go with reasons, and reasons to depend on what's true. So Williams himself seems less reluctant to say that the agent is irrational than some of the authors who have taken up his example.

⁹I am idealising somewhat here, since there may be cases in which a worldly rationalisation can be given but the action is not rational because the agent had conclusive reason to do something else. We might say that in such a case, the worldly rationalisation would not be

was a sensible thing to do, and that they did it sensibly. Psychologised rationalisations show something less: they make the action intelligible merely as subjectively rational (though they do not imply that the action was not also objectively rational). To understand how psychologised rationalisations explain actions, then, we'll want to say something about how we might understand subjective rationality.

Niko Kolodny offers a way of thinking about subjective rationality which seems to me particularly amenable to the idea, discussed above, that truth is privileged in our understanding of others. Kolodny argues that we should see the normativity of subjective rationality as, in a sense, merely apparent.¹⁰ He presents a 'transparency account' of subjective rational ought-statements,¹¹ on which statements about what an agent subjectively ought to do are in effect statements about the agent's perspective, rather than statements about what the agent actually ought to do. On this view, when we say that Bernard, given that he believes the stuff is gin, ought to drink it, what we mean is that, as things seem to Bernard, he ought to drink the stuff. Saying this is consistent with insisting that, as things actually *are*, he really shouldn't drink it. The mismatch between the 'ought'-statements is no more than the mismatch between the agent's point of view on the world and the way the world really is. Adopting Kolodny's approach, it seems that we can more or less collapse our two puzzling aspects of psychologised rationalisation—the notion of subjective rationality and the role of the agent's idiosyncratic perspective—into one. Understanding an agent's action as rational is just understanding that action as something that, from the agent's perspective, seemed the thing to do. The real question will then be how we should think about 'the agent's perspective', and what's involved in *taking* the agent's perspective, which coming to understand their action in this way would presumably require.

First, though, Kolodny's account requires some extension and refinement if it is to form the basis of a satisfactory account of psychologised rationalisation. Apart from the unexplained idea of 'the agent's point of view', to which we will return, I can see two significant limitations to Kolodny's account as it stands. First, his focus is exclusively on rational *requirements*, and so only gives an account of the subjective rational 'ought', of cases in which it seems to the agent that they have *conclusive* reason to do or not to do something. It's not at all obvious that we only ever rationally do things we *ought* to do. It seems to me very plausible that often, all that the reasons for which we act do is to make a range of possible responses appropriate or eligible, so that we can rationally choose which course of action to take. Reason needn't determine your course of action as the uniquely legitimate one in order that you

the whole story. Certainly if *we* were aware of the conclusive reason, we would want to know why the agent did not act on it.

¹⁰Kolodny (2005).

¹¹From here on, unless otherwise indicated, 'rational' will be used to mean subjectively rational.

take that course rationally, for a reason.¹² Even if this view turns out eventually to be mistaken, it seems to me better not to rule it out of our account of rationalising explanation, if we can avoid doing so.

Kolodny's account is of subjective rational requirements, and his claim is that when we say that an agent ought rationally to ϕ , we mean that, as it seems to the agent, they ought rationally to ϕ ; they believe they have conclusive reason to ϕ . So if we have a psychologised rationalisation of the form 'A ϕ -ed because she thought she had conclusive reason to', Kolodny has an account of the sense in which this explains the action as rational. How should we extend the idea so as to take into account explanations that explain the action as rationally eligible rather than rationally required? The obvious thing to do would be to say that an action is rationally eligible for an agent when the agent believes that she has sufficient reason, or a good enough reason, to ϕ . This suggestion, however, brings us to another problem with Kolodny's account, which is that it is overly intellectualised. As Kolodny presents things, the rationality of an action for an agent is determined by the agent's beliefs about what reasons they have. This is unsatisfactory when we come to apply the account to rationalising explanations of action, because in such explanations we typically advert not to such higher-order beliefs but to ordinary beliefs about how things are—the kinds of beliefs the contents of which would *be* reasons if they were true.

Perhaps it might be suggested that these explanations explain by indicating something that the agent believed to be a reason: we infer the higher-order belief, and this is the basis of our judgement that the action was rational. This seems misguided to me. It's from the reasons themselves that we both determine what to do and understand what others have done. Suppose for example that a trusted friend tells you that you have a reason to give her £50, and you believe her, but she doesn't tell you what the reason is. You believe you have a reason to give your friend £50 and a reason not to (since giving her the £50 will deprive you of the £50, and it's your money after all). What should you (rationally) do? It seems impossible to say, either from your perspective or from an observer's. The natural thing to do in such a situation would be to try and guess what the reason might be. Perhaps if your friend could tell you more about the reason—what sort of reason it was (what concerns it pertained to), how 'weighty' it was, and so on—you might be able to come to a decision, but this would hardly be any substitute for knowing what the reason *is*. By contrast, when one does know what the relevant reason is supposed to be—that is, when one thinks that p , in the case where the reason is the fact that p —one doesn't need the higher-order belief *that* it is a reason to ϕ in order to be moved to ϕ , nor (I think) to make this move a rational transition.¹³

¹²This is a dominant theme in Joseph Raz's work. See especially Raz (2000).

¹³Cf. Carroll (1895); Blackburn (1995, 2010). See also Arpaly and Schroeder (2012).

Kolodny considers a suggestion along these lines, which he credits to Pamela Hieronymi and Seana Shiffrin, with respect to his position on rational requirements. The suggestion is that

when we say ‘You ought to ϕ ; it would be irrational of you not to,’ we mean not (in general) ‘As it seems to you, you ought to ϕ ,’ but instead, ‘As it seems to you, something is so, and (although you may not have realized it) if that is so, then you ought to ϕ .’¹⁴

Kolodny says he has one misgiving about this suggestion, which is that he doesn’t think someone is irrational in failing to ϕ merely because they’re not aware that something they believe would be conclusive reason for them to ϕ if it were true. The misgiving may be onto something. Suppose that some ordinary person believes some mathematical truths which entail (by way of a complex proof) the truth of Fermat’s Last Theorem. Suppose this is enough to make those mathematical truths conclusive reasons to believe the Theorem. Being an ordinary person, our agent simply has no way of knowing that some things they believe would, if true, entail Fermat’s Last Theorem. It would be absurd to accuse this person of being irrational for failing to believe the Theorem; given that they cannot see the entailment, it would normally be irrational for them to believe it. So there does seem to be a problem with the proposed revision. On the other hand, Kolodny’s suggestion seems unpalatable as well. Where to go from here?¹⁵

In a recent paper, Kurt Sylvan proposes an attractive middle way between the two problematic accounts of subjective rationality.¹⁶ Sylvan aims to give an account of apparent reasons. Without going into all the fine details of Sylvan’s account, the key to his middle way is the idea of a ‘relevant reasons-sensitive competence’, which is ‘a competence to treat R-like considerations like objective reasons to do ϕ -like things only if they are, when true, objective reasons to do ϕ -like things’.¹⁷ Obviously there are a number of aspects of this idea that need unpacking. To simplify, though, we could think of the relevant competence as something like rational competence with the relevant subject-matter. Our agent in the example above is not irrational not to believe Fermat’s Last Theorem because, although he has conclusive reason to believe Fermat’s Last Theorem, he could not believe it *for* that reason, because he lacks the mathematical skill to run through the relevant proof. This obviously connects nicely with the explanatory role of rationality, and the idea that it’s a precondition of a reason’s explaining someone’s doing something that they had the capacity to do that thing for that reason. It also elucidates

¹⁴Kolodny (2005, fn. 47). I have changed the schematic letters to make them consistent with the rest of the text.

¹⁵For an interesting discussion of a similar dilemma about intellectualism in reason, see Lavin (2011).

¹⁶Sylvan (2015).

¹⁷Sylvan (2015, 599).

somewhat the sense in which the irrationality of failing to respond appropriately to (apparent) reasons is a failing in the agent: the agent fails to do not just something that they ought to do, but something that they were competent to do for the reasons that showed it to be something they ought to do. Adopting something like our simplified version of Sylvan's proposal, then, we can say that psychologised rationalisations show actions to be rational in the sense that the agent acted on a belief that, if true, would have constituted a reason for them to act as they did, and which they were competent to treat as such.

This approach to subjective rationality fits either a factivist or a Davidsonian belief-first account. Although it understands rationality in terms of reasons, this is not the same as understanding acting rationally in terms of acting for a reason. It gives a privileged role to what is the case (and to what ought to be done, and so on), as both the factivist and the Davidsonian do. It is a separate question whether an action's being explained by a reason is privileged as against an action's being explained by a rationalising belief.

4.3 Taking the agent's point of view

The suggestion about subjective rationality that we have in hand is one on which ascribing a belief to an agent explains their action by showing that, from their perspective, it was something that made sense for them to do. Understanding an action as subjectively rational, then, involves, in some sense, coming to see things from the agent's perspective.¹⁸ The question that immediately arises is what we mean by this. How we understand the idea of the agent's point of view, and of taking another's point of view, will partly determine how we conceive of the nature of the understanding that rationalisations provide.

Talk of the agent's 'point of view' with respect to rationalisation is, of course, somewhat metaphorical. In the sense of interest to us, it's not usually their point of view in the perceptual sense. The suggestion I want to make is that taking the agent's perspective, in the way relevant to rationalisation, and psychologised rationalisation in particular, involves imagining things being as the agent took them to be, and thinking about the imagined state of affairs more or less as the agent thought about what they took to be the actual state of affairs. 'Thinking in the same way' here will be a matter of exercising the same sorts of rational competences as the agent did. I thus want to appeal to a version of an idea familiar from the so-called 'simulation' approach to 'folk psychology', and particularly to something like Jane Heal's conception

¹⁸A similar idea to the one I will outline is presented in Roessler and Perner (2013), although it is not developed in quite the same way. In particular, Roessler and Perner see a greater disconnect than I do between the role of suppositional reasoning and the notion of simulation or co-cognition. The idea is also suggested, though not developed in detail, in Hornsby (2008).

of simulation as 'co-cognition'.¹⁹

When I am told that Karoline is calling a cab because there's a Tube strike on, I understand her action because I recognise that fact as a reason to call a cab. The explanation tells me something about *her* thinking, because her acting in this way presupposes that she was competent to respond to that fact as a reason to do that thing. Crucially, my coming to understand her through this explanation requires me to exercise a corresponding ability. If the fact stated explains Karoline's action as rational in virtue of being a reason for her to perform that action, I fail to see how it explains her action unless I too see the fact as such a reason. In this way, grasping the rationalisation of Karoline's action requires me to 'recreate' her thought, or to 'co-cognise' with her.

If we are not competent, or are not disposed, to think about the same fact in the same way, then this way of understanding the agent's doing what they did will be unavailable to us. As Davidson says, rationalisations 'make others intelligible to us only to the extent that we can recognise something like our own reasoning powers at work.'²⁰ Suppose we have a doctor, A, who knows that symptom S is a sign of disease D, which must be treated with treatment T. Suppose this is specialised knowledge to which *we* aren't privy. We ask the doctor why she's giving the patient T, and she replies that the patient is exhibiting S. Although this is a perfectly correct worldly rationalisation, it won't in itself make the doctor's action rationally intelligible to us in the normal way, since we lack the competence that the doctor manifests in doing what she does. Of course, in this context we normally wouldn't find the doctor's behaviour unusual in the first place, and would easily infer that S is a sign of some disease that T is used to treat, but we can easily imagine cases in which the rationalisation would fail more radically. In this case, we fail to grasp the rationalisation because we lack relevant medical knowledge, but competence with reasons can depend on many different things. A psychopath might not grasp a rationalisation of an action that appealed to reasons of altruism, someone from one culture might not grasp reasons depending on local convention in another, and so on. In such cases, to fully understand the person's action as rational, one must acquire the relevant competence. This is not to say that one who lacks a relevant rational competence will necessarily find such actions baffling; they simply won't be capable of the kind of understanding that rationalisations, distinctively, provide.

With this as background, the following way of accounting for our understanding of psychologised rationalisations becomes available. Grasping a worldly rationalisation requires us, in a sense to be specified, to share the agent's perspective. In the case of worldly rationalisation, you share the agent's perspective in that you both see things as they are—you are aware

¹⁹See Heal (2003).

²⁰Davidson (2004, 115).

of the same reason and both recognise it as such. What about those cases in which you don't share the agent's perspective? If A ϕ -s because she thinks that p , and you know that not- p , you can't come to understand her action by thinking as she did about the reason there was for her to ϕ , because there is no such reason. We don't want to say that you have to come to share the agent's belief. Coming to think that A ϕ -ed because p is not coming to understand A's ϕ -ing if it's not the case that p . To echo Stroud, we want to understand the agent's behaviour in the circumstances they are actually in, in the common world that we share. But we also don't want the kind of understanding provided by psychologised rationalisation to be utterly different in kind from that provided by worldly rationalisation.

We can bring these threads together by recognising a way of taking an agent's perspective that doesn't require that things seem to us as they seemed to the agent: perspective-taking achieved through imagination. To take the agent's perspective so as to grasp a psychologised rationalisation of their action, you take their perspective *in imagination*, aware that that is what you are doing. One way to think about this is that you imagine having their beliefs, but this idea needs clarifying in at least two respects. First, that there are different things we might mean when we talk about one's imagining having certain beliefs. What we don't mean is that one imagines *oneself*, possibly from a third-person perspective, having the belief in question, doing the kinds of things one would do if one had that belief. This is not a way of occupying the agent's perspective, thinking as they thought. I think what we do mean is: imagining things being as the agent believed them to be. When acting on the belief that p , one doesn't think about the fact that one believes that p . One thinks of the things one believes as simply the way things are. It is this attitude that one must imaginatively adopt to understand the behaviour of the believer, and one imaginatively adopts this attitude by imagining things to be as the agent took them to be.

The second point of clarification concerns the scope of what is imagined. To grasp a psychologised rationalisation it is not necessary to imaginatively adopt *all* of the agent's beliefs. For one thing, there might be cases in which we, as onlookers, have a better understanding than the agent does of how the proposition on the basis of which the agent acted is something that would, if true, have been a reason to act as the agent did, and so are in a better position than the agent to understand the agent's action as rational. Nomy Arpaly has argued that one can ϕ for a good reason even when one's higher-order normative beliefs imply that that reason is no reason to ϕ —she calls these cases of 'inverse akrasia'.²¹ Imagining things being as the agent took them to be in all respects would, in these cases, lessen our understanding. If we understand how the reason was a good one, we're in a better position to

²¹Arpaly (2002).

understand the agent's action than they are.

Another reason not to try to take on the agent's total perspective when attempting to understand their behaviour is again the thought that understanding someone else's behaviour, like understanding anything, involves seeing how it fits into the world as it really is. Understanding someone's behaviour as rational cannot require us to detach the agent's perspective completely from how things really are. In so far as the agent's perspective enters into our understanding of the action, it is as a way of seeing how their action was guided by rational thought, a kind of thought directed at responding appropriately to the circumstances.

4.4 Perspective-taking, 'simulation' and 'theory'

The above account draws on an idea from simulation theory—the idea of co-cognition. This is a theory that faces certain challenges. Here I want to consider just a couple of these, and show how the limited scope of the account I have given makes them less pressing than they might be for a view that made stronger claims. The first is an epistemological question, about the so-called 'inference from me to you'. The thought here is that for me to come to know something about your mind simply by thinking about the things you are thinking about, in the way that simulation theory seems to suggest we do, I need to assume that you and I are relevantly similar. Otherwise the fact that *I* would ϕ given that *p* gives me no ground for understanding *your* ϕ -ing because *p*. The second point is based on empirical research. It's argued that certain systematic errors in our predictions about what agents will do in certain circumstances are better accounted for by 'theory theory'—the view that common-sense psychology is in some meaningful sense a theory.²² The relevant cases are ones where we fail to predict that an agent's thought and action will be subject to some sort of systematic bias. For instance, in one widely-discussed study, subjects were asked to say which of a number of qualitatively identical items was of the best quality, and the subjects reportedly displayed a bias towards the rightmost item.²³ Stich and Nichols plausibly claim that 'few people who have not heard of the experiment predict that they would behave in the way that the subjects behaved', and suggest that the 'natural interpretation ... is that people's predictions ... are guided by an incomplete or inaccurate theory'.²⁴

I think that, when we recognise the proper scope of the above account of our understanding of rationalisations, these challenges can be defused. I am using the idea of co-cognition quite specifically as a way of thinking about the way we grasp rationalising explanations of action. My claims are thus

²²Stich and Nichols (1992).

²³Nisbett and Wilson (1977, 243f.).

²⁴Stich and Nichols (1992, 68).

doubly restricted: to the explanation of actions (not to prediction), and to a particular kind of explanation, rationalising explanation. The second restriction is significant because some of the influences on even rational agents' behaviour, and thus some of the factors that can enter into explanation of such behaviour, are non-rational. I will return to this point shortly. The first restriction is significant because discussions of theory theory and simulation theory, including objections to the latter, typically focus on predicting behaviour. If the idea of simulation couldn't adequately account for certain aspects of how people predict behaviour, this might constitute an objection to the account I have presented of explanation, *if explanation and prediction were symmetrical*. However, they are not, and I don't want to claim that we generally predict others' behaviour by using the same resources that I've suggested we use to grasp rationalising explanations of their behaviour. Such predictions would be pretty unreliable. That's not a problem for my account, because, again, explanation and prediction are not symmetrical. This point has long been recognised in the philosophy of science,²⁵ but the significance is overlooked if it is implicitly assumed that a single mechanism or capacity will account for our abilities both to predict and to explain what others do. Kristin Andrews²⁶ rightly challenges this assumption, arguing that an appreciation of the asymmetry between psychological prediction and psychological explanation is necessary for a proper understanding of folk psychology. As Andrews observes, if prediction and explanation are symmetrical, then 'whenever we make a correct prediction of behavior, we have at that moment an explanation of the behavior in hand.'²⁷ There are clear counterexamples to this claim. As Andrews points out, we very often use simple induction to predict how someone will act in a given situation, and such predictions may be quite reliable, but the bare fact that someone has acted a certain way in the past, or that people of a certain sort typically act in a certain way in a certain situation, doesn't generally constitute a good explanation of their doing so. Another point is that it's often very difficult indeed to predict someone's behaviour. Here is an example of Andrews':

[I]f we were walking down the street ... when a mugger confronted us, I would have very little idea of how either of us would behave. We might run, or hand over our money, or try to speak to the potential mugger, or one of us might try a self-defence move.²⁸

If you did any of these things, 'He was trying to mug me' might be a good enough rationalising explanation of your doing what you did; but knowing that someone is trying to mug you does not enable me to predict how you will react, unless perhaps I know you very well indeed.

²⁵See Salmon (1989) for an overview.

²⁶Andrews (2003).

²⁷Andrews (2003, 203).

²⁸Andrews (2003, 208).

We can draw at least two morals from Andrews' discussion that are of particular significance here. The first is that we shouldn't necessarily assume that we will be able to give a simple account of 'folk'-psychological explanation in general. I have been explicit throughout that I am interested specifically in *rationalising* explanations of action. I stressed the distinctiveness of this kind of explanation, and that there are many ways of explaining actions. I'm giving an account of one. The second moral is that we cannot assume that an account of how we understand a certain form of *explanation* of action must *eo ipso* give an account of how we *predict* actions. Considering an agent's reasons for acting can afford us a particular kind of understanding of why they did what they did. It enables us to understand their action as rational, and thus, if the action really was rational, makes possible a kind of understanding that we would not otherwise have. This is often a very good explanation, and just the kind of explanation we want. However, the fact that rationalisation can be a very good way of explaining actions does nothing to show that it's a good way of predicting them. The complexity of the normative realm and of individual characters and their circumstances means that considering someone's reasons for acting one way or another is often a very bad way to predict what they will do, as is clear if we consider cases in which there are a number of eligible options open to someone. Here, even assuming that the agent is completely rational and has all the relevant information won't tell us what they will do, because there are multiple courses of action any of which the agent could rationally choose. If we knew a great deal about the agent's character, maybe we could predict which they will plump for, but maybe not. Regardless, once the choice is made, they will have acted for a reason and that reason will explain why they did what they did.

We can see now that the issue of the 'inference from me to you' also shouldn't trouble us. It may be that to predict someone's behaviour purely by the co-cognitive method would *sometimes* work well, particularly if the agent is just like you, and it's true that it will often be problematic to assume that another person is sufficiently like you for this method to work. However, things look very different with regard to explanation. When an action is rationalised, the agent has already determined how to act, and the reason for which they did so is given in the explanation. To understand their action from such an explanation requires that you see how the reason was a reason to act as they did. The explanation *implies*, in virtue of being a rationalisation, that the agent was rational, that they were competent to respond to this kind of reason. The explanation may also clue you in to some further facts about the agent's character, values, and so on. But this is information the explanation *gives* to you, not information you must antecedently possess in order to understand the explanation. This partly explains why rationalisations of people's actions can tell us a lot about their character.

This doesn't yet speak to Nichols and Stich's objection. They might argue that the cases on which they focus are ones in which agents try to predict how someone will act based on the reasons for them to act, and the point is that there is a *systematic* mismatch between what's predicted and how the agent actually acts, and that this is better accounted for by theory theory than by simulation theory. This brings us back to the second scope restriction I mentioned: the account I have given is intended solely as an account of our understanding of rationalising explanations. But agents' behaviour can be influenced by non-rational factors. Sometimes we will need to appeal to such a non-rational factor to explain why the agent did what they did, and no rationalising explanation can be substituted. This seems to be the case with respect to the behaviour of the subjects in the studies Nichols and Stich discuss. The reason *why* the subjects selected the rightmost item as being of the best quality isn't something that was a reason *for* them so to select. As such, we can't understand their behaviour as rational. Their behaviour is, in this sense, unintelligible. The subjects' behaviour is properly explained not by a reason or belief, but by a certain sort of cognitive bias.

Stich and Nichols suggest that we make false predictions about these cases because we use a theory to predict people's behaviour, and in thinking about these cases we fail to take some relevant bit of theory into account. In a way, this is right. If we try to predict the actions of agents in these cases solely on the basis of what it would be rational for them to do, we fail to take into account a relevant factor, namely a bias that will tend to exert a non-rational influence on their behaviour.

Of course, the real question is why our own thought, when we think about things from these agents' perspective, doesn't reproduce the same bias. This is an interesting and important question,²⁹ but it is unclear whether it is one that should trouble the account I have given of rationalising explanation. The bias in question is not a bias of reasoning—the subjects don't take themselves to have a reason to pick the rightmost item. If, by contrast, we consider cases which clearly do involve biases of reasoning, such as those involving statistical reasoning heuristics, the situation will look different: those who reason with the same bias will take the behaviour of agents that exhibits the relevant biases to be quite rational.³⁰

I hope I have made clear the limited scope of the use I have made of the idea of 'simulation' or 'co-cognition' in giving an account, consistent with factivism, of how we understand rationalising explanations of action.³¹ I think this limited scope makes the account less vulnerable to certain forms of challenge than are more ambitious claims for simulation theory. In the fi-

²⁹For an extended discussion, see Short (2015).

³⁰Davies and Stone (1998).

³¹This limited scope seems to me to be largely in line with Heal's own claims on behalf of co-cognition.

nal section of this chapter I want to discuss another limit, of quite a different kind, on the role of the kind of imagination discussed above in our grasping rationalising explanations of action.

4.5 Desire and re-enactment

There is another sense in which we might need to consider things ‘from the agent’s perspective’ in order to understand their action as rational, that is not obviously accommodated by the account I have outlined, namely when something’s being a reason or apparent reason for the agent to do what they did depends on what we might call the agent’s ‘conative outlook’. I think it’s plausible that imagination will have a role to play here too, but it’s beyond the scope of this essay to explore that thought in any great detail. I’ll discuss it briefly here, primarily to set it aside.

Often, something’s being a reason for one to perform some action depends on one’s prior motivations, interests, projects, or values. Sometimes, although I disagree with someone about whether the consideration on which they acted was really a reason for them to do what they did, I can nonetheless understand their action as rational because, aware of their values, education, or whatever it is that led them to take the consideration in question to be a reason to do what they did, I understand *their* taking it to be such a reason. Rationalising explanations of action can thus be ‘psychologised’ in a different way from that already discussed, when we advert to some idiosyncratic feature of the agent’s evaluative or conative outlook, such as the agent’s taking something to be worthwhile, or wanting something, or being engaged in some project. They tell you, as it were, the light in which you need to view the consideration on which the agent acted in order to see it as a reason to do what the agent did. Call these explanations (somewhat crudely) ‘desire-citing rationalisations’. Does understanding explanations of this kind also require the exercise of imagination? If so, does it involve perspective-taking of the same kind that I’ve suggested is required to understand psychologised rationalisations? I’ll tentatively suggest that the answers to these questions are, respectively, Yes and No.

If grasping a desire-citing rationalisation involves exercising imagination, then it would seem at first blush not be the same sort of exercise of the imagination required to understand psychologised rationalisations, because to want or value something is not in general, in itself, for the world to seem a particular way to one—even though normally when one wants something it is *because* things are or seem a certain way. Consider for instance the following kind of case, in which a certain fact’s being a reason for someone to do something appears to depend upon an idiosyncrasy of the relevant kind.³²

³²The example, and its point, are inspired by Hampshire (1999).

Someone becomes attached to the work of a mediocre artist, and starts collecting their work. This person is under no illusions about the quality of the artist's work. It isn't terrible, but it's nothing special. It is fine art in two senses, but not the one that would mean it was worth seeking out if you didn't already have some attachment to it. However, our collector has formed an attachment to it, and so for him the fact that a given piece is a work of this artist is a reason to acquire that piece. Our collector doesn't want to acquire this artist's works because they seem to him to be of particular value compared to other artworks he might collect; he just likes them. To understand the collector's buying a given piece of this artist's work as rational, it seems we must take into account the peculiar attachment he has to this specific artist. Simply recognising that the work is one of the artist's will leave something out, given the banality of the artist's work; but we won't grasp the missing element by imagining the world to be some way other than it is, say the piece's being more valuable than it in fact is, since the agent is quite clear-eyed about its banality.

Perhaps all that's required here is an appreciation of the fact that sometimes people acquire reasons to do things because of things like attachment, and the recognition that this is such a case. Nonetheless, it seems to me plausible that there's a kind of understanding of an action that is only available to us if we can imagine wanting what the agent wanted, but that lacking this understanding needn't make their action rationally unintelligible to us. Suppose you too have had fancies similar to the collector's before, and can imagine falling for this artist's work in the way he has, and suppose that I, by contrast, find banality in art utterly repulsive. There might be a way in which you can understand the collector and I can't. You can, we might say, 'identify' with his attraction to the work in a way that I cannot. Both of us understand his action as having been done for a reason—I understand perfectly well the connection between his collecting the works of this artist, this piece's being a work of that artist, and his buying this piece—but there remains a deficit in my grasp of his psychology, that I might naturally express by saying something like: 'I just can't understand how you can buy this crap.' It seems to me plausible that this kind of understanding has something to do with whether one is able to imagine being motivated as the agent is: not just doing what they did, but being moved to do it, in the same way that they were, by the same considerations. In this case, understanding would depend not just on co-cognition construed as thinking about the same subject-matter, but something closer to Collingwood's idea of re-enactment.³³ Sadly, it would take us far beyond the scope of the present investigation to explore this thought further.

The same, or a very similar, kind of phenomenon can occur even when

³³See Collingwood (1946, esp. Part V). There might also be some connection here with the phenomenon of 'imaginative resistance'. See Liao and Gendler (2016) for an overview.

we agree that what the agent took to be a reason for acting as they did was no such reason at all. Apparently people used to give lungwort as a medicine to people with pulmonary infections because the leaves of the plant were thought to look like diseased lungs. It's not exactly clear whether that explanation is a rationalisation or not. It's not entirely unnatural to interpret the actions of these people as resulting from a kind of arational or irrational associative thinking, because clearly (clearly to us) the fact that a plant looks like a certain body part is no reason to think it will be of medicinal benefit to that body part. To understand their behaviour as rational (in at least the minimal sense), we need to understand why they took the one thing to be a reason for the other. Apparently the explanation is that these people ascribed to something called the Doctrine of Signatures, the idea that the physical form of a plant indicates its therapeutic value. Whether you can really understand the thought and action of these people as rational might depend on whether you can imagine finding the Doctrine of Signatures compelling. If you can't, you might find it hard to see these people as anything but fools.³⁴

The point of these diversions has been, primarily, to stress another restriction on the scope of the account of rationalising explanation that I have outlined. There are many questions it leaves unanswered, not least those concerning what it takes for something to be a reason for someone to do something, how imagination works, and what it is to believe that something is the case. I do hope, however, to have given, in outline, a plausible account, consistent with both factivist and Davidsonian approaches, of how worldly and psychologised rationalisations explain actions, and how we come to understand agents in the light of such explanations. In the next, final, chapter, we will consider what a developed factivist account might look like, and attempt to mount an argument for such a position.

³⁴It's possible that people employing something like the Doctrine of Signatures are in fact less foolish than this summary suggests: Bennett (2007) argues that the Doctrine is 'primarily a symbolic device used to transfer information, especially in preliterate societies', in other words that it is more a mnemonic device than a rule of inference. Score one for the principle of charity.

Chapter 5

The Explanatory Autonomy of Reasons

5.1 What is at issue?

We saw, in the last chapter, a way of thinking about worldly and psychologised rationalisations that was, on its face at least, consistent with both factivist and belief-first approaches. This picture took reasons to be facts, and took understanding such facts to be fundamental to our concept of rationality. Although it places such emphasis on the factive, though, we saw that the picture is quite consistent with a Davidsonian belief-first conception of rationalisation. This raises the question of what *is* at issue between the two approaches, and how the dispute might be settled.

To properly understand the debate, we should reiterate that factivism is essentially a negative position. The belief-first approach is characterised positively, and necessarily so, for the belief-first theorist makes a specific claim about the fundamental role of a certain kind of explanation across both forms of rationalisation. The belief-first theorist claims that worldly rationalisation rests upon, and can only be understood in terms of, psychologised rationalisation. If we understand factivism as simply the rejection of the belief-first thesis, then evidence against the belief-first theory thereby favours factivism. A factivist, on this construal, is anyone who claims that worldly rationalisation cannot always be understood in terms of psychologised rationalisation. If, then, we can find *a* worldly rationalisation, or a kind of worldly rationalisation, that we have good reason to think is not explicable in terms of a corresponding psychologised rationalisation, we will thereby have good reason to endorse the factivist approach. However, and for the same reason, there will then be a pressing question whether the point can be generalised such that the factivist approach gives us a distinctive way of thinking about rationalisation in general.

A factivist view will naturally be called 'disjunctivist', in that there is a

clear analogy between the factivist's denial that all rationalisation must be understood in terms of psychologised rationalisation, and the denial of the disjunctivist theorist of perception that all perceptual experience (veridical, illusory and hallucinatory) must be understood in terms of an element common between veridical experiences and certain corresponding illusory and hallucinatory ones—between the 'good case' and the 'bad case'.

However, there are different motivations for disjunctivism, which lead in turn to different kinds of disjunctivist theory. One such motivation is to defend a certain sort of view about object-perception, for instance that the experience one has when one sees a mind-independent object is partly constituted by that object. That position might be thought to be threatened by the observation that it is possible to have an hallucinatory experience subjectively indistinguishable from an experience of seeing a given object, and that one can have such an experience in the absence of any such object. The disjunctivist's response is that we cannot simply assume, because of the way we group these together as 'experiences', that they are occurrences of the same kind. This concept of 'experience' could be essentially disjunctive: roughly, to have an experience as of an object *o* could be either to perceive *o* or to merely seem to perceive *o*.¹ The challenge for this kind of disjunctivist is how to respond to the 'causal argument from hallucination', which purports to give reason to treat hallucinations and veridical perceptions as being of the same fundamental kind.² Things get very complicated here and I shan't get into it further, because this form of disjunctivism does not provide the appropriate analogy for the factivist view.

A better comparison with the factivist view is given by a disjunctivism motivated by epistemological concerns. The threat in this case is the argument from illusion. As John McDowell summarises that argument, it says that

since there can be deceptive cases experientially indistinguishable from non-deceptive cases, one's experiential intake ... must be the same in both kinds of case. In a deceptive case, one's experiential intake must *ex hypothesi* fall short of the fact itself, in the sense of being consistent with there being no such fact. So that must be true, according to the argument, in a non-deceptive case too.

So, the argument continues, when one acquires knowledge through one's capacity to tell how things are by looking,

we have to conceive the basis [of one's knowledge] as a *highest common factor* of what is available to experience in the deceptive

¹Hinton (1973).

²See in particular Martin (2004, 2006); Snowdon (2005).

and non-deceptive cases alike, and hence as something that is at best a defeasible ground for the knowledge³

'Defeasible', here, means something like 'non-conclusive'. The epistemological disjunctivist's concern is that if, in having an experience as of its being the case that p , one is aware of something less than its actually being the case that p , then to acquire knowledge of the external world one will need to make an inference from the grounds provided by one's experience. But this inference will need to be supported by some sort of theory, which theory would itself need to be acquired on the basis of perceptual experience.⁴ To avoid scepticism, this disjunctivist says, we must conceive of experience providing grounds in the 'good case' that it cannot provide in the 'bad case': *factive* grounds.

I suggest that this McDowellian epistemological disjunctivism provides the relevant analogy to the factivist view because its concerns are, to a great extent, the same. I'll substantiate this further in the final section. Indeed, I will suggest there that the relationship is actually rather closer than that of analogy. Both the factivist and the McDowellian disjunctivist hold that in order to make proper sense of a certain aspect of rational psychology, we must in some cases attribute a kind of contact between agent and fact which cannot be made sense of in terms of the conjunction of a non-factive analogue of that contact together with certain other conditions being met. This is quite different from the disjunctivism concerned with object-perception, which is stimulated by problems quite proprietary to its metaphysical picture of nature of perceptual experience. Because of this, it is not clear that the epistemological disjunctivist, or the factivist, must reject the idea that there is a relevant psychological (experiential) element in common between the good and bad cases.⁵ Factivism may, then, be consistent with a merely 'non-conjunctivist' position.⁶ On this approach, we simply reject the belief-first theorist's claim about the priority of psychologised rationalisation, and say that at least sometimes, when an agent A ϕ -s because p , the content of this explanation cannot be (non-trivially) captured by the conjunction of A 's ϕ -ing because she thought that p with further conditions. We need not add, as the object-perception disjunctivist would regarding the case that concerns them, that there is no unified element, picked out by psychologised rationalisation, that's common across cases in which worldly rationalisation is available and those in which it isn't.

At the end of this chapter, I will attempt to outline a (rough, conditional) case, based precisely on the kinds of considerations that motivate McDowell, for there being worldly rationalisations that cannot be reduced to any psy-

³McDowell (1998, 386).

⁴McDowell (1998, 391).

⁵Snowdon (2005); Byrne and Logue (2008).

⁶Cf. Williamson (2000, 44–8).

chologised rationalisations. First, however, I want to briefly consider some recent accounts of rational action that have approached a factivist account.

5.2 Hornsby

Jennifer Hornsby presents what she calls a ‘disjunctive conception of acting for reasons’, and connects the account she gives with McDowell’s disjunctivism about perceptual knowledge.⁷ Although Hornsby’s discussion is important, the position she presents does not directly address the issue on which belief-first and factivist theorists disagree, because her concerns are quite distinct from those on which we have been focussing.

As I read Hornsby, she is primarily arguing for a claim which I have mostly been taking for granted: that both worldly and psychologised rationalisations are legitimate forms of rationalisation, that both essentially make actions intelligible as rational, and that worldly rationalisations are factive. Hornsby’s target is not as such the claim that all worldly rationalisations can be reduced to psychologised rationalisations, but the claim that there *are no such things* as worldly rationalisations. Her thesis is that sometimes, when an agent acts in a rationalisable way, her action can be rationalised simply by citing worldly reasons, reasons for which the agent acts. Clearly, this must be granted if *our* question is even to be raised.

Strange as it may seem, though, this claim is rejected by authors with otherwise quite diverse views. Some have held that the reasons for acting and the reasons for which we act are just two different sorts of thing, the latter being not worldly reasons but attitudes like belief.⁸ Others claim that although our reasons are the always the objects of our attitudes rather than the attitudes themselves, it is always the attitudes and not their objects that explain our actions.⁹ Of particular note here is Jonathan Dancy, who thinks that worldly reasons do all the work, and that worldly reasons are facts (or ‘states of affairs’), but who also holds that rationalisations, even of the superficially worldly form, are non-factive.¹⁰ In other words, in the context of rationalisation, ‘A ϕ -ed because p ’ needn’t entail that p —or if it does, it is not because of anything internal to the explanation itself but merely due to the ‘factive pressure of the word “because”’, and the same explanation can be given in a non-factive way.¹¹ Appropriately enough, Dancy seems to be moved by a version of the argument from illusion. He rightly observes that a rationalisation explains an action as rational by showing how it made sense from the agent’s perspective, and that the possibility of error means we need

⁷Hornsby (2008).

⁸Smith (1994).

⁹Davis (2005).

¹⁰See Dancy (2000, 2011, 2014).

¹¹Dancy (2011, 350).

a way of revealing the light in which the agent saw things as a way of explaining why he did what he did, but without asserting that he was right to see things that way.¹²

Quite so. But Dancy takes the further much more dubious step of claiming that *all* rationalisations are essentially like this.

Hornsby's aim, I take it, is to challenge views of this sort.¹³ Hornsby stresses that in order for action to be explained by a worldly rationalisation citing a certain fact, the agent must have a certain sort of cognitive access to, or awareness of, that fact, and following a number of other authors,¹⁴ she suggests the relevant kind of access is provided by knowledge.

Building on this point, she proposes that if an agent ϕ -s because she believes that p , then either she ϕ -s because she knows that p and therefore because p , or she ϕ -s because she merely believes that p . This is the sense in which Hornsby's account is disjunctive, and it seems clear to me that this disjunctive account does the job of undermining Dancy's argument from illusion, showing as it does how an action's being rationalisable by facts about the agent's beliefs needn't preclude its also being rationalisable by worldly reasons. As we have seen, though, recognising the legitimacy of worldly rationalisation does not commit us to factivism.

5.3 Williamson

Perhaps there is a distinctively factivist picture lurking within Hornsby's discussion, however. Hornsby seems to suggest that it is in recognising a distinctive role for *knowledge* that we find the distinctiveness and autonomy of worldly rationalisations. One way we might develop this thought is to try to understand worldly rationalisation, or the notion of responding to a reason, explicitly in terms of knowledge. (I'm not suggesting that this is what Hornsby has in mind.) The idea that knowledge is autonomous with respect to belief has recently gained a great deal of traction in epistemology. 'Knowledge-first' theorists reject the traditional project of seeking to define knowledge in terms of belief, and take knowledge to be a quite distinctive state with an explanatory role not reducible to that of belief.¹⁵ The factivist says that acting for a reason is not to be understood in terms of treating-as-a-reason, and Hornsby tells us that the former requires knowledge. Combining the recognition of worldly rationalisation with the ideas that, first, worldly rationalisation presupposes knowledge and, second, knowledge cannot be

¹²Dancy (2000, 133).

¹³See also Hyman (2011); McDowell (2013), who respond directly to Dancy, the latter building on Hornsby's account.

¹⁴See Unger (1975); Hyman (1999); Williamson (2000).

¹⁵The *locus classicus* of this way of thinking is of course Williamson (2000). See McGlynn (2014) for critical discussion.

understood in terms of belief could give worldly rationalisation a kind of derivative autonomy. However, I think the prospects of arriving at an interesting factivist position by this route are actually rather dim.

Perhaps the most obvious way to expand the knowledge-first enterprise in epistemology to encompass a factivist approach to rational action would be to say something like the following. For someone to ϕ because they think that p requires, transparently, that they think that p . And as Hyman, Hornsby, and others argue, for someone to ϕ because p requires that they know that p . Given that the fact cited in explaining the agent's action in the former case is a fact about the agent's believing something, it is not unnatural to think that we should understand psychologised rationalisation in terms of belief. Similarly, one might argue, what is really explaining the agent's p -ing when we give a worldly rationalisation is that the agent knows that p . One might suggest that we should understand acting for a reason and acting on a belief essentially in terms of the psychological operation of states of knowledge and belief respectively. If this is right, and if knowledge is autonomous with respect to belief as the knowledge-first epistemologist claims, then worldly rationalisation will be, derivatively, autonomous with respect to psychologised rationalisation.

The major challenge for this approach will be to explain how we should understand the two forms of rationalisation in terms of knowledge and belief. It is not at all clear that we can say *what it is* to act for a reason, or *what it is* to act for an apparent reason, in terms of knowledge and belief respectively. A natural way to do so would be to posit some distinctive kind of rational causation, and to say that (for the worldly case) someone ϕ -s because p just in case their ϕ -ing is caused in the relevant way by their knowing that p . It will need to be explained what the relevant kind of causation is—if it is left primitive, then we are effectively leaving the notion of acting for a reason primitive, and so not as such explaining acting for a reason in terms of knowledge. The prospects for this project, though, look very poor. Notoriously, theories of this sort run into the problem of 'deviant causal chains': counterexample cases in which the specified conditions for the 'right kind of causation' are met, but in which, intuitively, the agent does not act for the relevant reason.¹⁶ The deviant causal chain problem is remarkably similar to the Gettier problem in epistemology. The failure of the traditional epistemological project to satisfactorily answer the latter problem is a major part of the motivation for knowledge-first epistemology: the inability of accounts of knowledge in terms of belief to avoid the Gettier problem is taken as cause for serious pessimism about the prospects of that project. It is hard to see why someone who responds in this way to the Gettier problem would not respond in a similar way to the problem of deviant causation, thus taking a

¹⁶See Mayr (2011, chap. 5) for an overview of various failures to solve the problem.

dim view of any attempt to analyse acting for reasons in terms of a specified kind of causal connection.

However, if our method of expanding knowledge-first epistemology into a factivist theory of rational action isn't reductive in something like the way considered above, then it's not clear that it will have anything distinctive to contribute to our enquiry concerning the two forms of rationalisation and the connection between them. If we look for an account based on weaker claims than the reductive ones just considered, there are a couple of aspects of the connections between worldly rationalisation and knowledge, and between psychologised rationalisation and belief, that we might focus on. One is merely the idea that an agent's being in the relevant state is a necessary condition for the availability of the relevant rationalisation. This, together with a knowledge-first epistemology, would not as such constitute a factivist theory. Perhaps the correct belief-first theory of worldly rationalisation is just that a worldly rationalisation is available when an agent does something on the basis of their belief that p , whilst knowing that p . How informative such a theory would be would depend, of course, on whether we could give some kind of substantive positive account of doing something on the basis of a belief, but that we could is by no means ruled out by the claim that knowledge is autonomous with respect to belief.

The other thing that the knowledge-first theorist might want to say is that one's ϕ -ing because p just is one's ϕ -ing's being caused in the right way by one's knowledge that p , but that the 'right kind' of causation is not something of which we could give an independent or reductive account.¹⁷ This would be a factivist view since it would entail that we couldn't give an independent or reductive account of worldly rationalisation in terms of psychologised rationalisation, but the account wouldn't really gain any content from its token association with knowledge-first epistemology. We would still need an argument for *why* we shouldn't think of the right kind of causation in the knowledge case as being explicable in terms of a kind of causation common between the cases. In other words, we would still be without an argument for factivism. It is hard to see how such an argument would come from the association of the factivist view with knowledge-first epistemology—the association would be little more than an embellishment of the factivist picture.

5.4 Hyman

One place where I think a developed and distinctive factivist view can be found is in the work of John Hyman. Hyman argues not just that knowledge is a necessary condition on someone's responding to a reason, but that knowledge should be understood essentially in such terms. For Hyman, knowing

¹⁷This would be, in a way, a sort of knowledge-first, factivist Davidsonian view.

that p simply is having the ability to be *guided* by the fact that p , where fact-guided behaviour is exactly the kind of behaviour distinctively explicable by worldly rationalisation.¹⁸ So for Hyman, knowledge itself is to be understood in terms of the kind of behaviour explicable by worldly rationalisation.

In itself, this doesn't force factivism, since acting for a reason or being guided by a fact might itself be accounted for in terms of acting on the basis of a belief, or for an apparent reason. Hyman's account is factivist in virtue of its combination of his account of knowledge with a distinctive account of belief: believing that p , Hyman suggests, is being disposed 'to act (think, feel) as one would if one knew that p , or as one would if one were guided by the fact that p '.¹⁹ Hyman's theory thus gets to psychologised rationalisation from worldly rationalisation, *via* knowledge and belief. It is factivist in that it holds psychologised rationalisation to be explicable in terms of worldly rationalisation: knowledge is defined in terms of doing things for reasons, belief is defined in terms of knowledge. Presumably psychologised rationalisation is explained in terms of belief: one's ϕ -ing is explicable by a psychologised rationalisation when it manifests one's disposition to behave as one would if one knew.²⁰

Here might be a good juncture at which to consider the case for taking knowledge to be required for an agent to respond to a reason. Hyman and Hornsby make much the same argument, which takes the form of an inference to the best explanation.²¹ We can begin with some basic observations. For someone to act because p , it must seem to them that p , in much the same way as it must for them to act because they think that p . Moreover, it must be the case that p , since, in general, ' q because p ' entails both that q and that p . However, the conjunction of these conditions is not enough to make possible a worldly rationalisation of someone's acting in terms of the fact that p . The mere truth of what one believes doesn't guarantee that there is any explanatory connection between that truth and one's action. A potential fix might be that one's belief must be justified, or based on good reasons. As Hyman and Hornsby observe, though, such an account fails for worldly rationalisation just as it fails for knowledge, as is shown by Gettier cases. Hornsby gives the following example:

Edmund ... believes that the ice in the middle of the pond is dangerously thin, having been told so by a normally reliable friend, and ... accordingly keeps to the edge. But Edmund's friend didn't want Edmund to skate in the middle of the pond (never

¹⁸Hyman (1999, 2006, 2010, 2015).

¹⁹Hyman (2015, 173).

²⁰It's worth pointing out that, for Hyman, these are all 'connective' rather than reductive, analyses—so his aim is not to put worldly rationalisation first in the same sense in which the belief-first theorist puts psychologised rationalisation first. See Hyman (2006, 908–909), Hyman (2015, 42).

²¹Hyman (1999); Hornsby (2008).

mind why), so that he had told Edmund that the ice there was thin despite having no view about whether or not it actually was thin. Edmund, then, did not keep to the edge because the ice in the middle was thin. Suppose now that, as it happened, the ice in the middle of the pond was thin. This makes no difference. Edmund still didn't keep to the edge because the ice was thin. The fact that the ice was thin does not explain Edmund's acting, even though Edmund did believe that it was thin, and even though the fact that it was thin actually was a reason for him to stay at the edge.²²

The unavailability of worldly rationalisation in such cases calls for explanation, and Hyman and Hornsby suggest that the best explanation is that, for one to ϕ because p (in the sense of worldly rationalisation), one must know that p .

This argument has recently been challenged, however. Nick Hughes and Dustin Locke (independently) appeal to cases structurally similar to Alvin Goldman's 'fake barns' case,²³ and argue that in such cases an agent can act for a reason he does not know. Here is Hughes:

Henry is out hiking. He's lost, and the weather is turning nasty. The situation is getting serious. He sees what he believes to be a hiker's hut in the distance, and feels relieved. In fact, unbeknownst to Henry, he is in fake hiker's-hut county—an area where there are only a handful of real huts, and many hut-facades designed to look exactly like real huts to passing hikers. Henry justifiably and truly believes that the structure in the distance is a hut, but he does not know this.²⁴

Hughes suggests that in this case, 'Henry feels relieved because there is a hut in the distance' would be a legitimate, and genuinely rationalising, explanation of Henry's feeling relieved. If that's correct, then while knowing that p might well be sufficient to put one into a position to do things because p , it would seem that it is not, in general, necessary.

Hyman is aware of Locke's paper²⁵ (which, again, makes much the same argument as Hughes's), but doesn't respond to the argument directly. He does discuss fake-barn cases, however, and takes it to be 'clear' that in these cases worldly rationalisations are *not* available.²⁶ This is fair enough. Intuitions are intuitions, and these cases seem very close to the boundary of the applicability of worldly rationalisation, whether we think they fall inside

²²Hornsby (2008, 251).

²³See Goldman (1976).

²⁴Hughes (2014, 461).

²⁵See Hyman (2015, 231, n. 22).

²⁶Hyman (2015, 156.).

or out.²⁷ Nonetheless, it would be nice to be able to appeal to more than mere intuition here. One possible line of response would be to challenge the Hughes–Locke judgement directly, providing both diagnosis of their mistake and an explanation of why worldly rationalisation does after all fail in these kinds of cases.²⁸ In the present context, though, with our focus on the relationship between worldly and psychologised rationalisation, it would be better, I think, to simply say the following. What is required for an agent to be in the position to respond to the fact that p is that the agent be *aware* that p , or *have access to* the fact that p , where awareness of or access to a fact is a factive cognitive condition. It may be that being aware of a fact just is, after all, knowing that fact. Or perhaps something more is required; perhaps knowing that p is a matter of both being aware that p and believing that p with a certain sort of normative warrant.²⁹ The relationship between awareness and knowledge is a good question, but our question is about the relationship between worldly and psychologised rationalisation, and we will do better, if we can, to address the latter question without taking a stance on the former. And for our purposes, it seems we could retain most aspects of Hyman’s account, simply substituting ‘access’ for ‘knowledge’.³⁰

Hyman’s account is ambitious. His account of rationalisation entails, but is also perhaps somewhat stronger than, the non-conjunctivist factivist thesis. We’ve seen a threat to one aspect of the account—the definition of knowledge—and seen that its factivist character can be retained whether that challenge is accepted or rejected. However, the account’s factivist status, and its suitability to account for rationalising explanation, depend equally on its treatment of belief, and Hyman merely presents this as a plausible suggestion, rather than explicitly arguing for it.³¹ Without the account of belief, the factivist status of Hyman’s picture is not secured. A crucial question will be whether his account of belief does proper justice to its rationalising role, in particular if that role is fleshed out in terms of the notion of the agent’s point of view as developed in the previous chapter. But let’s just take Hyman’s account as a model— it’s one potential way of developing a thoroughgoing factivist account. Let’s turn, finally, to the question of whether the factivist approach is motivated as such.

²⁷Cf. the send-up of fake-barn intuition-pumping in Gendler and Hawthorne (2005).

²⁸Littlejohn (2014) has a crack.

²⁹Compare the account of memory-knowledge in the final chapter of Peacocke (1986).

³⁰Hyman gives other reasons for finding the equation of knowledge and access attractive, as discussed in Hyman (2015, especially chap. 8). It’s a nice question whether these would still apply if we understood knowledge in terms of access as suggested. Trying to answer that question will have to be left for another time.

³¹Quite understandably, since the positive account of belief makes up a very minor part of his discussion, which is primarily concerned with knowledge.

5.5 An argument for factivism

An argument for the factivist approach will seek to identify a case in which we want to give a kind of worldly rationalisation that cannot be reduced to the corresponding psychologised rationalisation. The case I want to discuss is something of a departure from what has been the primary focus of the present essay, the rationalisation of action. As indicated earlier, it concerns the epistemology of perception, and is about believing, rather than acting, for a reason. One challenge will be to show how the point, if successful, generalises to action as such. I'll say something briefly about this at the end.

Another caveat is that the argument's suasive force will be, to a significant degree, conditional, in that it rests on substantive and controversial—but hopefully not implausible—assumptions. It's perhaps best seen as a development of McDowell's epistemological project,³² and as such is unlikely to convince those who lack McDowellian inclinations. Nevertheless, success even within this limited scope would be a significant result. We have not yet seen a cogent case for factivism, and it will be instructive to see what one might look like. So: on to the case.

I want to think about the justification that one can have for believing that p in virtue of its visually seeming to one that p . (Call such a belief a 'perceptual belief'.) As long as we are not radical sceptics, we will want to say that one can often be justified in believing that p on the basis of its looking to one as if p . But it's a somewhat fraught question why and how such beliefs are justified. I want to suggest that the factivist is in a position to give an attractive answer to this question—an answer that is not available to the belief-first theorist.

So our first assumption is that perceptual beliefs are often justified in virtue of one's having the perceptual experience in which they are, one way or another, grounded. Our second assumption is that a certain principle about the justification of beliefs is correct: whenever someone who believes that p is justified in believing that p , they have a justification for believing that p , and this justification can be represented abstractly in the form of an argument to the conclusion that p . For this justification to be what in fact justifies the agent's belief, it must somehow correspond to the actual cognitive basis of the agent's having that belief.³³ Within our framework, it seems natural to say that the abstractly-formulated argument must correspond in a suitable way to a correct rationalisation of the agent's believing what they do. In a way, the demand that justified beliefs have a justification representable in abstract argument form is just the demand that justified beliefs be susceptible to satisfactory rationalisation.

³²See for instance McDowell (1998, 1994, 1995).

³³Compare Harman (1964), who helpfully distinguishes abstract inference from particular acts of inference, and discusses how we should think of the two as being related.

There is a puzzle about how to apply this principle to the justification of perceptual belief. That is: When it looks to me as if p , and I justifiably believe that p as a result, what is the justification of my believing that p ? A natural thought is that the justifying premise is that it looks to me as if p . The problem with this is that it's very hard to see how such a premise could ever afford sufficient justification for my belief about the external world. That p clearly doesn't follow from its looking to me as if p . Suppose I believe that there's a barn over there because it looks to me as if there's a barn over there. Its looking to me as if there is a barn over there is quite consistent with there being no such barn. It might be a mere barn-façade. I might be hallucinating. If an operative justifying premise for my perceptual belief is to be a proposition about how things look to me, we need some further explanation of why I am entitled to infer from things looking a certain way to their being that way. Given that what we are looking for is an abstract representation of my justification for holding the perceptual belief, we have two options. The first is to argue that there is a general, valid rule of inference from propositions of the form 'it looks to me as if p ' to ' p '. This seems a clear non-starter. The alternative is to add further premises. Perhaps my justification for believing that p is that (i) it looks to me as if p , and (ii) the best explanation for its looking to me as if p is that p . But if the second premise is to justify my believing that p , it must play the right kind of role in grounding that belief, and it must be legitimate for it to play that role. It's natural to think that I'd need to at least be justified in believing that the best explanation of its looking to me as if p is that p . It's very hard indeed to see how I could have warrant for accepting such a principle about the explanation of appearances without having antecedent warrant for having certain sorts of general beliefs about how things work in the external world, and it's hard to see how I could have *that* warrant except through perceptual experience. Similar problems will arise, it seems, for any substitute for (ii) that's meant to play a similar role of bridging the gap between non-factive appearances and worldly facts.³⁴

Some authors, faced with this puzzle, favour rejecting the principle about justification. Perhaps perceptual beliefs are *immediately* justified, where A is immediately justified in believing that p just in case A is justified in believing that p , but is not so justified in virtue of 'some relation this belief has to some other justified belief(s)' of A 's.³⁵ I think that this is in fact the correct response; we shouldn't think of my belief that there is a barn over there as being justified by inference from any other beliefs. However I think we must also retain the principle that I must have a *justification* for my belief, in the sense discussed above, and this is something that the authors I have in mind reject. James Pryor defends the rejection of the principle by saying that having a justification for believing that p doesn't mean that you must 'always be able

³⁴Compare McDowell (1998, 1994, 1995).

³⁵Alston (1983, 74).

to offer *reasons* ... in support of your belief.³⁶ Pryor is right to distinguish between a belief's being justified and the activity of justifying a belief, but he fails to consider a third thing, *there being a justification* for a belief. What he says gives us no reason to doubt that, when your belief is justified, there must be such a justification to which your belief is suitably connected. That doesn't imply that you, the believer, must be in a position to say what the justification is. It's we, the epistemologists, who can't rest content with the idea that your belief is justified by nothing. Compare the principle that if someone acts rationally there must be a rationalisation of their action. I might be unable to tell you why I did what I did (perhaps I forgot); but if there's simply no story to tell, should we not conclude that my action was, after all, not rational? Clayton Littlejohn rejects the principle on the basis that it's false as applied to action: it can be reasonable for us to do some things—idle doodling is one of his examples—for no reason at all, not even an apparent one.³⁷ However, there's a significant disanalogy here. It can be reasonable to idly doodle for no reason when there is no reason not to idly doodle. Idle doodling doesn't, in general, incur any significant costs. Given the potential costs of false belief, though, we might still think that one always needs sufficient reason to think a proposition is true in order to justifiably believe it. Moreover, many authors have argued that you ought only believe what is true, including Littlejohn himself.³⁸ It's not far from that idea to the thought that it would be epistemically irresponsible to believe that *p* without doing so on the basis of some (apparent) reason for thinking that *p* is true. Doodling is something that it can make sense to do just because I feel like it. Believing I have hands isn't.

If we can't get a good justificatory argument from its looking to me as if *p*, what sort of argument can we give? The answer I'd like to recommend is that the relevant premise is simply: *p*. The argument employs the repetition rule. In the good case, when I see that *p* and believe that *p* as a result, I believe that *p* because *p*, and this explanation is genuinely rationalising. To say that, we simply need to say that seeing that *p* is a way of being aware of the fact that *p*, and that seeing that *p* is antecedent to believing that *p*. Both points seem particularly plausible if we accept the Hughes–Locke analysis of the fake barns-type case: Henry sees that there is a hut and thus has access to that fact, but doesn't know it.³⁹ If he discovers that he is in fake hut country, he will be rationally obliged to suspend judgement, and so refrain from doing anything further in response to that fact; but his suspending judgement does not mean he is no longer in a (cognitive, epistemic) position to respond to that fact. In

³⁶Pryor (2000, 535).

³⁷Littlejohn (ms).

³⁸For example Wedgwood (2002); Shah (2003); Littlejohn (2012); Engel (2013); Whiting (2013).

³⁹This is consistent with McDowell's account, which takes perception to ground knowledge. It's not compatible with Williamson's claim that seeing that *p* is simply knowing that *p* in a certain way.

the bad case, the agent believes that p not because p , but because it looks to them as if p . This is not the reason for which they believe, though—it is a psychologised rationalisation, which corresponds to the worldly rationalisation just as a belief-citing psychologised rationalisation corresponds to a worldly rationalisation. Here, we understand the agent's response by considering literally how things looked from their point of view.⁴⁰

In both good and bad cases, the form of the justifying argument is: p , therefore p . Of course, this won't be a cogent argument to someone who doubts whether p , but the point of the justifying argument is not to convince such a person. The point is to explain how one's belief is based on something that justifies one in believing it. There could hardly be a better justification for believing that p than the very fact that p . The key question is whether we can make sense of the idea of such a belief being based on such a fact.

Obviously we can't say that the agent believes that p because she believes that p . That wouldn't even be an explanation, rationalising or otherwise. We must instead think of seeing that p as a way of having access to the fact that p that's independent of believing that p , and we must hold that one can believe something for a reason that one does not already believe. Littlejohn rejects the idea that one's reason for believing that p can be the fact that p itself, on the grounds that being aware that p involves the exercise of conceptual capacities, which exercise is 'distinctive of belief'. Since conceptual capacities are exercised also in supposition, imagination, linguistic comprehension, and so on, I take this to be a non sequitur.⁴¹

It would be too quick to conclude that a belief-first theorist cannot endorse the picture just on the basis that perceptual awareness is pre-doxastic. There is no need for a belief-first theorist to take all psychologised rationalisation to be belief-citing, and there is a non-belief-citing psychologised rationalisation to be given in the perceptual case: the subject believes what they believe because things look to them to be that way. The argument for factivism is that the worldly rationalisation that we are considering—one of the form 'A believes that p because p ', when A sees that p —cannot be captured in terms of this 'looks-citing' psychologised rationalisation. It cannot be so captured because we need to explain the way in which the appearance rationalises the belief by appealing to the way in which the fact rationalises the belief when the agent sees that things are as they are.

This can be brought out by considering the phenomenon of defeat. Henry, initially, is justified in thinking that the thing over there is a hut. After he finds out that he is in fake hut country, however, he loses that justification. Why? Not everything that looks like a hut in fake hut country is a fake hut. There

⁴⁰Williamson (2000, 198) suggests that the looks-fact *is* the agent's reason. If what I've said is right, that would suggest that such subjects' beliefs are unreasonable. I take it to be an advantage of the view I'm suggesting that it lacks this consequence.

⁴¹Littlejohn (ms).

are real huts too. That Henry is in fake hut country is no reason to believe that that thing is not a hut. It's a reason not to trust appearances. But why should Henry's being in fake hut country be a reason for him not to trust appearances? It can't be that it undermines one of the premises in the justification of his belief. It might if one of the premises were, say, that that thing's being a hut is the best explanation of its looking like a hut—the fact that Henry is in fake barn country would falsify *that* premise by placing on the table an equally good alternative explanation. But we have dismissed such accounts of Henry's justification. We want to say that Henry's only relevant premise is that that's a hut. The explanation of the defeat must be that once he finds out that he is in fake hut country, Henry is no longer entitled to base his beliefs on that premise.

Proponents of the idea that perceptual beliefs are immediately justified by the subject's experience appeal to the phenomenal character of perceptual experience in order to explain the immediate justification that such experience can provide for one's beliefs.⁴² When it looks to one as if *p*, one '*seem[s]* to ascertain that a proposition is true';⁴³ it is, for one, as if one is seeing that such-and-such is the case. The phenomenology of one's experience need not change, however, when one learns that appearances are unreliable. What changes is that one now has some reason to think that one's experience may be illusory. The factivist has an explanation of why this defeats one's justification: it suggests that one might not be believing what one believes *because* what one believes is true. That's not something we could say if we held that the fundamental way of rationalising such beliefs is by saying that one believes what one does because it seems to one to be true—because even when one is aware of the defeating condition, it continues to seem that way to one.

An adequate rationalisation makes sense of a belief as rational. When I am aware of a defeating condition, its looking to me as if *p* can no longer make sense of my believing that *p* as rational. This, I am suggesting, is because I am no longer in a position to think of myself as believing that *p* because *p*. If we want to say that one's justification for believing that *p* in the case of perceptual belief is simply that *p*, as I have suggested we might, then that seems to be the only way to make sense of defeat. It is, however, an explanation that only the factivist is in a position to give.

Supposing this argument is accepted, where have we got to? We are not yet at a thoroughgoing factivist account like Hyman's, for we might think that perceptual belief is a special case. In particular, it might be held that the considerations applying to theoretical rationality don't carry over to practical rationality and the rationalisation of action. The factivist will have to give some account of why we should generalise the point. I can't give an argument here, but I would like to indicate a possible way forward. Suppose we

⁴²See Pryor (2000, 2004); Silins (2014).

⁴³Pryor (2004, 357).

think, with Williamson and especially Hyman, that there is a distinctive kind of cognitive relation which an agent must bear to a fact in order for that fact to potentially rationalise that agent's doings, and we think of worldly rationalisations as implicitly invoking the notion of guidance by a fact which presupposes such a relation.⁴⁴ If we are so inclined, then we might take the case of perceptual awareness as both a paradigm of that relation and a demonstration of that relation's not being reducible in terms of any non-factive attitude. That might help secure more general significance for the factivist argument sketched above. These remarks are obviously very far from a making demonstrative case, or even anything that could properly be called an argument, but I think there may be something more to be said for the line of thought at which they gesture.

⁴⁴Perhaps the idea of there being such a unified relation stands to do significant explanatory work—see Williamson (2000); Hyman (2014, 2015).

Bibliography

- Alston, W. P. (1983). What's Wrong with Immediate Knowledge? *Synthese*, 55(1):73–95.
- Alvarez, M. (2010). *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford University Press, Oxford.
- Andrews, K. (2003). Knowing Mental States: The Asymmetry of Psychological Prediction and Explanation. In Smith, Q. and Jokic, A., editors, *Consciousness: New Philosophical Perspectives*, pages 201–219. Oxford University Press, Oxford.
- Anscombe, G. E. M. (1963). *Intention*. Basil Blackwell, Oxford, 2nd edition.
- Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. Routledge & Kegan Paul, London.
- Arpaly, N. (2002). *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press, Oxford.
- Arpaly, N. and Schroeder, T. (2012). Deliberation and Acting for Reasons. *Philosophical Review*, 121(2):209–239.
- Austin, J. L. (1950). Truth. *Proceedings of the Aristotelian Society Supplementary Volume*, 24:111–128.
- Bennett, B. C. (2007). Doctrine of Signatures: An Explanation of Medicinal Plant Discovery or Dissemination of Knowledge? *Economic Botany*, 61(3):246–255.
- Blackburn, S. (1995). Practical Tortoise Raising. *Mind*, 104(416):695–711.
- Blackburn, S. (2010). The Majesty of Reason. *Philosophy*, 85(01):5.
- Byrne, A. and Logue, H. (2008). Either/Or. In Haddock, A. and Macpherson, F., editors, *Disjunctivism: Perception, Action, Knowledge*, pages 57–94. Oxford University Press, Oxford.
- Carroll, L. (1895). What the Tortoise Said to Achilles. *Mind*, 4(14):278–280.
- Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78(2):67–90.

- Collingwood, R. G. (1946). *The Idea of History*. Oxford University Press, Oxford.
- Collins, A. W. (1997). The Psychological Reality of Reasons. *Ratio*, 10(2):108.
- Comesaña, J. and McGrath, M. (2014). Having False Reasons. In Littlejohn, C. and Turri, J., editors, *Epistemic Norms: New Essays on Action, Belief, and Assertion*, pages 59–79. Oxford University Press, Oxford.
- Crane, T. (1991). All the Difference in the World. *The Philosophical Quarterly*, 41(162):1–25.
- Dancy, J. (2000). *Practical Reality*. Clarendon Press, Oxford.
- Dancy, J. (2011). Acting in Ignorance. *Frontiers of Philosophy in China*, 6(3):345–357.
- Dancy, J. (2014). On Knowing One's Reason. In Littlejohn, C. and Turri, J., editors, *Epistemic Norms: New Essays on Action, Belief, and Assertion*, pages 81–96. Oxford University Press, Oxford.
- Davidson, D. (1980). *Essays on Actions and Events*. Oxford University Press, Oxford.
- Davidson, D. (1991). Three Varieties of Knowledge. *Royal Institute of Philosophy Supplement*, 30(1991):153–166.
- Davidson, D. (2001). *Inquiries into Truth and Interpretation: Philosophical Essays, Volume 2*. Oxford University Press, Oxford, 2nd edition.
- Davidson, D. (2004). *Problems of Rationality*. Oxford University Press, Oxford.
- Davies, M. and Stone, T. (1998). Folk Psychology and Mental Simulation. In O'Hear, A., editor, *Current Issues in the Philosophy of Mind*, pages 53–82. Cambridge University Press, Cambridge.
- Davis, W. A. (2005). Reasons and Psychological Causes. *Philosophical Studies*, 122(1):51–101.
- Engel, P. (2013). Doxastic Correctness. *Proceedings of the Aristotelian Society Supplementary Volume*, 87(1):199–216.
- Fodor, J. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, 28:97–115.
- Fodor, J. (1991). A Modal Argument for Narrow Content. *The Journal of Philosophy*, 88(1):5–26.
- French, C. (ms). The Rationalizing Role of Seeing an Object. Unpublished manuscript.

- Gendler, T. S. and Hawthorne, J. (2005). The Real Guide to Fake Barns: A Catalogue of Gifts for Your Epistemic Enemies. *Philosophical Studies*, 124(3):331–352.
- Goldman, A. I. (1976). Discrimination and Perceptual Knowledge. *Journal of Philosophy*, 73(20):771–791.
- Hampshire, S. (1999). The Reason Why Not. *The New York Review of Books*, April 22.
- Harman, G. H. (1964). How Belief is Based on Inference. *The Journal of Philosophy*, 61(12):353–359.
- Heal, J. (2003). *Mind, Reason and Imagination*. Cambridge University Press, Cambridge.
- Hinton, J. M. (1973). *Experiences: An Inquiry into Some Ambiguities*. Clarendon Press, Oxford.
- Hornsby, J. (2008). A Disjunctive Conception of Acting for Reasons. In Haddock, A. and Macpherson, F., editors, *Disjunctivism: Perception, Action, Knowledge*, pages 244–261. Oxford University Press, Oxford.
- Hossack, K. (2007). *The Metaphysics of Knowledge*. Oxford University Press, Oxford.
- Hughes, N. (2014). Is Knowledge the Ability to Φ for the Reason That P? *Episteme*, 11(04):457–462.
- Hyman, J. (1999). How Knowledge Works. *The Philosophical Quarterly*, 49(197):433 – 451.
- Hyman, J. (2006). Knowledge and Evidence. *Mind*, 115(460):891–916.
- Hyman, J. (2010). The Road to Larissa. *Ratio*, 23(4):393–414.
- Hyman, J. (2011). Acting for Reasons: Reply to Dancy. *Frontiers of Philosophy in China*, 6(3):358–368.
- Hyman, J. (2014). 'The Most General Factive Stative Attitude'. *Analysis*, 74(4):561–565.
- Hyman, J. (2015). *Action, Knowledge, and Will*. Oxford University Press, Oxford.
- Jackson, F. (2009). Primeness, Internalism, Explanation. In Greenough, P. and Pritchard, D., editors, *Williamson on Knowledge*, pages 109–121. Oxford University Press, Oxford.
- Jackson, F. and Pettit, P. (1988). Functionalism and Broad Content. *Mind*, 97(387):381–400.

- Kalderon, M. E. (2011). Before the Law. *Philosophical Issues*, 21:219–244.
- Kolodny, N. (2005). Why Be Rational? *Mind*, 114(455):509–563.
- Lavin, D. (2011). Problems of Intellectualism: Raz on Reason and its Objects. *Jurisprudence*, 2(2):367–378.
- Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, 50(3):249–258.
- Lewis, D. (1983). *Philosophical Papers, Volume I*. Oxford University Press, Oxford.
- Liao, S.-y. and Gendler, T. S. (2016). The Puzzle of Imaginative Resistance. Forthcoming in *The Routledge Companion to the Philosophy of Literature*, eds. Noël Carroll and John Gibson.
- Littlejohn, C. (2012). *Justification and the Truth-Connection*. Cambridge University Press, Cambridge.
- Littlejohn, C. (2014). Fake Barns and False Dilemmas. *Episteme*, 11(04):369–389.
- Littlejohn, C. (ms). Knowledge and Awareness. Forthcoming in *Analysis*.
- Loar, B. (1981). *Mind and Meaning*. Cambridge University Press, Cambridge.
- Martin, M. G. F. (2004). The Limits of Self-Awareness. *Philosophical Studies*, 120:37–89.
- Martin, M. G. F. (2006). On Being Alienated. In Gendler, T. S. and Hawthorne, J., editors, *Perceptual Experience*, pages 354–410. Oxford University Press, Oxford.
- Mayr, E. (2011). *Understanding Human Agency*. Oxford University Press, Oxford.
- McDowell, J. (1985). Functionalism and Anomalous Monism. In LePore, E. and McLaughlin, B. P., editors, *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, pages 387–398. Basil Blackwell, Oxford.
- McDowell, J. (1994). Knowledge By Hearsay. In Matilal, B. K. and Chakrabarti, A., editors, *Knowing From Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, pages 195–224. Kluwer, Dordrecht.
- McDowell, J. (1995). Knowledge and the Internal. *Philosophy and Phenomenological Research*, 55(4):877–893.
- McDowell, J. (1998). Criteria, Defeasibility, and Knowledge. In *Meaning, Knowledge, and Reality*, pages 369–394. Harvard University Press, London.

- McDowell, J. (2010). Tyler Burge on Disjunctivism. *Philosophical Explorations*, 13(3):243–255.
- McDowell, J. (2013). Acting in the Light of a Fact. In Bakhurst, D., Hooker, B., and Little, M. O., editors, *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*, pages 13–28. Oxford University Press, Oxford.
- McGlynn, A. (2014). *Knowledge First?* Palgrave Macmillan, London.
- Miracchi, L. (2014). *Getting Things Done*. PhD thesis, the Graduate School—New Brunswick, Rutgers, The State University of New Jersey.
- Nisbett, R. and Wilson, T. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84:231–259.
- Peacocke, C. (1986). *Thoughts: An Essay on Content*. Basil Blackwell, Oxford.
- Peacocke, C. (1993). Externalist Explanation. *Proceedings of the Aristotelian Society*, 93:203–230.
- Prior, A. N. (1971). *Objects of Thought*. Oxford University Press, Oxford.
- Pryor, J. (2000). The Skeptic and the Dogmatist. *Noûs*, 34(4):517–549.
- Pryor, J. (2004). What’s Wrong with Moore’s Argument? *Philosophical Issues*, 14:349–378.
- Raz, J. (2000). *Engaging Reason: On the Theory of Value and Action*. Oxford University Press, Oxford.
- Raz, J. (2011). *From Normativity to Responsibility*. Oxford University Press, Oxford.
- Roessler, J. (2014). Reason Explanation and the Second-Person Perspective. *Philosophical Explorations*, 17(3):346–357.
- Roessler, J. and Perner, J. (2013). Teleology: Belief as perspective. In Baron-Cohen, S., Lombardo, M., and Tager-Flusberg, H., editors, *Understanding Other Minds*, pages 35–50. Oxford University Press, Oxford, 3rd edition.
- Salmon, W. C. (1989). *Four Decades of Scientific Explanation*. University of Minnesota Press, Minneapolis.
- Scanlon, T. M. (2003). Metaphysics and Morals. *Proceedings and Addresses of the American Philosophical Association*, 77(2):7–22.
- Shah, N. (2003). How Truth Governs Belief. *Philosophical Review*, 112(4):447–482.
- Short, T. (2015). *Simulation Theory: A Psychological and Philosophical Consideration*. Psychology Press, Hove, East Sussex.

- Silins, N. (2014). The Agony of Defeat? *Philosophy and Phenomenological Research*, 88(3):505–532.
- Smith, M. (1994). *The Moral Problem*. Blackwell, Oxford.
- Snowdon, P. (2005). The Formulation of Disjunctivism: A Response to Fish. *Proceedings of the Aristotelian Society*, 105:129–141.
- Stalnaker, R. (1989). On What's in the Head. *Philosophical Perspectives*, 3:287–316.
- Stalnaker, R. (1990). Narrow Content. In Anderson, C. A. and Owens, J., editors, *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*, pages 131–145. CSLI Publications, Stanford.
- Stich, S. P. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. MIT Press, London.
- Stich, S. P. and Nichols, S. (1992). Folk Psychology: Simulation or Tacit Theory? *Philosophical Issues*, 7(1 & 2):35–71.
- Strawson, P. F. (1950). Truth. *Proceedings of the Aristotelian Society Supplementary Volume*, 24:129–156.
- Stroud, B. (2000). *The Quest for Reality*. Oxford University Press, Oxford.
- Sylvan, K. (2015). What Apparent Reasons Appear to Be. *Philosophical Studies*, 172:587–606.
- Textor, M. (2011). Knowing the Facts. *Dialectica*, 65(1):75–86.
- Thompson, M. (2008). *Life and Action: Elementary Structures of Practice and Practical Thought*. Harvard University Press, Harvard.
- Unger, P. (1975). *Ignorance: A Case for Scepticism*. Clarendon Press, Oxford.
- Wedgwood, R. (2002). The Aim of Belief. *Philosophical Perspectives*, 16:267–297.
- Whiting, D. (2013). Nothing but the Truth. In Chan, T., editor, *The Aim of Belief*, pages 185–203. Oxford University Press, Oxford.
- Williams, B. (1979). Internal and External Reasons. In Harrison, R., editor, *Rational Action*, pages 17–28. Cambridge University Press, Cambridge.
- Williamson, T. (2000). Scepticism and Evidence. *Philosophy and Phenomenological Research*, 60(3):613–628.
- Yablo, S. (1992). Mental Causation. *The Philosophical Review*, 101(2):245–280.
- Yablo, S. (1997). Wide Causation. *Philosophical Perspectives*, 11:251–281.
- Yablo, S. (2003). Causal Relevance. *Philosophical Issues*, 13:316–328.