The University Of Sheffield.

Familiar Paths: The Rationality of Habitual Action

William Christopher Hornett

The University of Sheffield

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

2021

FAMILIAR PATHS: THE RATIONALITY OF HABITUAL ACTION

William Christopher Hornett
Department of Philosophy
January 2021

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Acknowledgements

There are at least two ways of writing acknowledgements. One is to thank only those people who contributed quite directly to the thesis and its writing. The other is to thank really as many people as possible who have contributed in any way to one's education, work, and quality of life. When I think about the sheer number of people to whom I am grateful, the first option just seems impossible. So the second it is.

I want to start with my family. Throughout not only this PhD, but my whole life, my parents Mike and Maria have been about as supportive and lovely as parents could be. I owe them both a huge amount for their emotional and financial support, their love, and everything they have taught me to do and be. As a surprise to no-one, I think I got my pernickety, argumentative nature from my Dad. What might surprise him is that this has contributed about as much to this work and my love of philosophy as anything I have ever read. So, thank you. I want to thank my Mum for always being a couple of missed calls away, for her good humour, and her love. So, thank you too. I also want to thank my sister, Grace, for so many fun times, cakes, and for putting up with so much nonsense. Finally, I want to thank my Grandmother, Alice, and all of the rest of my big-ish, mad-ish family for all the love a Grandson, Nephew, and Cousin could need. This thesis is dedicated to all of you.

Since this is the final bit of my 'formal education', I also want to thank some teachers without whom I may not be here. My thanks go to Erika Klingler, whose friendship was so important throughout my time at school, and who gave me an invaluable sense of self-confidence. They also go to Velmar McGregor, who fostered what was to become my interest in philosophy, but which at the time I can't believe was anything but a nuisance, and Thomas Adamu, who saw me through Kingsdale safe and sound. I also need to thank, in the deepest way, Tim Kearns, who told me, with a characteristically knowing look, that 'you'll enjoy philosophy boy, it'll be great'. Not only did he teach me how to read philosophy carefully, he knew better than I what I would get out of it. Without his initial nudging suggestion, I don't think I would ever have gotten to this point. Thanks, Kearnsy. I also want to thank Angus Grogono, whose enthusiastic, ingenious, and insightful teaching was the best, the *best*, introduction to philosophy that anyone could possibly hope for. It can be a hard discipline to orientate yourself in, and he did such a good job of helping me see the wood for the trees. Finally, I couldn't thank Mr. Grog without thanking Yasmin Hylton. It was immense fun discovering philosophy together, getting our heads around art, God, politics and chaos theory. Philosophy is best done with friends, and I was lucky enough to learn that very early on with her.

At the University of York, I had a number of amazing lecturers and seminar tutors who I learned so much from. Special thanks go to Wilson Underkuffler, whose energy was infectious, and Keith Allen who was a great teacher, and has been supportive even since leaving York. Dorothea Debus was a constant throughout my degree, supervising a large amount of my work, and providing me with invaluable emotional support and intellectual stimulation. I am happy to say that I am still thinking about the issues I worked on with her, and that they are liberally scattered throughout this thesis. Finally, I want to thank David Efird, who sadly died last year. David taught me metaphysics,

and his Metaphysics of Christianity module is probably the best course I have ever taken. More, he taught me how to hold two incompatible thoughts in my head at once without falling over, something I benefited from many times whilst writing this thesis.

During my PhD I have had the honour of being supervised by two philosophers I greatly admired during my undergraduate degree, and who I very much still do. Komarine Romdenh-Romluc has been a solid, consistent and supportive supervisor throughout. She has guided me through my PhD brilliantly, especially when I have been struggling to find a focus, a voice, or even a project worth working on. She has waded (and I mean *waded*) through material that stretches all the way from the metaphysics of facts to the sociology of gender, as well as almost everything in between. As someone prone to over-complications, I have found her clear-headed and common-sense approach vital, and it has saved me from myself many a time. More, an awful lot of the work I have done in this thesis has been driven by a desire to avoid a puzzled look on Komarine's face. So if any of this thesis is clear at all, that is why.

I have also had the distinct pleasure of being supervised by Helen Steward, whose work is still some of my favourite philosophy out there. She has taught me to see how the tiniest details in the formulation of an argument or a view connect up with the big picture, and to see which bits of a picture *matter*. Helen's ability to home in on what matters has been both inspiring and crucial, and her comments have often radically shaped the thesis even up until our final supervision together where she told me to get rid of 'all that quasi-formal stuff'. This turned out to be perfect advice, which improved the thesis immeasurably. Our conversations have always been enjoyable and I have learned so much from her. So I want to thank both Komarine and Helen for all their brilliant work, without which this thesis wouldn't be half as good.

When I got to Sheffield, I was lucky to find many brilliant friends who have shaped my time here for the better. I want to thank James Gagen, Kevin Lacourse, Rebecca Straughan, Cathy Mason, Brendan Kelters, Laura Towers, Patrick Connolly, and Radivoj Stupar for being fun and stimulating people to do an MA with. Brendan, Cathy, Radivoj and Patrick I owe even more, not only for their continued friendship, but also the many conversations (both philosophical and normal) and drinks (both caffeinated and alcoholic) which have animated the last few years.

Indeed, I have benefited from countless conversations over equally countless drinks with so many friends, and I can't do better than name them here: Adriana Clavel-Vázquez, Ashley Pennington, Alice Murphy, Anna Klieber, Alex Duval, Ahmad Fattah, Brendan Kelters, Barney Riggs, Cristina Roadevin, Charlie Crerar, David Strohmaier, Dan Herbert, Giulia Casini, Graham Bex-Priestly, Henry Roe, Isela González Vázquez, Jack Herbert, Jingbo Hu, James Lloyd, Josh Black, Kayleigh Dougherty, Keira Chapman, Lewis Brooks, María José Pietrini, Nadia Mehdi, Neri Marsili, Richard Hassall, Siobhan Moriarty, Simon Barker, Trystan Goetze, Viktoriia Kononova, Yuhan (Felicity) Fu. It has been a fantastic community, and I have learned and grown so much by being a part of it.

But some people deserve a special thanks. I need to thank Emma Bolton and James Lewis for being such good friends, for being such brilliant and fun philosophers to chat to about issues all across the board, and for always being such a good laugh. It was great being in *Bob Stern Present James Lewis and the Popular Misconceptions*, drinking tea, playing music, and eating Bilash at 2am. It'll never get better than that. Special thanks go to Robbie Morgan and Rosa Vince, who have been constants for me the whole time I've been in Sheffield, and who have always been up for some day-

drinking, meme-making fun. And thank you to Will Morgan and Matt Cull for always being around for a puzzling chats about weird corners of metaphysics. I have also been lucky to have had some of the very best friends and comrades over the last few years. One way or another, I think I may have spent more time with Catherine McAndrew, Guy Cowman-Sharp, Joe Diviney, Roan James, and Tanju Cakar than almost anyone else! With countless meetings, pints and picket lines, I can't say they've helped me get my thesis done, but I can say that it would have been much more boring without them.

I also, of course, must thank many of the staff at the department. The department has always been both very friendly and a hot-bed of enthusiastic philosophy, so it has been a wonderful place to learn how to do philosophy. I want to thank: Bob Stern, Chris Bennett, Dom Gregory, Eric Olson, Jenny Saul, Josh Forstenzer, Jimmy Lenman, Jules Holroyd, Megan Blomfeld, Paul Faulkner, Steve Makin, and Steve Laurence. Special thanks go to Luca Barlassina for a lot of teaching advice, help with papers, and for conversations which have made me that bit more naturalistically inclined. Thanks also to Niall Connolly for lots of great conversations and King Cobras, both of which I have benefited greatly from. Special thanks also go to Yonatan Shemmer whose door has always been open whenever I haven't known something about practical reason, and also for kindly agreeing to be my internal examiner, which means the viva will be a lively affair I am sure! Also, thank you to the administrative staff – Kate Atkinson, Joanne Renshaw, Sally Weston, Siri Romare, and Anne-Marie Frisby – for making things work so smoothly, and putting up with my tardiness.

Finally, my thanks go to everyone who has either read any part of this thesis and given me comments on it, or read anything that was *supposed* to be a part of the thesis but never made it: Kayleigh Dougherty, James Laing, Patrick Connolly, Jingbo Hu, as well as David Olbrich, Parisa Moosavi, Aaron Henry who have all commented on my work at conferences. In particular, I need to thank Jingbo Hu for reading nigh-on the whole thesis and giving me many deeply interesting comments. I have tried to take everything on board, however I am sure to have missed things.

I have been lucky enough to have had my PhD funded by the White Rose College of the Arts and Humanities, and I am grateful for their financial support throughout my PhD. I also had the pleasure of working with Clare Mac Cumhaill and Rachael Wiseman for the Women in Parenthesis project in Durham for a month, which was a refreshing break from my usual work, and a fascinating foray into the history of philosophy. Special thanks go to Amber Donovan and Sascha Lawson-Frost for making that such a good occasion, and organising the most stimulating reading group I have ever been to.

I have also been lucky enough to present my work at a whole host of different conferences, so I want to thank anyone and everyone who heard me speak or asked questions at: 'Reason and Agency' – the 16th Annual Graduate Conference at the University of Toronto; 'Empathy, Direct Perception and Other Minds' at the Centre for Subjectivity; the 2018 Aristotelian Society Open Session; the Northern Phenomenology Network; the Intersubjectivity Club workshop at UCL; the 2019 Aristotelian Society Postgraduate Session; the European Society of Philosophy and Psychology; and no less than seven Graduate Seminars at the Sheffield Philosophy Department. All of these events have been very helpful and stimulating. Of particular importance for me was a conference I did not speak at: the British Society for the History of Philosophy's annual conference in 2018 on the topic of habit. I think I thought I knew about habit before going to this conference, but I saw a great number of talks which proved to me I was wrong. I want to single out Tyler Haddow's marvellous

paper on habituation, which he allowed me to read a draft of, and which has had immeasurable impact on my thinking since then. Indeed, it was Tyler who first suggested to me that a habit-bearer's familiarity with courses of action might play a central role in a theory of habit, and that has been a guiding light for me ever since, becoming the substantive view I defend in Chapter's 6 and 7. So I want to thank him for spurring me to develop this view, in hopefully a somewhat original direction.

Finally finally, I have to thank Andrea Blomkvist. I don't think I can express just how supportive, silly, caring and fun she has been the whole time we've been together. Writing and finishing a thesis is tricky, and doing it in a year like this is even harder. I can't imagine what it would have been like doing it without Andrea, and I can't believe how lucky I am to have had her around to read, discuss, and support me through pretty much every idea in this thesis. She has taught me so much, and put up with so much talk of reasons and habits – I just cannot thank her enough. Her humour and friendship remains really essential to me. So, finally, thank you too.

ABSTRACT

What contribution do habits make to the explanations of what habit-bearers do when they act habitually? More specifically, do they make a *rational* contribution by shaping the landscape of the habit-bearer's reasons? Many think the answer must be 'no', since habits are mere behavioral dispositions which manifest in automatic, mindless behavior. In this thesis, I argue that this is false. Habits make a rational contribution by figuring in what I call 'broad rationalising explanations' when they manifest. That is, they have some rational bearing on the reasons for which a habit-bearer acts when they act habitually. I call this 'the Rationalising View'. Further, I argue that habits do this because they are tendencies to do things for the reason *that doing this is familiar*, where this reason is unavailable to people without the habit-bearer's history of action. Therefore, I elucidate the rational role and metaphysical nature of habit by connecting it to the notion of familiarity, and showing the theoretical benefits of doing so.

Chapter 1 argues against Gilbert Ryle's view of habit as an internally simple disposition to do things mindlessly. In Chapter 2, I argue for a framework for thinking about rationalising explanations within which we can assess arguments for and against the Rationalising View, and I outline my argument for that thesis. Because the argument requires some claims about intention, in Chapter 3 I consider the connection between reasons and intention. I argue for the Rationalising View in Chapter 4, and defend it from objections in Chapter 5. In Chapter 6, I reject some models of the Rationalising View in favour of one which centres on familiarity. In Chapter 7, I argue that this model does explanatory work in the theory of habit, particularly in giving us an account of the force of habit.

TABLE OF CONTENTS

INTRODUCTION

Contemporary philosophers of action have little interest in, and apparently no use for, the notion of habit. This may be surprising for a number of reasons. Firstly, the philosophy of action is, on the face of it, concerned with action and agency in all of its complexity. Therefore, to get a full picture of agency, we must not be limited to basic notions like belief, desire, and bodily movement. Rather, we must deal with the veritable smörgåsbord of puzzling things we find when we think about action, from akrasia and addiction, to deliberation and indecision, to skill and sociality. By and large, philosophers of action do concern themselves with this huge variety of agential phenomena. It is odd, therefore, that habits have been so neglected.[1]

The surprise is deepened when we remember that a vast swathe of the literature is devoted to questions about action-explanation: are actions caused, and if so, how and by what; what are reasons for action; how do we explain actions; and what are the proprietary antecedents of intentional actions? Given this interest in action-explanation, it is strange that habit – one of the most ordinary and commonplace factors in everyday action-explanations – simply goes missing from discussions.

Another reason it may be surprising is that straightforward analytic philosophy seems to be the only tradition in which habits get so royally side-lined. Phenomenologists take them to be of immense importance for understanding agency and perception, as do the classical pragmatists.[2] David Hume (1777) thought habit and custom central to understanding cognition and society, and Aristotle (2002) seemed to think habit essential to learning the virtues. So it is somewhat odd that analytic philosophers of action so strongly demur from the historical trend.

I think the reason is fairly simple: contemporary philosophers have, in general, a very low opinion of habit. They reveal this attitude in the casual, dismissive ways they talk of "mechanical

[1] There is a small but growing recent literature on habit which is a sign that this is changing. See Delacroix (2017), Douskos (2017b, 2017a, 2018a, 2018b, 2019a, 2019b), Owens (2017), Peters (2014), Pollard (2003, 2005b, 2006a, 2006b, 2008), Small (forthcoming, 2020), and Romdenh-Romluc (2011, 2013).

[2] The *locus classicus* for phenomenological engagement with habit is Maurice Merleau-Ponty's *Phenomenology of Perception* (2012). For contemporary phenomenological discussions, see Ingerslav (2020), Magrì (2018), Moran (2011), Romdenh-Romluc (2011, 2013), and Sachs (2014). All of the classical pragmatists are deeply interested in habit (Dewey, 2007; James, 2000; Lekan, 2007; Levine, 2015b; West & Anderson, 2016). This is not to say that the phenomenologists and pragmatists got habits right – I think they were often rather misguided. The point is that they take habits *seriously* in accounting for aspects of agency.

habit", "brute habits", "blind habit", "sheer habit" and "mere habits".[3] Robert Brandom speaks of "the thoughtless jostling of the habitual" (Brandom, 1994, p. 403), and John McDowell distinguishes virtuous action from "the outcome of a blind, non-rational habit or instinct, like the courageous behaviour – so called only by courtesy – of a lioness defending her cubs" (McDowell, 1979, p. 331). Harry Frankfurt says that some "patterns of interest or of response may be manifestations only of habits or of involuntary regularities of some other kind" in contrast to the things which a person *really* cares about (Frankfurt, 1982, p. 260). Manifestations of habit are presumably also often to be located, by those who follow Frankfurt in distinguishing between 'actions proper' and 'mere activities', in the category of mere activities. When one habitually drums one's fingers on the table, idly and inattentively, one is not performing an action; one is "only being active" (Frankfurt, 1977, p. 58).[4] Finally, and most comically, Mark Johnston even connects some habitual performances with cases of 'collapsed affect':

> "There you are teaching *Phil 287: The Philosophy of Sport* for the eighth time. Somewhere in the middle of the semester as you are lecturing you begin to hear the words coming out of your mouth as if you were a detached auditor overhearing remarks while waiting in a bus station. You are lecturing on automatic pilot in a way that allows your attention to drift elsewhere, eventually to alight upon your own performance. You find it devoid of value and of dis-value. You may even notice that the students are quite taken with your philosophical analysis of the Olympic ideal. You couldn't care less, one way or the other. Your habitual performance is hollow because however professional it might be, it is not prompted by your sense of the appeal of anything in it or connected with it." (Johnston, 2001b, pp. 192–193)

There is no indication that Johnston takes all habitual action to be like this. However, the way he connects habit to automaticity, lack of attention, and behaviour which is quite disconnected from anything one values or cares about is quite apt for the way that philosophers of action picture habits. They are forces within us which manifest in ways that bypass our rational capacities, values, and which may not even manifest in action. They are properties we share with ('mere') animals; they are blind to reason; they 'thoughtlessly jostle' us.

My project in this thesis is to dislodge this picture of habit. I think it is constructed through a series of assumptions and prejudices about habit, often unstated, which do not ultimately stand up to scrutiny. It is presumed, for example, that habitual actions are automatic, either in the sense that they are mindless because inattentive, or in the sense that they cannot also be the results of deliberation. It is presumed, often, that what is done habitually either cannot be done intentionally

[3] The quotes are from Bäckström & Gustafsson (2017, p. 42), Pettit (2004, p. 4), Lavin (2011, p. 370), McDowell (1979, p. 338), and Ryle (1970, p. 41) respectively. They are representative of the sorts of epithets that get attached to 'habit' very often.

[4] Distinctions of this sort are also accepted by David Velleman (2000) and Christine Korsgaard (2009).

or often is not done intentionally. Similarly, it is sometimes said that what is done habitually is not done for reasons. Whether these assumptions are true is rarely questioned.

In order to dislodge this picture, I want to try to answer some questions about the contribution habits make to action-explanations. To appreciate these questions, consider a small story. I have a habit of going to *Gilmour's Café* for an 11am coffee. When I go there, I do so for a number of reasons: the coffee is tasty; the staff are friendly; it's close and cosy. But when I go there out of habit, my habit of going there also figures in an explanation of why I do. My friend Alice has never been to *Gilmour's* before, but she has heard the coffee is good and the place is cosy, so one day she decides, for those reasons, to go there. Having never been before, though, Alice does not have a habit of going there which could help explain why she goes for the first time.

In the story, even if Alice and I go for our 11am coffees for the same reasons, there is a difference in the explanations: in my case, there is a habit at work. We can ask, though: does my habit make a contribution to the rationalising explanation of my going to *Gilmour's*? Does the fact that I went there out of habit, but Alice did not, mean I possess reasons to drink there which Alice does not; does it alter the landscape of my reasons; or does it have some other rational bearing on my having of, and acting for, reasons? Is it what we might call a 'rational difference-maker'?

Over the course of this thesis, I will argue that the answer to these questions is 'yes'. When habits manifest, they *do* add something to a full rationalising explanation of why the habit-bearer does what they do. They are not sorts of merely causal or dispositional forces which make their contribution outside the sphere of the agent's rationality. This, in rough outline, is what I call 'the Rationalising View of Habit', and it is the purpose of most of this thesis to defend it.

Now, some may think that the only items involved in rationalising explanations are *reasons*, the facts which favour doing something, the normative force of which is appreciated by agents who thereby act *for* those reasons. So what can be meant by the idea that habits – a sort of property – can contribute to a rationalising explanation? One of the central points I will make is that there are really two sorts of rationalisations: narrow and broad. A narrow rationalisation explains why someone *A*-s by stating their own reason for *A*-ing – the fact which, from their perspective, favours *A*-ing (Alvarez, 2010; Dancy, 2000). In contrast, I argue, a broad rationalisation is an explanation of why they *A* which states facts about the agent's perspective which make it intelligible that they *A*-ed for the reason they did. For example, the narrow rationalisation of why Sally ran from the bear is the fact that a bear was chasing her. The broad rationalisation includes the fact that Sally knew a bear was chasing her, that she did not want to be eaten, and that she feared the bear. The broad rationalisation has its rationalising power by showing how properties of Sally – her knowledge, desires, and fears – bear on the narrow rationalisation. When the facts which make up an explanation make essential reference to a property which is rationally relevant to a narrow rationalisation, I will say that the property *figures in a broad rationalisation*.

With this in hand, we can state the Rationalising View of Habit as the view that, whenever a habit manifests in someone's *A*-ing such that they *A* habitually, their habit figures in a broad rationalisation of why they *A*. Therefore, just as Sally's knowledge, desires and fears are rationally

relevant to her running from the bear, my habit is rationally relevant to my going to *Gilmour's*. So there *is* a difference in the broad rationalisations of my coffee-going and Alice's.

The Rationalising View of Habit I want to defend is in fact quite general. It says that habits, when they manifest, have *some sort of* rational bearing on one's having of, or acting for, reasons. It leaves it open what that bearing is. So if we decide to accept the Rationalising View, we may ask the further question of *how* habits figure in rationalising action-explanations. Are they like virtues in being sensitivities to independent reasons, like pains in grounding reasons and providing knowledge of them, or do they have an entirely distinctive explanatory profile?

I argue that, habits are, indeed, explanatorily distinctive, with a distinctive rationalising role. To accommodate this, I argue for the view that habits are tendencies to act for a certain reason, where habit-bearers would not have this reason if they lacked their habits. The reason in question is that doing the habitual in the usual context *is familiar*. Familiarity with going to *Gilmour's* grounds a reason to go there – that getting an 11am coffee at *Gilmour's* is familiar – and my habit of going there for coffee is a tendency to go there for that reason. So the way that habits figure in rationalising explanations is rather different from the ways other properties do: they figure by manifesting as the habit-bearer's response to a reason that is *idiosyncratic*, in that it depends on features of the agent, such as their own history of action. Further, I argue that these reasons are typically revealed in experience, in the 'feeling of familiarity'. This experience represents situations as being contexts in which one is familiar with the habitual thing, brushing one's teeth, interrupting people, biting one's nails, or putting on the left shoe first. The experience thereby provides one with knowledge of a reason for action, and it exerts a motivational force. That, I claim, is the basis for the force of habit. I call this 'the Familiarity View'.

If I am right, or on the right track, then the picture of habit which analytic philosophers of action implicitly work with is very wrong. Far from being forces alien to our rationality and subjectivity, habits are intimately connected to reasons for action and to our perspective on those reasons. Habits are not blind or brute or mechanical; habitual action is not the effect of thoughtless jostling. Acting out of habit is acting for a special sort of reason, where this reason is connected in special ways to certain psychological forces, and habits themselves are tendencies to act for those reasons.

The Rationalising View of Habit is a view about the role of habits in action-explanations, and it is a view that says that habits have a *rational* role in broad rationalisations of what someone does when they do it habitually. This issue should be distinguished from other issues about the relations between habit and reason. In particular, I will not be concerned, as William James (2000, pp. 73–78) was, with the question of whether it is rational to have habits, or what is good about having habits. It may very well be good to have habits because they reduce our cognitive load, even if, when one acts habitually, one's habit does not figure in a rationalising explanation of what one does.[5] It is compatible with habits being good for us that they still function as sorts of causal compulsions, albeit ones that we should be happy to have.

[5] This is one of way of reading Michael Bratman's remarks on habit in (Bratman, 2018, Chapter 7).

It will be useful to say something about the scope of my discussion. It is limited to what we may call 'habits of action', and I do not consider, say, habits of emotion or habits of perception. It is not that I deny that there are such habits – I think there may be. But I think that they are, firstly, perhaps less central to our understanding of habit. Secondly, more prosaically, I have found that I actually have something to say about habits of action, and that what little I want to say about habits of emotion and perception really derives from what I say about action. So I have found it better to focus on the things I have something to say about.

I also want to set aside what I would like to call 'habits of style' – habits whose manifestations are the distinctive, perceptible *ways* of acting which characterise the style with which one does things. Zadie Smith, in a beautiful comparison of Fred Astaire and Gene Kelly, says:

> "[Fred] was not actually that tall, he only appeared as if he were, and when moving always appeared elevated, to be skimming across whichever surface: the floor, the ceiling, an ice rink, a bandstand. Gene's centre of gravity was far lower: he bends his knees, he hunkers down. Kelly is grounded, firmly planted, where Astaire is untethered, free-floating." (Smith, 2018, p. 137)

One immediately knows what Smith is getting at here – each dancer's distinctive *style*. And it is equally obvious that we all carry with us styles and manners of doing things far more mundane than Broadway dancing. There's the way you brush your teeth, the way you speak when you're proud; the way you wear your hat, the way you sip your tea. It is very likely that there are habits of styles; that some of the ways which characterise the aesthetic features of our actions are acquired and sustained by habituation.

Unfortunately, I will not be discussing these habits, even though styles are so intimately connected to agency. The main reason is that it is very hard to say what styles *are*, and what relations they bear to better understood categories in action-theory, such as action, activity, and thing done. Not knowing quite what a style is makes it hard to say how to account for the habits which manifest in them.[6]

The fact that there are probably habits which fall outside my topic creates a small tension which I want to acknowledge. I present my arguments and views throughout the thesis as if they give us a totally unified theory of habit. However, this is very possibly not the case. For example, although the Rationalising View of Habit as I stated it above could apply to habits of emotion, perception, or style, the Familiarity View seems not to, since it has embedded within it the idea that habits are tendencies to do things *for reasons*. It is at least a difficult question whether the perceptual experiences which manifest one's habits of perception can be had 'for reasons', and some have argued that some emotions cannot be responses to reasons (Velleman, 1999). Further, my argument for the Rationalising View runs via the claim that everything done habitually is done intentionally. But this even harder to connect to emotions, experiences, and styles. When I get angry at the news

[6] My thanks to Kayleigh Doherty for sparking my fascination with these styles.

out of habit, I do some things intentionally (roll my eyes, swear) but *my getting angry* is not intentional. Styles, too, very often seem to slip beneath the radar of intention. And intention bears little relation to perception in this regard. So it is not clear that one could argue in the way I do for a Rationalising View of Habit which applies to habits of emotion, perception, and style.

I do not think this is disheartening or problematic. I am content that my arguments and conclusions, intended as they are to apply only to habits of action, capture the phenomenon fairly well. In fact, I have hope that there *is* a way of arguing for a fully general Rationalising View of Habit, and that the argument will not be too different to the considerations I give here. However, it must be said that this is not the purpose of this thesis, and that this is all I will say about habits of emotion, perception, and style.

So, my topic is the much more limited notion of a habit of action. However, the name is liable to mislead. It is not *actions* we are in the habit of 'doing'; rather, we are in the habit of doing certain *sorts* of things. I have habits of biting my nails, singing silly songs, and watching dogs play in the park. But these are types of things one can do, rather than the actions which are one's doings of those things. These types are repeatable: I can sing silly songs on Tuesday, and then again on Wednesday. They are also multiply instantiable: both my partner and I can be watching dogs play in the park at the same time, and thereby be doing the same thing. These things therefore bear the hallmarks of *properties*. Actions, on the other hand, are the events or processes which these are properties *of*. And events and processes are individuals, or particulars, of a kind that occur over time. They are unrepeatable: my singing of the song yesterday is not the same event as my singing of the song today. Actions are also not multiply instantiable: my watching of the dogs is numerically distinct from my partner's watching of the dogs.

This distinction between actions and the things one does was most famously articulated in these terms by Jennifer Hornsby in her excellent book *Actions* (1980), and it is of paramount importance to my thesis. Throughout, I will use locutions like 'when one does something habitually' and 'if one *A*-s out of habit' and these should be understood as claims about things done rather than actions; about *what* one does, rather than *one's doing* of those things. Whenever I use schematic, capitalised and italicised letters like '*A*' and '*B*', the letters stand in for things an agent (sometimes the anonymous *S*, sometimes Sally or Alice) does, did, or may do. I will not often have use for talk of actions, but when I do, I will name the individuals in question with Greek letters like 'ϕ' and 'ψ'. Now, granted, Hornsby's distinction is not universally accepted. There is a live debate about whether actions should be individuated coarsely or finely (Cohen & Hornsby, 1982; Goldman, 1971; Hornsby, 1981; Sandis, 2015; Thalberg, 1971; Thomson, 1979). That granted, I have little to add to Hornsby's arguments, so her view stands as an assumption in my thesis.

Hornsby's distinction is the first of three assumptions about actions which I do not have the space to defend but which will sometimes appear in my thesis. The second assumption is the thesis that if one *A*-s by *B*-ing, then there is an event, ϕ, which is one's *A*-ing and one's *B*-ing. If I switch on the light by pressing the switch, then there is an event, ϕ, and ϕ is my pressing of the switch and my switching on of the light. The thesis is associated with G.E.M Anscombe (1979, 2000, pp. 45–46), Donald Davidson (Donald Davidson, 1980a), and Jennifer Hornsby (1980, pp. 6–10). Again, there is

a live debate about this,[7] however, I do not have the space to argue for it here so it will stand as an assumption. It will do little argumentative work in my thesis, although it will quite often affect my precise formulations of points and arguments.

The third assumption is that actions are either causes or causings which are distinct from what they cause or are the causings of. For example, if I raise my arm, then my arm rises. The first is something I do; the second is something that happens as a result of what I do. Instantiating these two properties are two distinct occurrences: my raising of my arm, and my arm's rising. My assumption is that these *are* distinct occurrences and that the first either caus*es* the second, or is the caus*ing* of the second. My arm's rising is the causal result of my raising of my arm. Following David-Hillel Ruben's useful (if un-pretty) nomenclature, we may call this view 'two-particularism about actions', since it says that actions involve two occurrences standing in a causal relation (Ruben, 2018b). Once again, the view is associated with Hornsby who argues that the two occurrences are events which stand in a causal relation (1980, pp. 1–46). In contrast, Maria Alvarez and John Hyman argue that the action is the causing of an event (Alvarez & Hyman, 1998). I will not choose between these ways of cashing out this 'two-particularism', but I do assume that the view itself holds.[8] Once again, this view does little in my argument, however I do make use of it in Chapter 3, Section 3.

The final thing I want to say is about the nature of habits, rather than actions. Throughout, I effectively assume that habits are properties which are involved, somehow, in explanations. Some philosophers, Bill Pollard for example, have said that habits are rather *patterns of behaviour* which extend over time (Pollard, 2005a, p. 73, 2006a). Whatever patterns are, if they are supposed to extend through time, they cannot be properties, since properties are not spatio-temporal items.

But I do not think that habits are patterns of behaviour. I think that gets things exactly the wrong way around for a few reasons. Firstly, habits get invoked in *explaining* patterns of behaviour; the regularities we find in what people do are sometimes the manifestations of habit. But this is hard to understand if the habits *just are* the patterns. Secondly, the notion of a pattern is probably too thin to do the explanatory work we need. Compare the Humean view of laws defended by David Lewis (Lewis, 1981, 1994), on which laws of nature do not govern causal interactions; rather, they describe them. That view may be true, but it does rule laws out of explaining *why* the planets move as they do, and *why* rocks roll down hills. The laws just describe the patterns of correlation between things, and the correlations may, in time, fail to hold. The laws do not explain why the correlations hold as they do. Pollard's view of habits as patterns is similar. Patterns, like correlations, are cheap: they are easy to come by, and have no particular explanatory weight. But whilst it may be possible to agree with Lewis that laws need not explain or constrain, I think it is quite intolerable as a conception of habit. Habits *do* figure in explanations of why people walk to work by certain routes, or drum on the table when they are bored. So we cannot think of habits as patterns of behaviour if they are to explain those patterns. Rather, I think we had better see them as properties which figure

[7] See Sandis (2010) for discussion.

[8] It has recently come under attack in Ford (2014), Haddock (2005), Ruben (2018b, 2018a), and Small (2016). Helen Steward, in her (2018) review of Ruben's (2018b) book provides reasons to doubt the cogency of views which deny 'two-particularism'.

in explanations by manifesting in their bearer's doing something. At least, I think I have given reasons to think this assumption is a good one.

***

My thesis has seven chapters. In Chapter 1, I discuss Gilbert Ryle's (1949/1970) conception of habit as a 'single-track' disposition to do things which do not express the habit-bearer's intelligence. I single out Ryle for a few reasons, primarily because he is one of the few philosophers who actually articulates and defends the conception of habits which I showed is so prevalent at the outset. I also think that Ryle's views have been influential, occurring, as they do, in that most influential Chapter 2 of *The Concept of Mind*. The point of this chapter is to argue that Ryle is wrong in thinking that habits are mechanistic dispositions which produce unintelligent behaviour in such a way that contrasts with skills. Doing so allows me to set aside a number of prejudices about habit, which allows me to pursue my argument for the Rationalising View of Habit in the rest of my thesis.

In Chapter 2, I introduce the question of the Rationalising View of Habit and try to provide a framework within which we can state it and understand how to argue for (or against) it. The framework is a conception of rationalising explanations. I argue that the reasons for which we act are facts rather than mental states, and that, in the first instance, rationalising explanations are explanations which simply state the fact which motivated an agent to do what they did. However, I also argue that this is only one, narrow, sort of rationalisation. There is also a broader kind, which states facts about the agent's perspective which make it intelligible that they acted for the reason they did. Further, I argue that aspects of the agent's perspective – their beliefs, knowledge, desires, and fears – *figure in* the broad rationalisations when they are ineliminable from a full explanation of what they did. On the basis of this framework, I state the Rationalising View of Habit as the view that, when a habit manifests in *S*'s *A*-ing, the habit figures in a broad rationalisation of why *S A*-s. At the end of the chapter, I outline the following sketch of my argument for the Rationalising Conception:

1. Habits are mental properties (broadly construed);
2. If mental properties figure in explanations of why someone does something intentionally, then they figure in broad rationalisations of why they do that thing;
3. Whenever habits figure in explanations by manifesting, they figure in explanations of why the habit bearer does something intentionally.

C) Therefore, whenever habits figure in explanations by manifesting, they figure in broad rationalisations of why the habit-bearer does that thing.

I defend Premise (1) in Chapter 2, but defending the next two premises take up the next three chapters. In Chapter 3, since the argument runs via some claims about intention, I focus on two issues: firstly, what it is to do something intentionally; secondly, whether one does something intentionally if and only if one does it for a reason. About the first issue, I defend a broadly

Anscombean conception of intention in terms of non-observational practical knowledge of what one is doing. I motivate the view, and then defend it from two significant objections: that what is done intentionally can outrun our knowledge of what we are doing, and what is done intentionally outruns what can be known without observation. I then defend the biconditional connecting acting for reasons with acting intentionally and argue that, if it is true, we can use it to defend Premise (2) above.

Chapters 4 and 5 both defend Premise (3). In Chapter 4, I develop a positive argument for the claim that everything done habitually is done intentionally, and I show how this easily entails Premise (3). My argument employs a fairly complicated strategy. The idea is that habits are partially individuated by what they are habits of doing, but we cannot say how a habit is individuated unless we say what the habit-bearer must do *intentionally* if the habit fully manifests. I argue this entails that everything done habitually is done intentionally. Premise (3) then follows quickly: whenever a habit manifests, what it manifests in is something the agent does intentionally. This chapter concludes my positive case for the Rationalising View of Habit.

In Chapter 5, I defend the controversial conclusion of the previous chapter from three powerful objections. Firstly, that there are cases of 'habit-slips', where agents act against their standing intentions because their habit kicks in (Romdenh-Romluc, 2013). Secondly, that sometimes the explanatory demands of sociology (say) require that what we do habitually is done unintentionally. Finally, that there are sub-intentional actions, and that some of those actions are also manifestations of habit (O'Shaughnessy, 1980; Steward, 2009). Having fended off those objections, this chapter concludes my positive case for the Rationalising View of Habit.

In Chapter 6, I turn my attention the question of *how* habits figure in broad rationalisations. I consider a number of models which may suggest themselves, and reject them. In doing so, I develop five constraints on a plausible model of the Rationalising View, and I suggest that the Familiarity View I described earlier fits them well. To do this, I spend some time developing a conception of familiarity, the feeling of familiarity, and their relations to reasons for action.

The argument in Chapter 6 is not conclusive, and is not supposed to be. However, I think there is another argument for the Familiarity View. Therefore, in Chapter 7, I argue for the Familiarity View by showing that it can explain the fact that habits are developed by repetition; that they make us relatively insensitive to alternative courses of action; and that there is a *force* of habit. A popular alternative explanation is given in terms of the apparent 'automaticity' of habitual action, the fact that what is done habitually is not the result of deliberating. I argue this fails because acting habitually and acting from deliberation do not rule each other out. I use this as a springboard for my alternative explanation in terms of familiarity. I end with a very brief conclusion, summarising my arguments and conclusions.

CHAPTER 1

HABITS, SKILLS, DISPOSITIONS, AND INTELLIGENCE

I want to begin my thesis with an exploration of some important themes in contemporary discussions of habits which range over questions about the 'mindedness' of habitual actions, the relations between habit and skill, and the metaphysics of habit. Although these are broad themes which cover much philosophical ground, there are important connections between them. For example, if one thinks that habits are sorts of behavioural dispositions which can be manifested in only one way, then they may seem like mere routines which, lacking any variation, could not manifest in intelligent or rational action. This view of their metaphysics therefore encourages a certain view of their relation to agency and rationality.

My overarching goal in this thesis is to provide an account of habits whose core is the view that habits figure in rationalising explanations of why we do things, when we do them habitually. This view is rather different from the account of habits on which they are mindless dispositions, an account which seems to hold quite an attraction for philosophers, for example those I mentioned in my introduction. Given the prevalence of that view, as well as the prevalence of the somewhat prejudicial ideas about habit on which it is based, it is important to show that it is mistaken for my project to get off the ground. This chapter's role is therefore to clear the ground for me to develop my view of habit by showing that the traditional way of conceiving of habit's nature gets things very wrong.

I will spend most of this chapter addressing Gilbert Ryle's arguments concerning habits in *The Concept of Mind* (1949/1970), mostly because they are fairly well developed and embody the approach which I aim to tackle. However, I also focus on Ryle because of the historical role he has played in affecting the way that philosophers have come to think about habits. In taking on Ryle, I hope to therefore undermine some of the historical inertia that his theory of habits has generated.

I want to approach these issues by assessing a way of contrasting habits with skills which Ryle proposed and whose influence is displayed in the work of, for example, Victoria McGeer (2018) and Stina Bäckström and Martin Gustafsson (2017). Somewhat independently, the contrast has also been made by Julia Annas in her discussion of practical expertise in *Intelligent Virtue* (2011) and her paper 'Practical Expertise' (2012). The contrast promotes an image which favours skill over habit in allowing that skilled actions display an agent's intelligence, whereas habitual ones do not. This

contrast is underpinned by a metaphysical claim, that skills and habits are both sorts of behavioural dispositions which differ only in complexity. Skills, they think, are very complex dispositions; habits, in contrast, are very simple ones.

However, the contrast between habits and skills is sustained, in this picture, by the assumption that they also share important features. In particular that, despite their differing complexities, both are dispositions which contribute to action-explanations in just the same way. These background assumptions are just as important for the picture as the contrasts are.

I will argue that this picture is, in all respects, mistaken. Habitual actions do display intelligence; habits are internally complex; habits and skills figure very differently in action-explanations; and, whatever dispositions are, habits and skills are not both dispositions.

In Section 1, I explain and motivate Ryle's views, and break them down into the aspects he thinks distinguishes skills from habits (complexity and mindedness), and the aspects he thinks they share (dispositionality and explanatory role). Then, I argue that habits are internally complex (Section 2) and that habitual actions cannot be ruled mindless (Section 3). In Section 4, I argue that if we leave things there, feeling that we have vindicated habits, we will have simply collapsed the categories of habit and skill. But we must avoid this, since habits and skills perform very different explanatory roles (4.1) and they are very different sorts of modal properties (4.2). My aim is that, by the end of this chapter, very little, if anything, should be left of Ryle's picture.

Section 1 – Dumb Habits and Intelligent Capacities

The general project of Ryle's *Concept of Mind* is to argue against a kind of Cartesian view on which the mind is fundamentally 'internal', unobservable, and somehow imputes intelligence to 'external' bodily behaviours. This view is most famously attacked in Chapter 2, called 'Knowing How and Knowing That', which is concerned to argue against the view that bodily actions are only imbued with *intelligence* by being suitably related to theoretical operations such as the consideration of propositions. This Cartesian view makes "the absurd assumption [...] that a performance of any sort inherits all its title to intelligence from some anterior internal operation of planning what to do" (Ryle, 1970, p. 32). This means that bodily behaviours are only derivatively mental, and they derive their mentality from some internal propositional theoretical states and processes.

Ryle casts his discussion in terms of the intelligence of behaviours. However, Ryle is not interested in the ordinary, narrow sense of 'intelligence'. When Ryle talks about the intrinsic intelligence of a behaviour, he means that that behaviour is itself a manifestation of the person's mindedness. This is why his list of intelligence predicates is so extensive. He is interested in behaviours which are clever, sensible, careful, methodical, inventive, observant, quick-witted, cunning, scrupulous, stupid, dull, silly, humourless, unmethodical, uncritical, dense, rash, unwise, and uninventive (Ryle, 1970, p. 26). Whilst it may sound strange to say that 'stupid' is an intelligence predicate, Ryle's point is that someone can manifest their intelligence by doing something stupid, like looking at their phone when driving. Therefore, manifesting intelligence is a matter of

manifesting one's 'qualities of mind' in one's behaviour, qualities of mind which need not be desirable.

Ryle's objection to the Cartesian view is his famous regress argument, but I will not discuss it here.[9] His own view is that we replace Cartesianism with a view on which a person's bodily actions are intrinsically intelligent in virtue of their being exercises of intrinsically intelligent capacities or skills. This means that, for Ryle, there are two questions to get straight: firstly, what is a capacity; secondly, what is intelligence? The first question was, for Ryle, fairly simple to answer: a capacity is a kind of disposition:

> "When we describe a glass as brittle, or sugar as soluble, we are using dispositional concepts, the logical force of which is this. The brittleness of a glass does not consist in the fact that it is at any given moment actually being shivered. It may be brittle without ever being shivered. To say that it is brittle is to say that if it ever is, or ever had been, struck or strained, it would fly, or have flown, into fragments." (Ryle, 1970, p. 43)

Ryle has an awful lot to say about the logic of dispositional concepts, but some of the central points are contained in this passage. Firstly, a disposition is a *modal property* – to say that some actual object has a disposition is not to say what it is doing but to say what it *would*, *could*, or *might* do. Dispositions are characterised by an existential independence from their manifestations – a glass that never breaks is still fragile; a rubber band that is never stretched is still elastic. Secondly, the way that Ryle analyses the modal nature of dispositions is by means of a subjunctive conditional with the form of '*if* (stimulus condition), *then* (manifestation of disposition)'. For example, 'if the vase were dropped, then (all things being equal) it would shatter' is one of the conditionals which characterises the vase's fragility.[10]

Thirdly, Ryle makes it clear that he thinks of skills and ordinary dispositions like fragility and solubility as members of a single metaphysical kind: "[skills] are certainly second natures or acquired dispositions [...]" (Ryle, 1970, p. 41). Therefore, skills are also analysed in terms of a set of subjunctive conditionals.

> "What is involved in our descriptions of people as knowing how to make and appreciate jokes, to talk grammatically, to play chess, to fish, or to argue? Part of what is meant is that, when they perform these operations, they tend to perform them well, i.e. correctly or efficiently or successfully" (Ryle, 1970, p. 29)

[9] Particularly useful discussions of this are Hornsby (2012) and Small (2017b).

[10] The *ceteris paribus* condition is to exclude cases where conditions might stop a disposition being manifested on some occasion once the stimulus condition has been applied (Bird, 1998; Lewis, 1997). These complicated problems do not concern me here, so I will generally omit the clause.

So, to say that someone is a skilled jokester is to say something (very roughly) like this: if they were to tell a joke, then (all things being equal) their audience would laugh.

Fourthly, Ryle makes a distinction between two types of disposition: single-track and multi-track dispositions. Single-track dispositions are those which can be characterised by a single conditional which pairs one type of stimulus condition with one type of manifestation. A multi-track disposition is one which must be described by more than one such conditional.

If Ryle thinks of skills as dispositions whose modal nature is just like that of fragility and solubility, something must account for why dispositions like the skill of tennis playing, or skiing, or French-speaking are intelligent and those like fragility and solubility are not. This brings us to the second question: what is intelligence?

As I noted earlier, Ryle thinks that an activity is intelligent when mental conduct predicates apply to it. And he thinks that an activity is intrinsically fit for those intelligence predicates when the agent self-consciously applies standards, rules, or norms of the activity *in and by acting*. For example, a clown and I might fall over in just the same way with respect to our spatio-temporal trajectories. Since I simply trip, my falling does not display my intelligence because I am not conducting myself in accordance with any standards. The clown, however, is self-consciously trying to meet the standards of safety and comedy. This self-conscious care makes the clown's falling intelligent in that it is constitutive of her conducting herself in ways that she intends to. This indicates that an intelligent *capacity* must be a disposition to behave in ways which involve self-conscious adherence to the standards of the activity. Since Ryle's sense of 'intelligence' is broad enough to cover capacities of agents whose manifestations are 'illogical', 'unobservant' and 'unwise', and so which we would not deign to call skills, we can say that skills are, for Ryle, basically the intelligent capacities to succeed at meeting the standards of the activity one has a capacity to engage in.[11]

Finally, it is important to Ryle's view of skills and intelligent capacities that they be multi-track dispositions (Ryle, 1970, pp. 42–46). His thought is that skills are paradigmatically flexible capacities which allow agents to respond in varied ways to different situations. For example, a skilled climber can climb many rock-faces and could climb some rock-faces in different ways. A less skilled climber has fewer options open to them. And someone who in exactly one condition could climb a rock-face in exactly one way we are not liable to call skilled at all – theirs is a mechanical disposition. So, Ryle's thoughts about the involvement of care and attention in skilled activity are underpinned by an aspect of his metaphysics of skill.

1.1 – Ryle's Contrast

Now that we have Ryle's conception of skills, we can assess his contrast between skills and habits. Firstly, Ryle claims that skills are exercised with care and attention whereas habits are not: "When we describe someone as doing something by pure or blind habit, we mean that he does it automatically and without having to mind what he is doing" (Ryle, 1970, p. 42). An agent exercising a skill, for example "a mountaineer walking over ice-covered rocks in the high wind [...] thinks what

[11] Ryle effectively says as much in the previous passage from (Ryle, 1970, p. 29).

he is doing" (ibid.). Ryle's thought is that habitual actions do not involve the agent's self-consciously trying to satisfy the standards of their activity, a heed which characterises skilled action. He thinks that the relative automaticity of habits rules out this care, since when an agent acts habitually he may "be quite unaware that he has done it" (Ryle, 1970, p. 106). Ryle's thought is that habitual action is characteristically absentminded and so cannot be characterised by the self-consciousness requisite for intelligent action. This receives some support from the fact that one can fail to do what one planned because one acts habitually instead, for example when intending to cycle to the park, but absentmindedly cycling to work out of habit.[12] In acting from habit, the habit-bearer is seemingly running on rail-road tracks rather than following rules.

Secondly, skills are flexible multi-track dispositions – they allow us to engage with different situations in a variety of ways, and the same skill can manifest in many different activities. My guitar-playing skill can be exercised on stage or at home; under pressure or whilst relaxed. And it can also manifest in many different ways, for example, in playing this tune or that. In contrast, Ryle thinks that habits are single-track dispositions. He says that "to be a smoker is just to be bound or likely to fill, light, and draw on a pipe in such and such conditions [...] the actualisations of which are nearly uniform" (Ryle, 1970, p. 43). If habitual actions are 'nearly uniform', and we can specify the conditions under which one smokes in narrow terms, then this suggests that we can describe habits with simple, unitary, conditionals. But then they are not like skills, since they lack the internal complexity and flexibility characteristic of skill. Since that complexity is the metaphysical basis for intelligent performances, and since habits are single-track dispositions, habitual actions are not intelligent.

The final contrast between habits and skills is somewhat secondary for Ryle. He argues that habits are acquired by drill whereas skills are learned by training. Ryle casts this distinction in terms of both the feedback the learner gets from the teacher and of what counts as success. "[Training] involves the stimulation by criticism and the example of the pupil's own judgement. He learns how to do things thinking what he is doing, so that every operation performed is itself a new lesson to him how to perform better" (Ryle, 1970, p. 42). This kind of feedback, feedback that stimulates thoughts in the learner about how and why they are being taught to act the way they are, is very different from the kind of feedback gained from, say, an army drill. There, the feedback involves someone shouting at you until you salute correctly in the right circumstances. Relatedly, what counts as success is different in training and in drilling. In training, success is measured at least partly by whether the learner understands *why* they are supposed to do certain things, so that they can take it upon themselves to improve according to the standards of the activity. In drilling however, success is measured simply by whether one has formed the habit.[13]

For Ryle, these contrasts are of paramount importance for thinking about the nature of intelligence. They tell us that he thinks an intelligent capacity's exercise must involve self-

[12] See Romdenh-Romluc (2013). I discuss habit slips in more detail in Chapter 5.

[13] For excellent discussions of the role of the distinction between drilling and training in Ryle's general philosophical outlook, see Bäckström & Gustafsson (2017). For a very interesting extension of Ryle's discussion to problems in free will and moral responsibility, see McGeer (2018).

consciousness, care and attention; it must involve training and an understanding of why the capacity has the standards it does; and it must be flexible and adaptable to a variety of different situations. Insofar as habitual actions fail to meet these conditions, they are not displays of agents' intelligence; insofar as habits produce such actions, they are not intelligent dispositions. Ryle sums up his position thusly:

> "When we say that someone acts in a certain way from sheer force of habit, part of what we have in mind is this, that in similar circumstances he always acts in just this way; that he acts in this way whether or not he is attending to what he is doing; that he is not exercising care or trying to correct or improve his performance; and that he may, after the act is over, be quite unaware that he has done it." (Ryle, 1970, p. 106)

For Ryle, then, there are deep differences between habits, skills and their relative manifestations in action. However, it is important for my discussion that we see a certain structure in the *way* the contrast is drawn and what it presupposes. For, whilst Ryle contrasts habit with skill, he sees the contrast as being a contrast *within a common kind*: learned behavioural dispositions which explain their manifestations in fundamentally the same way.

This is brought out in the fact that, given what has been said about the contrasts, habits and skills must compete for explanatory space in action-explanations. This is because, since habits are unintelligent dispositions to do things unintelligently, then habitual actions which they figure in explanations of must also be unintelligent. Complimentarily, intelligent actions must be explained by intelligent capacities or skills, and therefore cannot be explained by habit. This means habits and skills can never figure in the explanation of the same action – they compete for explanatory space. But if they compete for explanatory space, then this can only be because Ryle conceives them as having the same explanatory *role* in action-explanation.[14] After all, items with different explanatory roles can generally co-exist in explanations.

Therefore, for all their differences, on Ryle's view, habit and skill share both a metaphysical structure as behavioural dispositions *and* an explanatory role in the aetiology of action.[15] His contrasts do not distinguish between ways that habits and skills *explain* behaviour. They are the same kind of *explanans*, albeit with different characters (Ryle, 1970, pp. 42–43). This means we can characterise the picture in terms of four claims, two of which state supposed similarities between habits and skills, and two of which state contrasts:

Similarities

A) Habits and skills figure in action-explanations in fundamentally the same way;

[14] I want to be clear that I am trying to draw out an implication of what Ryle says. Ryle himself is not explicit on this point.

[15] This commitment to sameness of explanatory role is not made explicit by Ryle and so he is silent on exactly *what* that role is. Later, I will argue that habit and skill play different roles so it doesn't matter what answer Ryle might have given to this question.

B) Habits and skills are dispositions;

Contrasts

C) Habitual actions do not involve self-conscious attention to the standards of the activity, but skilful actions do;
D) Habits are single-track dispositions, but skills are multi-track.

The first two claims make it so that the second two claims are contrasts within a single metaphysical and explanatory kind. But the second two claims signal that habits and habitual actions are unintelligent in contrast to skills and skilled actions. I will continue to call the second pair of claims 'the contrasts' for ease of reference. Importantly, attacking only the second pair of claims would potentially collapse any distinction between skills and habits; attacking only the former pair leaves untouched the core of Ryle's view that habits do not, whereas skills do, express intelligence. In this chapter, I am concerned to deny all of these claims. In the next section I argue against (D), the claim that habits are single-track. In *Section* 3, I argue, against (C), that habitual actions cannot be ruled mindless.

Section 2 – The Complexity of Habit

Firstly, it is worth asking why we should think that habits are characterised by a single conditional. Ryle gives no argument for it. All he says is that habitual actions are "nearly uniform", and he implies that we can describe them with a single conditional (Ryle, 1970, p. 43). But this is insufficient for establishing the claim that habits are single-track, which is in fact quite a strong claim.

Luckily, we can construct a plausible sounding argument for this claim. We usually specify habits with descriptions like 'the habit of running in morning' or 'the habit of interrupting when bored' which have the canonical form 'the habit of *A*-ing in context-type *C*'. Now, it seems plausible that a habit of *A*-ing in *C* can only be manifested by one's *A*-ing in *C* – my habit of running in the morning cannot be manifested by an evening run or a morning walk.[16] But this means that habits can only be manifested in one way, by *A*-ing, and in one context, *C*. Therefore, habits seem to be given by a single conditional: for example, if it were the morning, then I would run. And if habits are like this, then they are single-track.

But there is a problem with this argument, stemming from issues about how to specify the conditions, '*C*', which partly individuate the habit. Let's take Ryle's own example of the habit of smoking. How could we represent this habit with a single conditional? Notably, Ryle fails at this task. He says "My being a habitual smoker [is] my permanent proneness to smoke when I am not eating,

[16] Christos Douskos argues for this in (Douskos, 2019a, 2019b). In Chapter 5, I will argue against this claim, but for now I want to let it stand for the sake of argument.

sleeping, lecturing or attending funerals, and have not quite recently been smoking" (Ryle, 1970, p. 43). The implied subjunctive conditional is:

> *Disjunctive Conditional*: *If* I were not eating, or not sleeping, or not lecturing, or not attending funerals, *then* I would smoke.

The fact that this conditional's antecedent is a disjunction puts Ryle in a difficult position. For a disposition to be single-track, it must be specifiable with a single subjunctive conditional. If we have a disposition which is characterisable with a single conditional, but where the conditional has a disjunctive antecedent, then it is hard to see how there could be any multi-track dispositions. For, for any disposition putatively characterised by a number of conditionals with different antecedents and consequents, we could just collapse the antecedents and consequents into big disjunctions that could then be linked by a single big conditional.

But then, the apparently principled metaphysical distinction between single-track and multi-track dispositions would disappear, since any 'single-track' disposition can be made into a 'multi-track' disposition by a quick semantic trick concerning the logical equivalence of a set of simple conditionals with some big disjunctive conditional. However, Ryle's view depends on there being a principled, *metaphysical* distinction. So he must rule out the use of this semantic trick in the formulation of conditionals specifying single-track dispositions. Therefore, *Disjunctive Conditional* should be decomposed into more, and simpler, conditionals, meaning that Ryle's habit of smoking is multi-track.

It is important to see that I am not just exploiting a weakness in Ryle's formulation; his is not a slip of the tongue. It really is true that *those* are the conditions under which a habitual smoker smokes. But the fact is that they are best treated as many separate conditions, and rather than a single condition. However, this is obscured if we abstract from the habit's numerous stimulus conditions and use terms like 'such-and-such conditions' or 'appropriate conditions'. After all, that is one way of getting a non-disjunctive conditional for Ryle's habit:

> *Appropriate Conditions*: If conditions were appropriate, then I would smoke.

But this is misleading at best. Although the conditional itself does not embed a disjunction, the term 'appropriate conditions' is effectively defined disjunctively, so that *Appropriate Conditions* is logically equivalent to *Disjunctive Conditional*. This makes it just as hard to see how there could be any genuine multi-track dispositions, since, for all the antecedents of the different conditionals which jointly specify a multi-track disposition, we could just define a term like 'appropriate conditions' as being the disjunction of those antecedents, and then formulate a single conditional logically equivalent to the set of the simple conditionals. Again, this is no option for Ryle since it threatens the substance of the metaphysical distinction between single-track and multi-track dispositions.

However, this line of thought seems to call the metaphysical distinction into question regardless of what Ryle does about it. After all, the problems I have raised above seem to derive from the fact that the metaphysical distinction between types of dispositions is drawn in purely semantic terms, since it refers only to the complexity of certain conditionals. But we can freely move between characterising a disposition with one logically complex conditional and many logically simple conditionals so long as they are logically equivalent. The ease with which we can move in the semantic domain suggests that the complexity of conditionals is hardly a guide to the metaphysics of dispositions. So, if there are distinctions between kinds of dispositions (which I am sure there are), the distinction between dispositions in terms of 'the number of their tracks' is not a deep one.

If the distinction is not significant, or if it is merely semantic, then it cannot uphold Ryle's intended metaphysical claims about the differences between the complexity of habits and skills. But if it cannot do that, then Ryle's argument, that habits and habitual actions are unintelligent because habits are single-track dispositions, fails. For we should not draw substantive conclusions about the intelligence of habits and habitual actions from the claim that habits fall on one side of a shallow semantic distinction.

There is another sort of argument one might give for (D) which is suggested by Annas's (2011, 2012) discussions of habit and skill. Annas identifies an agent's having a habit with their having a certain routine. On her view, I have a habit of driving to work the same way if it is a relatively automatic routine of mine – if it is the route I drive regularly and without thinking about it. Then she says that "It is central to routine that the reaction to the relevant situation is always the same; this is why routine is predictable and dependable, which is often useful" (Annas, 2012, p. 102). The thought is that routines introduce predictable, dependable, inflexible, and highly regular sorts of behaviour. And if that is the case, then they cannot exhibit the kind of complexity characteristic of skills. Annas does not phrase this in terms of dispositions, and it is hard to see how one could turn her argument into anything as strong as claim (D), that habits are single-track dispositions but skills are multi-track. However, it is clearly intended to support a weaker version of (D) and has the same spirit as Ryle's more hard-nosed claim.

I think the right response to this can be found in Nathan Brett's excellent paper 'Human Habits' (1981). He argues that just because habits introduce an element of uniformity to our activities, habitual actions are not exact replicas of one another:

> "It is quite possible to make it a habit to go jogging or play a game, or simply get some exercise, to solve a puzzle, say a prayer, read a book, or eat green vegetables, etc. – once a day. Obviously this does not mean that every jaunt, exercise or puzzle, each prayer, or book, or salad must be an identical copy of every other." (Brett, 1981, p. 368)

The uniformity habit introduces is most often at the level of broad types of actions or activities, a uniformity that gives some structure to the kinds of things we do, and when and where

we do them.[17] However, we simply labour the point far too hard if we insist, as Annas does, that these imposed uniformities are essentially inflexible. My habit of reading a book in bed is manifested in my activity, most nights, of reading a book. But I read for different lengths of time, at different speeds, different numbers of pages, and different books, for example when I have finished one novel and pick up another. My habit is flexible enough that all of these things I do can be things I do habitually, and that my doings of these things are thereby habitual actions. The sense in which there is a routine is far too coarse-grained for it to licence anything like claim (D), even the weaker version that Annas's argument is intended to suggest. Therefore, although Annas is right that habits impose routines on us, we must be careful not to over-exaggerate their rigidity.

I have been objecting to the particular arguments that I think makes (D) seem plausible. However, I have also given positive reasons for thinking of habits as, at least for the most part, 'multi-track', if only in the sense that habits are internally complex and exhibit much more flexibility than philosophers often allow for. Habits are not, or at least are not typically, rigid sorts of routines as Ryle and Annas envisage. Habits very often have a wide variety of types of context in which they are manifested, and they can be manifested in a variety of ways. Therefore, not only is the specific claim in (D) false, but the whole contrast it embodies is misguided.

Section 3 – The Mindedness of Habits

So much for one half of Ryle's contrast. I now want to argue that nothing about habits entails that habitual actions are mindless or that habitual actions lack intelligence properties; (C) is false.

Firstly, we should take a moment to register an awkward tension in Ryle's view. Remember, that for an action to be intelligent in Ryle's sense is just for it to manifest an intelligence property. And *any* intelligence property will do. So, if Ryle's claim is that habitual actions are genuinely unintelligent dispositions to do things unintelligently, then he must claim that, not only are habitual actions not careful, attentive, and considered, they are not even stupid, rash, inconsiderate, or unwise.

However, I have cast (C) as the claim that habitual actions are not intelligent in a more limited sense: they do not involve self-conscious attention to the standards of the activity. But, for Ryle, showing that (C) is true is not sufficient for showing that habitual actions do not display a habit-bearer's intelligence, their qualities of mind. So, (C)'s truth is insufficient for showing that habits are unintelligent dispositions to do things unintelligently – that habitual actions are not 'minded' in the way that skilled ones are – a claim that Ryle certainly wants to make.

Indeed, the stronger claim is patently false. If I habitually look at my phone when driving, I manifest my lack of consideration, incaution, and stupidity; if I habitually make goofy faces at babies

[17] Merleau-Ponty makes the same point when he says: "[t]he subject [...] acquires the power of responding with a certain *type* of solution to a certain *form* of situation. The situations may differ widely from case to case, the responding movements may be entrusted sometimes to one effector organ and sometimes to another, and situations and responses resemble each other in the different cases much less through the partial identity of elements than by the community of their sense." (Merleau-Ponty, 2012, p. 143 my italics).

at the park, I manifest my silliness. These qualities of mind are displayed in my behaviour; there is no real comparison between these cases and the behaviour of a clock, a mechanism whose behaviour is genuinely unintelligent. So the strong claim about habitual actions' lack of mindedness is clearly untenable.

Despite this, the weaker contrast in (C) may still stand. Perhaps habitual actions cannot display the sorts of qualities of mind that *skilled* actions do. But I do not think so. Again, Brett's essay speaks well to this point. Brett says that "Clearly behaviour which is the outcome of deliberation and choice is not (in that respect) the product of habit. But if behaviour is the exercise of habit, is it true that the behaviour itself [...] must be done without care and attention?" (Brett, 1981, p. 364).[18] Brett considers a dentist's habit of washing his hands between every patient as a matter of hygiene. He says that it is true that "[if] the dentist must still pay attention to the fact that his hands are not clean [...] and hence, that this is a situation that calls for handwashing, then he has not made this a matter of habit" (ibid). To that extent, Brett agrees with Ryle (his explicit target) that habitual actions must be executed without attending to the facts that make this execution appropriate. However, Brett points out that this does *not* entail that the dentist's action did not involve care, attention, and the kind of heed to the standards of the activity which Ryle paints as the hallmark of skilled action, the kind of intelligence relevant to (C). It is unlikely that a dentist in the habit of washing his hands will simply rinse them – he will scrub them carefully. Brett then rightly argues that "[We] can even speak (without obvious fallacy) of making it a habit to wash with great care and attention. And this means that care and attentiveness are *necessary conditions* of the exercise of some habits. If you do not pay attention as you wash, you have not made it a habit to wash with care" (Brett, 1981, p. 365).

The fact that Ryle misses this is somewhat astonishing, since there are a whole host of habits which are impossible to characterise independently of the quality of mind that its manifestation expresses. The habit of looking at my watch to tell the time is a habit of *attending*. The habit of making a hot water bottle for my girlfriend who is ill is a habit of *taking care* of her. My habit of looking around to see if anyone is following me on a dark night is a habit of *being alert*. Nothing about the way these actions are *initiated* – without conscious deliberation or thought – (or anything else for that matter) should force us to think that *the actions themselves* lack the qualities of mind that Ryle rightly thinks are so important. Brett is quite right that "[something] can be done automatically (i.e. without deciding to do it) without being done as an automaton would do it" (ibid.). So (C) is false.

Brett also pursues a further argument against Ryle. He urges that habits are very often "ingredients" in skills like "being a good driver" – "the habit of keeping one's eye on the ball is essential to being a good ball-player" (ibid.). I think this is too strong – I might be a brilliant ball-player without having a *habit* of keeping my eye on the ball. I might just always keep my eye on the ball. Since habits are not mere regularities but are explanatory of behaviour, we need not think that all such regularities involve habits. But a weaker claim than Brett's *is* true: for some agents who

[18] In Chapter 7, I will argue that Brett's assumption here is false: one can do something habitually and from deliberation. For now, since Brett is simply conceding to Ryle that habitual actions are in some sense automatic, we may treat that point as conceded for the sake of argument.

have a particular habit, having that habit is such that, if they lacked it, their skilled performance would suffer. Maybe it is not essential for the skill of tennis-playing that I keep my eye on the ball *out of habit*, but if, as a matter of fact, *I relied on that habit*, then losing it would make my playing worse.[19]

Ryle's view cannot accommodate the fact that habits can play a positive role in the development and exercise of skills. The problem is that Ryle thinks that habits push out skills, so they cannot be seen to contribute positively to skill. But since they do (or, at least, can), it is false that all habitual actions lack the self-consciousness associated with skilled action.

Section 4 – Habits, Skills, and Action-Explanation

I have argued that claims (C) and (D) are false by attending to the details of the manifold of examples of habitual behaviour as it occurs in human life. Habits are not single-track in the sense of being internally simple, and they cannot be not ruled out from being intelligent and producing intelligent action. For philosophers keen to vindicate habits from the kind of mechanistic view of them that Ryle proposes, these are certainly very important claims to deny.

However, equally important to Ryle's picture as those contrasts are his assumptions about what habits and skills *share*, their similarities. These are embodied in (A) and (B): respectively, that habits and skills figure in action-explanations in fundamentally the same way; and that habits and skills are both dispositions. If we were to rest the argument here, but leave (A) and (B) untouched, we may be tempted to think that habits and skills are just the same sorts of things.[20] However, I think that this view is seriously mistaken. It is mistaken because there are very real and important differences in the explanatory role that habits and skills play in action explanations. In Section 4.1, I argue that habits and skills play very different roles in action explanations. Then, in Section 4.2, I argue that are not both dispositions. So (A) and (B) are false too.

4.1 – The Explanatory Roles of Habits and Skills

In this section, I will argue against (A), the view that habits and skills figure in action-explanations in the same way. The view I will defend is that, whereas skills help explain how it is possible for someone to do something, habits help explain why someone does something (when they do it habitually). I will defend this view by considering and rejecting Christos Douskos's recent, interesting account of their explanatory roles (Douskos, 2019a, 2019b). This is because, whilst (A) has the status of an assumption for Ryle, Douskos *argues* that habits and skills share explanatory

[19] See Will Small's (forthcoming, 2020) for a development of the idea that habits contribute to skills.

[20] This view is held by Levine (2015b, 2015a), Merleau-Ponty (2012), Morris (2012, pp. 65–66), Romdenh-Romluc (2011, p. 789), and Sachs (2014, Chapter 5).

duties. So I will respond to Douskos by showing that his arguments fail, and that the view he arrives at is problematic. Then I will develop my alternative.

The discussion will focus on the way that reference to habits and skills functions in answers to certain sorts of explanatory questions. This is because the answers to questions that ask for explanations of things (what linguists call *wh*-questions) are explanations, and different sorts of *wh*-questions ask for different sorts of explanations. 'What is going on here?', 'How did you get up there?' and 'Why are you doing that?' all ask for different sorts of answers, each of which explains what is going on in a different way: 'I'm trying to fix the roof'; 'I used the ladder'; 'The tiles are loose'.[21] The thought that guides my (and Douskos's) discussion is that, if habits and skills figure in answers to different sorts of *wh*-questions, then they provide different sorts of explanations, and contribute differently to action-explanations. If they figure in answers to the same sorts of questions, performing the same role, then (A) is true, and they have the same explanatory role.

Douskos begins by motivating a 'Simple View' which he ultimately rejects. On the Simple View, habits figure in answers to *why*-questions, and skills figure in answers to *how*-questions. For example, if asked 'Why are you singing that tune?' I may mention my habit. My habit figures in an explanation of why I did that thing. In contrast, my capacity to sing does not give a good answer to the *why*-question. Instead, on the Simple View, Douskos says that skills figure in answers to *how*-questions because, when someone has a skill, their skill determines the means they take to do what they want to do (Douskos, 2019b, p. 988). For example, where one previously had to work out how to tie shoelaces, one's skill now 'takes over', solving the problem itself by getting one to take the relevant means given one's ends.

Now, Douskos thinks the Simple View is false, and his argument is interesting. He thinks it is false because many habits are what he calls 'habitual routines': habits of *B*-ing when *A*-ing, where *B*-ing is one's means of *A*-ing. For example, Kate's habit of taking a certain route when going to work is a habitual routine. Taking that route (*B*-ing) is Kate's means of getting to work, and, when she is going to work (*A*-ing), she habitually takes that route (*B*-s). Douskos's argument is that when habitual routines manifest, they both figure in an explanation of why one *B*-s *and* how one *A*-s. Kate's habitual routine helps explain *why* she takes the route she does, and it explains *how* she got there. He makes a similar point regarding skills: if Kate makes a skilful move in a game, her skill both helps explain *how* she performed the move (it explains her taking the means she did), and *why* she did so: Kate made that move because she was so skilled. So, on Douskos's view, habits and skills can both figure in answers to both *why*-questions and *how*-questions, so they can have the same explanatory roles.

I think there are some mistakes in Douskos's argument. Firstly, we should take care about the variety of *how*-questions. In particular, there are two kinds of *how*-questions which Douskos

[21] This approach is closely related to a question-theoretic conception of explanation associated with Achinstein (1977), Cross (1991), and Van Fraassen (1980). However, one can deny these theories (for example, that the objects of explanations are indirect interrogative questions, or that a theory of explanation is primarily a theory of the pragmatics of certain kinds of pairs of question-answer utterances) whilst accepting that *wh*-questions typically ask for explanations and that a good way of finding out whether two items have different explanatory roles is by looking at what sorts of questions they figure in answers of.

problematically conflates. Following Stephen Yablo[22], I will call these *means*-questions and *manner*-questions. A *means*-question asks for the means by which something did, will, or might happen. For example:

(*Means-Question*)
'How do I make an omelette?'
- 'First you get the eggs out, then you crack them into a bowl, whisk them...'

The answer states some things you have to do in order to make an omelette. That is, the answer really lists some act-types which, when you chain them appropriately together results in your having made an omelette. Now, we sometimes call the things you have to do in order to make an omelette a *way* of making an omelette. After all, two people may make omelettes differently, for example, if one person uses cheese and the other doesn't. And if they do, then they make omelettes *in different ways*.

Contrast this with a *manner*-question:

(*Manner-Question*)
'How did she sing?'
- 'Beautifully and quietly'.

So, *manner*-questions also ask for the ways someone did (may do, will do) something. But *these* are not the same *sorts* of ways as the ways one makes an omelette! Whereas the ways which are means are act-types, 'beautifully' and 'quietly' are not act-types, or things you can do. Plausibly, they are *ways of doing* things you can do.

Manners and means are both sorts of ways of doing something, but they are different kinds of ways. Without trying to work out a theory of ways, let's use 'manner' and 'means' as terms

[22] *Means*- and *manner*-questions are the most commonly discussed *how*-questions in the literature (Cross, 1991; Ginzburg, 2011; Sæbø, 2016). However, in his Whitehead Lectures (2018), Yablo identifies and distinguishes between no less than five different sorts of *how*-questions:

Means-questions: 'How do I get to Carnegie hall? – You go straight then turn left.'
Method-questions: 'How did she get up there? – With a ladder.'
Manner-questions: 'How did he speak? – Softly, sadly.'
Merit-questions: 'How did you sleep? – Badly'
Status-questions: 'How did he abandon his intention to apply for a Green Card? – By leaving the country.'

What is striking about all of these questions is that, whilst they differ from each other to varying degrees, their answers are all what we may call *ways* that something may happen or be done. This indicates that the notion of a *way* contains hidden complexities, perhaps not always appreciated by some philosophers who make use of them (Stanley & Willlamson, 2001). For more on hows and ways in connection with modality, see Yablo (1996).

denoting two different types of ways, and let's assume that we have an intuitive enough grasp of the distinction.[23]

Douskos says that skills figure in answers to *how*-questions, and he means *means*-questions. That is why he says that skills resolve the question, for an agent, of how to do something (Douskos, 2019b, pp. 988; 992). But it seems false that skills figure in good answers to *means*-questions:

'How do I get to Carnegie Hall?'
- 'By being able to walk'
- 'By being able to read a map and get around accordingly'
- 'By being a good musician'

All those answers attempt to answer the question by making reference to a skill or capacity. The answers use different methods, by directly adverting to different skills (for example, map-reading, walking, and musicianship skills). But none seem to work; they all sound like poor jokes. So it looks like skills do not figure in good answers to *means*-questions.

One might wonder if the fact that the questioner is asking how *to do* something, rather than asking for an explanation of *how something happened*, is skewing the results here. After all, we are interested in explanations, not advice.

But, firstly, the question does ask for an explanation, and it does use a *means*-question construction. Therefore, on Douskos's view, it must be possible to give some answer adverting to a skill. Secondly, putting the questions in past tense is no help:

'How did you make that chocolate cake?'
- 'By being a really good baker'

This is a joke, not an explanation. The fact is that it is just very hard to see how one can advert to a skill in a genuinely good answer to a *means*-question. In contrast, it is very easy to answer a *manner*-question by adverting to someone's skill:

'How did she climb that boulder?'
- 'Very skilfully.'

That's a perfectly good answer. So although skills can figure in answers to *how*-questions, of these two types, they can only figure in answers to *manner*-questions, not *means*-questions. But since Douskos's claim was that skills figure in answers to *means*-questions, his intended rendering of the claim that skills can answer *how*-questions is false.

[23] Yablo (2018) argues that the distinction between manners and means collapses somewhat when we turn our attention to determinates and determinables, for example the relation between *being scarlet* and *being red*. This may be true, but the concern with determinates and determinables is quite far from the current topic.

The problem is not that Douskos is unaware of *manner*-questions. The problem is that he conflates them with *means*-questions:

> "To *A* skilfully is to *A* in a skilful way, just as to *A* elegantly is to *A* in an elegant way. Manner adverbs are apt to feature in answers to 'How?' questions, the same sort of question the answer to which may feature an adverbial prepositional phrase specifying a means or way of doing something: 'by *W*-ing'. (Douskos, 2019b, p. 988)

But this is just a mistake. If we speak only in the overly general terms of *how*-questions, then it may seem that both sorts of ways – means and manner – may feature in the answer to any old *how*-question. Then it looks like skills can figure in explanations of one's taking a certain means to a given end. But we have seen that they cannot. So we should deny that skills provide us with answers to *means*-questions; they are only pertinent to *manner*-questions.

There is a further problem with Douskos's treatment of skills and *how*-questions. He is led to think that skills figure in answers to *means*-questions by the fact that a skilled person need not deliberate about how to go about doing something – their skill 'solves it for them', if you will. But this is not pertinent to whether or not skills figure in explanations of how (by which means) someone does something. The way that *means*-questions are *resolved by agents* is not relevant to what counts as an answer to the *means*-question 'How did they do it (what means did they take)?' The ways that *means*-questions are resolved by agents are themselves answers to *different means*-questions, such as 'What means did they take *to resolve the question of what means to take*?'

An example may help. If two people solve a puzzle by putting the pieces together in the very same order, but one had to deliberate hard and the other found it easy due to their skill, then one question of how they solved the puzzle is 'How (by which means) did they solve the puzzle?'. The answer is the same in both cases, because the way they solved the puzzle was by taking the very same steps – first put down this piece, then that, then this one. In this sense, they do not differ in how they solved the puzzle. But the answers to a second *means*-question do differ: 'Which means did they take to resolve the question of where to put the puzzle pieces?' For the person who had to deliberate, the answer is 'By deliberating' – that is the means he took. But for the other agent, there seems to be no answer at all. If she just saw where the pieces went and put them there, then there was no means that she took, the employment of which could explain how she decided to put the pieces where she did. Actually, the fact that she was able to solve the puzzle so easily, without having to employ any means, seems to be a part of why we call her puzzle-solving skilled. The example shows that how someone resolves a practical question does not bear on the answer to the question of the means they took to their end.

I have been arguing that Douskos is wrong to think that skills figure in answers to *means*-questions, and therefore figure in explanations of the means a person takes. What about his claims about habits? After all, my goal in this section is to show that habits and skills do not figure in answers to the same sorts of questions, and therefore that they have different explanatory roles. Might it still be true that habits can still figure in answers to the same questions as skills?

Douskos makes the case that habits can provide answers to *how*-questions, but, again, he means *means*-questions. Therefore, even if he is right, this would not show that habits and skills can play the same explanatory role, since I have argued that skills cannot figure in answers to such questions. That's not the conclusion Douskos is after.

On top of this, though, we can ask whether it's true that habits provide answers to *means*-questions:

'How (by which means) did Kate get to work?'
- 'She has a habit of going to work by that route.'

But this answer is totally incongruous, indicating that it fails to provide a proper explanation. Douskos clearly thinks that cases of habitual routines like Kate's are supposed to provide clear examples of habits figuring in explanations of the means a person took to doing something else. But if this is supposed to be an explanation of *how* Kate did something, why won't it fit as the answer to the appropriate *means*-question?

The reason is that it is not really an explanation of how Kate got to work: it is an explanation of *why she went to work by that route*. Sure, explaining that she *A*-ed by *B*-ing, where she *B*-ed out of habit means that Kate's habit figures in an explanation of something about the means by which she *A*-ed. But that is not sufficient for explaining *how* she *A*-ed, since what it is about the means by which she *A*-ed which Kate's habit helps explain is *why she B-ed*:

'Why did Kate take that route to work?'
- 'She has a habit of going to work by that route.'

That is perfectly sensible, in contrast to the previous question-answer pair. I would venture that Douskos's mistake in this case is to confuse the sameness of subject matter of two explanations for sameness explanation. That is, the explanation of why Kate *B*-ed has the same subject matter as the explanation of how Kate *A*-ed – they both concern Kate's *B*-ing. However, this does not mean that they explain the same thing, since they explain different things about Kate's *B*-ing: one explains why she *B*s; the other explains how she *A*s.

Douskos is wrong, then, when he argues that skills and habits can both figure in answers to the same questions: habits can only figure in answers to *why*-questions; skills cannot figure in answers to *why*-questions or *means*-questions, but they can help answer *manner*-questions.

This is important for our discussion of the Rylean picture, since it helps dislodge the idea that habits and skills compete for explanatory space in action-explanations. But I think we can do better, for I think we can work our way towards a positive proposal about the explanatory differences between skills and habits.

Part of my positive proposal involves making a further claim about the explanatory role of skill. I want to claim that skills figure in answers to *how-possible*-questions because they are, as Ryle rightly says, sorts of *capacity*. The connection between capacities and *how-possible*-questions comes

out in Aristotle's discussion of the explanatory power of capacities. In *Metaphysics Theta*, when arguing against the view of the Megarians that there are no such things as possibilities or capacities, Aristotle argues that:

> "[I]f what is deprived of a capacity is impossible, it will be impossible for what is not happening to happen [...]. For [...] what is seated will always be seated; for being seated it will not get up; for it is not possible for something not capable of getting up to get up." (Aristotle, 2006, p. 4)

Aristotle's argument is that the explanatory role of capacities (and therefore skills) is to explain how certain types of events and processes are possible. They do not explain *why* they are actual (when they are), since that is for other sorts of entities, such as reasons or causes. Capacities explain why it is true of someone not currently singing that they *can* or *could* sing; they explain how it is possible that they sing. Capacities are therefore what Barbara Vetter calls a "localized modality" – a kind of modality that attaches to particular things within a world, in contrast to the way that global modalities expressed with sentential necessity and possibility operators attach to the worlds themselves (Vetter, 2015, p. 2).

I think that this yields a good understanding of the explanatory role of skills. Skills are uncontroversially sorts of capacities: a tennis-playing skill is a capacity to play tennis; being a skilful typist is having a capacity to type that meets a certain standard. Therefore, having skills makes it possible to do certain things, and so they can figure in answers to *how-possible*-questions:

'How was he able to play such a hard tune?'
- 'Well, he's a very good guitar player.'

'How was it possible that she beat him at tennis?'
- 'Her skill is that much greater, I guess'

I will not develop this view of skill further. There are clearly interesting questions here about what marks skills out from other capacities, their relation to evaluations of performances, and so on. But the important point for me is that we have isolated skill's explanatory role in action-explanations, and that its role is very different from that of habit's, thereby showing that (A) is false.

However, we can go one step further. Whereas, according to Ryle's claim (A), habits and skills fight for explanatory space, pushing each other out, I think the view of skill and habit I have just defended allows us to see how they often provide *complimentary* explanatory contributions. The thought is that if habits figure in explanations of why someone does something, then it must be possible for the person to do it. And sometimes, what makes it possible for them to do it is their skill. Therefore, there must be action-explanations which appeal both to a person's habit and their skill. More precisely:

1. To have a habit of *A*-ing, one must have the capacity to *A*.
2. For many habits of *A*-ing, the capacity to *A* is a skill.
3. For any habit of *A*-ing where the capacity to *A* is a skill, any manifestation of that habit is also the manifestation of a skill.

C) Therefore, many manifestations of habits are also manifestations of skills.

Premise (1) follows from the Aristotelian view of capacities, since if one *A*-s habitually then one can *A*, which in turn entails that one has a capacity to *A*. Premise (2) simply seems impossible to deny. My habit of playing 'Take Five' whenever I pick up the guitar is a habit of doing something – guitar playing – where my capacity to do that is a skill. Premise (3) follows pretty swiftly, and the conclusion is utterly antithetical to the Rylean view I have been attacking.

So, the truth is once again very far from Ryle's claim. It is not just false that habits and skills figure in explanations in the same way, and therefore push each other out as (A) claims, it is very often the case that a full explanation of how someone came to do something out of habit requires both that we cite their skill *and* their habit.

4.2 – Habits and Dispositions

So far, I have argued against both of Ryle's contrasts between habit and skill, (A) and (B), and I have argued that his implicit assumption about their explanatory roles, (C), is also false. My final job is to show that (B) is false, that habits and skills are not both dispositions as Ryle claims they are. There is a lot that one can say about this, and there is a particularly interesting debate about whether skills and abilities are dispositions.[24] However, I will only consider habits at any length for two reasons. Firstly, if I can show that habits are not dispositions, then (B) is false because then, whatever is true of skills, it is false that they are *both* sorts of dispositions. Secondly, it will be important at various points later in the thesis that habits are not dispositions, so I will focus on them.[25]

The argument against thinking of habits as dispositions is fairly simple. On every view of dispositions, it is consistent with ascribing a disposition to an object that the object never does what it is disposed to.[26] A fragile glass may never break; a malleable piece of clay may never deform. However, it is inconsistent with ascribing someone a habit of *A*-ing in *C* that the person never has, does, or will *A* in *C*. After all, habits are formed by repetition, and if we stop repeating them, then they can wane over time. Someone cannot *get* the habit of throwing scrunched up balls of paper in the bin when bored without sometimes actually being bored, scrunching up paper, and trying to get it in the bin. And if they stop doing that when they're bored (they make paper airplanes instead),

[24] See Jaster (2020), Small (2017a), and Vetter (2019).
[25] The fact that habits are not dispositions will crop up, for example, in Chapter 6.
[26] This is a feature of accounts as different as Bird (1998), Fara (2005), Lewis (1997), Molnar (2003), Mumford (1998), and Vetter (2015).

they will lose their habit. So it is a requirement on having a habit that one actually do what one is in the habit of doing, at least sometimes.[27]

What, then, *are* habits if they are not dispositions or capacities? I think they are what we may call 'tendencies' or 'propensities'.[28] This claim unites the thoughts in the previous paragraph and previous section. A tendency may help explain why someone did something. For example, Bill's tendency to get angry at the news helps explain why he gets angry when he does; Jill's tendency to pick her nose helps explain why she does that too. But tendencies must also sometimes actually manifest: Bill and Jill may only have their tendencies if they actually do those things sometimes. The less someone does something, the less willing we are to say that they tend to do it, or that they have a tendency to do it. In that respect, tendency-ascriptions must invoke *patterns* of behaviour, whereas neither disposition- nor capacity-ascriptions do (since neither need be manifested in order to be instantiated).

Interestingly, tendencies are often ascribed using what linguists call 'habituals', sentences of the form '*S A*-s in *C*': Mary smokes when she gets home; Peter sings when he's in the shower (Carlson, 2005; Ferreira, 2016). Habituals say something about what an object normally, or typically, does. This is why ascribing a tendency to an object is inconsistent with the object never doing the relevant thing. Of course, sentences which ascribe *habits* canonically have the form of a habitual. Sometimes, we add that Mary (say) smokes *out of habit* when she gets home. But what that indicates is that Mary's habit of smoking is operative in an explanation of why she smokes when she gets home, thereby distinguishing Mary's case from one in which there is mere behavioural regularity. If we don't wish to add that information, we may ascribe a habit to Mary by saying that she smokes when she gets home.

It looks, then, like habits are tendencies, not dispositions. I have argued that skills are capacities and have left it up to the reader to decide whether capacities are dispositions. I don't think they are, but I will not argue the case here. The important point is that (B) is false: habits and skills are not both sorts of disposition.

Conclusion

I have argued that all of Ryle's claims are false: habits are not unintelligent dispositions to do things unintelligently; habits are not 'single-track' or internally simple; habits and skills do not share an explanatory role; and since habits are a kind of tendency, habits and skills are not both kinds of disposition. Therefore, nothing of Ryle's picture stands.

In fact, I think that my arguments have shown much more than this. I have argued that habits are often internally complex, and do not typically impose highly rigid routines on our practical lives; that habitual actions are often skilled actions, and (probably) always display a person's

[27] Another way of putting this is that we may stipulate a case where someone who is never bored nonetheless has the *disposition* to throw paper balls in the bin *if* bored. But we cannot stipulate that the person has the *habit* of doing so. So disposition-ascriptions come apart from habit-ascriptions precisely (or at least) on this point.

[28] I use 'tendency' and 'propensity' interchangeably throughout the thesis.

intelligence. They can also help contribute to a person's skill in some domain, and one can even have a habit of doing things attentively and carefully.

These are important results. Pictures like Ryle's are quite predominant in contemporary philosophy, even if philosophers do not spell out their view in as much detail, and with as much self-consciousness, as Ryle did. But, as I said in my introduction, pictures like Ryle's constitute a layer of preconceptions about habits which philosophers often seem to adhere to. This is the reason why philosophers so often preface any mention of habit with 'blind', 'dumb', 'mere', 'mechanical' or 'automatic'. So, although I have been attacking Ryle's view in particular, I think that, really, something like his view (if not only his *attitude*) is shared by a large number of philosophers. That it is so widely shared makes it difficult for me to put forward my own view, on which habit and habitual action integrated very deeply with rationality, intention, and reasons. So the results of this chapter represent an initial vindication of my project: we simply get the phenomenon of habit very wrong if we focus myopically on the routine nature of habits, or the fact that some things we do habitually we do absentmindedly. Of course, these are features of habit and habitual action, but we cannot let them blind us to the rich, complex nature of the phenomenon. Indeed, the fact that we have been misled by focussing too narrowly on those aspects can serve as an invitation to broaden our thinking about habit, and think much harder about its relation to things which we do not traditionally associate with it, such as reasons for action and intention. The rest of this thesis attempts to take up this invitation, and finds that it is much more importantly embedded in the framework of our rational capacities than any view or attitude like Ryle's could ever accommodate.

CHAPTER 2

RATIONALISING EXPLANATIONS

I have a habit of going and drinking coffee at *Gilmour's Café* every morning. On most occasions when I go there out of habit, I go there partly for the reason that it is cosy, close, and that the coffee is good. In fact, *Gilmour's Café* is popular, and most people drink coffee there for those sorts of reasons. This is because these are reasons which pretty much anyone can have, whether they have my habit or not. In fact, my friend Alice – who has never been there before – went there yesterday for just those reasons. So, when the possession of these reasons contributes to explaining someone's drinking at this café, it does so independently of whether the agent has a habit of drinking there. But when I drink at the café, a part of what explains why I do that is the fact that I have a habit of going there.

There are two competing ways to think of my habit's explanatory role, here. The first is to say that the fact that I have a habit of going to *Gilmour's* adds nothing to the set of reasons I have to drink there, and does nothing to alter the landscape of these reasons. My having this habit does not mean that I have any more reasons than Alice does to drink there, and brings with it no new sensitivities which rationally bear on my possession of reasons. Therefore, my habit plays no role in a *rationalising* explanation of my going there when I do so out of habit. The habit figures in an explanation of what I do in some other, perhaps purely causal, way. No comparison between the rationalising explanations of my drinking there and Alice's drinking there hinges on the fact that I am drinking there out of habit. It is not a 'rational difference-maker'. We can call this the 'Non-Rationalising View of Habit', or 'The Non-Rationalising View' for short. Philosophers rarely consider the question whether habits might be rational difference-makers, partly, I think, because so many think that habits are mindless compulsions. In the previous chapter I argued that much of the basis for that picture is wrong. Nonetheless, I think that that the Non-Rationalising View is likely to be the default view for most philosophers.[29]

The second way of thinking about habit's explanatory role is that it *is* a 'rational difference-maker'. That is, when I drink at *Gilmour's* out of habit, the fact that I have a habit of drinking there

[29] David Velleman is an important exception here. In *Practical Reflection* (1989) he argues that habits provide us with reasons for action. I discuss Velleman's view in Chapter 6. Similar sympathies are also found in remarks of Donald Davidson's (1980d, p. 225) and Mark Johnston's (2001b, p. 188).

adds something to a full rationalising explanation of why I go there. So, insofar as I have a habit of drinking at this café and Alice does not, either I possess reasons to drink there which Alice does not, or the landscape of my reasons is appreciably altered by my having the habit, or it has some other rational bearing on my having of, and acting for, reasons. I will call this 'The Rationalising View of Habits', or 'The Rationalising View', for short.

The main goal of this thesis is to argue for the Rationalising View. This chapter is the first stage of my argument. In it, I will defend a conception of reasons for action and rationalising action-explanations. After all, it is no use defending a Rationalising View of Habit if we don't even know what a rationalising explanation *is*. This chapter therefore lays down a framework within which we can state the Rationalising View more precisely, and within which we can assess arguments for and against it. My discussion of reasons for action is very selective, since what I am trying to do is draw out some theses which are important to the way I want to frame my discussion of habits and how I want to state and defend the Rationalising View. Therefore, there is much of interest about reasons and rationalising explanation which I leave aside.

It will be helpful to point out two views which I will defend in this chapter which are particularly important for the rest of my thesis. Both are views about the *explanans*, or explainers, in rationalising action-explanations. Firstly, I argue that reasons for action are facts which favour doing certain sorts of things, and that when someone is motivated to do something for a reason, that reason explains why they do it, but only because it seemed to them to favour doing it. This is really the heart of my account of rationalising explanations, since it connects the reason for which someone did something with the agent's (fallible) perspective on their own reasons. Rationalising action-explanations therefore explain why someone did something by giving *the person's own* reason for doing it. Strictly speaking, though, on the view I defend, only reasons – a class of *fact* – explain things.

The second important view is that, although reasons are the only explainers, other sorts of item (such as desires, beliefs, pains, and virtues) can *figure in* rationalising explanations when they are rationally relevant to a person's having of, or acting for, a reason. For example, although my reason for taking the pain killer is the fact that my knee hurts, my *pain* figures in a rationalising explanation of why I take the pill because the pain grounds the reason for action and provides me with knowledge of it (by being the a first-personally available phenomenal occurrence of a reason-grounding thing, a pain). Similarly, my belief that I have a reason to go outside figures in a rationalising explanation of why I go outside, even though it is not the reason for which I do that. The phrase '*x* figures in an explanation', then, is a term of art for me.

These two views are particularly important because they mean we can state the Rationalising View of Habit precisely. For example, if I am right that only facts can be reasons but that other items may figure in rationalising explanations, this suggest that we should state the Rationalising View as, roughly, the view that habits figure in rationalising explanations of why people do things, when they do them habitually, by being rationally relevant to their having of, or acting for, reasons. In turn, this gives us a framework in which we can assess arguments for, and against, the Rationalising View.

Without my views in place, other, much more vulnerable, statements of the Rationalising View may seem attractive. For example, we may be tempted to state it as the view that habits are reasons for action, or habits are motivating reasons, or something similar. Now, one may hold these views if one wishes, but I think they are quite misguided because they get the metaphysics of reasons and rationalising explanations wrong. Therefore, I have opted for a strategy whereby I first defend some very general views about rationalising explanations on their own merit, and then use those views to formulate the theses and questions about habit.

This chapter is split into five sections. In Section 1, in order to orient the discussion, I articulate some basic distinctions one finds in discussions of reasons. In Section 2, I argue that reasons are always and everywhere facts, and defend this view against the classic problem of error that derives from cases where agents seem to act on a false belief. In Section 3, I briefly argue that the *explananda* of action-explanations are neither actions nor act-types. Instead, I suggest that the *explananda* are some sort of semantic item which takes what an agent does on some occasion as its topic. In Section 4, I make an important distinction between narrow and broad rationalising explanations. I argue that although the reasons for which one does something narrowly explain why one acts, all sorts of items figure in broad rationalising explanations. In particular, I argue that mental properties are especially apt for figuring in such broad explanations. Finally, in Section 5, I put all this together in articulating the Rationalising View, before pointing out how I will proceed to defend it in subsequent chapters.

Section 1 – Explanatory, Normative, and Motivating Reasons

Rationalising explanations are a sub-class of explanations of why things happen, so one place to start is to compare them to other sorts of explanations. For example, think about scientific or meteorological explanations. One good meteorological explanation is this: the reason the weather is so grey outside is that the air pressure has dropped. However, notice that the fact that the air pressure has dropped is not the *weather's* reason for becoming grey.[30] The weather is not the kind of thing that can *have* reasons, even though reasons are involved in explanations about the weather. This tells us both that reasons have a role in explanations that have nothing to do with rationality or rationalisation, and that there is an important distinction between things which can 'have' reasons and things which cannot. Some philosophers call reasons which are involved in explanations of why things happen 'explanatory reasons'.[31]

Reasons for action and belief seem to be different from the reasons involved in meteorological explanations in a couple of ways. Firstly, reasons for action and belief have a normative force. They favour doing or believing things, and (immense subtleties aside) to do what a reason favours because it is favoured by that reason is to do something right or good or rational.

[30] I owe this way of putting things to (Alvarez, 2010, p. 30).

[31] This characterisation, and what follows, owes much to Alvarez (2010) and Raz (2009).

Such reasons set rational standards for action, and so philosophers have called them 'normative reasons'.

Secondly, unlike the weather, people *are* the kinds of things which can 'have' reasons in the sense that we feel the rational force of certain facts with respect to what we should do and think – we can understand that a fact is a normative reason for us to do something. Thirdly, that our behaviour is governed to a degree by whether we feel the rational force of normative reasons means that we can fail to abide by them. Therefore, we can have normative reasons to act which we ignore or fail to live up to. But this means that, unlike so-called explanatory reasons, normative reasons need not actually figure in any explanations at all.

Now, sometimes, in taking a reason to have rational force, we act for that reason. As such, there seems to be a class of reasons which are both normative and explanatory. But they are only explanatory in virtue of being taken up by an agent as something which speaks in favour of acting in a certain way. Philosophers sometimes call these 'motivating reasons', and it is explanation by so-called motivating reasons that we initially home in on when we talk about rationalising explanations of why we think what we think and do what we do. This is because it is explanation by motivating reasons which makes it intelligible why an agent did what they did from the agent's own point of view. Knowing a person's own reasons for doing something makes their doing it intelligible to us as a case of (at least putative) rationality. For example, if Sally ran because there was a bear chasing her, that there was a bear chasing her makes Sally's running a putative candidate for being rational only because Sally appreciated that fact as speaking in favour of running *and* the fact actually explained why she ran (Stout, 2009).

An important part of this picture of so-called motivating reasons is what Doug Lavin calls the Cognition Requirement: "in standard rationalising explanations of a belief or intention by appeal to a reason, the connection between encountering a fact and forming the relevant belief or intention [...] holds precisely in virtue of the fact that the subject appreciates the rationalising force of the fact at issue" (Lavin, 2011, p. 370). In short: if a motivating reason explains why someone acted, the person must have treated the reason *as* a reason. Otherwise, the explanatory connection bypasses the agent's rational faculties, which makes the reason 'merely explanatory'. In such a case, a reason may explain why someone does something, but it will not provide much of an insight into their own motivations. For example, the fact that Jill is shy may explain why she sits at the back of the class, but the fact that she is shy is not what moves her; it is not Jill's own reason. Jill's own reason may be that she is less likely to be asked a question sat at the back. There are clearly interesting connections between these two reasons, some of which I will discuss further in Section 4. The point for now is that they explain Jill's behaviour in different ways: one is merely explanatory; the other gives Jill's own reason.

This three-way division can already help us sharpen the Rationalising View of Habit. For someone sympathetic to the Non-Rationalising View, it will be natural to think that my habit's role in explaining why I go, on some occasion, to *Gilmour's Café* is captured by giving a merely explanatory reason: that I have a habit of going there. My habit stands to my coffee trip as the air-pressure stands to the weather, or as shyness stands to Jill's sitting at the back. The thought is that

the fact that someone has a habit is clearly relevant to explaining why they do certain things, but that it plays no role in any rationalising explanations of what they do. Put another way, explanations citing someone's habits are explanations which provide us with no insight into the habit-bearer's own reasons for action. This is not to say that the Non-Rationalising View must deny that I may go to the café for reasons when I go out of habit. That would be very implausible – I go there for the reason that I like the coffee, the staff are funny, and it is close by. What the view *is* committed to is denying that any of the motivating reasons I have for going there are affected in any way by my habit, that the habit provides me with a new motivating reason, that it alters my appreciation of an existing reason, and so on. Therefore, in contrast, the Rationalising View must be the view that my habit's explanatory role is not captured by a merely explanatory reason, and that it somehow affects my having of, or acting for, *motivating* reasons.

I think this starts to clarify what is at issue in stating and deciding between the two competing conceptions of habit. However, I think much more must be said before we can properly articulate the Rationalising View. In particular, I will defend a view of the ontology of reasons (Section 2), a view of what they explain (Section 3), and my view that some items may figure in explanations without being reasons (Section 4). Only then can I say more fully what the Rationalising View is and how one might defend it (Section 5).

Section 2 – Reasons are Facts...

In the previous section, I introduced a three-way division amongst reasons – explanatory, normative, and motivating. Many philosophers take this as a division amongst *kinds* of reasons (Parfit, 1997; Smith, 1994). However, I take a monistic view on which reasons are all fundamentally of the same kind, and where the three-way division really marks a distinction in the *roles* that reasons can play in explaining or setting standards for things.[32] On my view, all reasons are facts, however, some of these facts explain things (explanatory and motivating reasons) whereas others favour things (normative). And amongst the reasons which explain things, some do so without having any connection to value (they are merely explanatory), whereas others do so by favouring believing or doing something and being treated as so favouring (motivating). In general, this seems an attractive position, since it preserves unity within the class of explanations whilst connecting some of those explanations with normativity in the cases of motivation by normative reasons. It is also a simple, parsimonious view, and I think it should be something of a default which would require a special argument to reject.

My way of introducing the divisions has also been 'externalist'. It is externalist in the sense that reasons are everywhere seen as facts, and never seen as psychological states or propositional attitudes like believings or desirings which are 'internal' to the agent.[33] It is the fact that a bear is

[32] In this respect, as should be obvious, I am really indebted to Maria Alvarez's *Kinds of Reasons* (2010).

[33] Proponents of this view include Alvarez (2010), Dancy (2000), Raz (2009), Sandis (2013), and Stout (2009). This is not to say that this view is externalist in Bernard Williams's sense, where someone's having a reason to do something

chasing her which is Sally's reason for running, and the fact that the air pressure has dropped which explains the weather's turning. Although Sally must believe that a bear is chasing her for her to treat that fact as a reason for her to run, it is not her believing of that fact which is her reason for running. The externalism and monism are connected, since a very plausible reason for denying monism is the apparent attractiveness of the view that motivational reasons are psychological states like believings or desirings, or pairs of those states (Davidson, 1963; Smith, 1994). I will call that view 'psychologism about motivating reasons'.

As I have said, it is important to me to defend this monistic and externalist conception of reasons since it informs how I want to formulate and defend the Rationalising View of Habit. Now, I am not claiming (though it may be true) that the monistic and externalist view of reasons is the only or best way of formulating the Rationalising View. But as I have said, it will be much easier to articulate the view once we have a clear conception of central notions like that of a reason, and so I will be upfront in defending what I think is the right conception of such things.

Therefore, I will offer four brief arguments in favour of externalism specifically regarding reasons which motivate us to do act. I focus on motivating reasons for two reasons: firstly, they are central to a picture of *rationalising* explanations because they are agents' own reasons; secondly, because it is motivating reasons which many philosophers give psychologistic, non-externalist accounts of. After giving these four arguments, I will then defend externalism from a common worry about the factivity of explanations. Here are the arguments.

Firstly, reasons must be able to be premises in practical and theoretical reasoning, for we deliberate by weighing and assessing our reasons.[34] This means reasons "must be the sort of thing that can be thought or said on behalf of an act" (Darwall, 1983, p. 31). But psychological states cannot be 'thought or said' on behalf of anything. One can say that one believes something by saying 'I believe the ice-cream is in the freezer' and say that one desires something by saying 'I want the ice-cream'. But the content of what is said in both cases is a proposition *about* what one believes or wants. Expressing thoughts about one's beliefs and desires is not the same as 'thinking or saying' the beliefs and desires themselves.

Can't we think or say *the contents* of our psychological states? For example, for the belief one would say 'the ice-cream is in the freezer'. Sure. But in saying what one believes, one just utters a sentence expressing the proposition *that the ice-cream is in the freezer*. That is not a psychological state either; it is a proposition. The case of desire is different, but no help for someone wanting to say that motivating reasons are desires. This is because to say what one desires in this case is just to describe the desire's object: 'the ice-cream'. But that is not even a proposition, but a definite description. (Of course, in a different case one might have a propositional desire *that I eat the ice-cream*. But not all desires are like that, and it is still no help: expressing that proposition isn't thinking or saying a psychological state either!).

---

does not depend on their psychology (Williams, 1979). My externalism is a view of the ontology of reasons – a view of which items they are. Whether some of these items' status as a normative or motivating reason must depend on desires is an orthogonal issue.

[34] Reasons are not the *only* things that we assess in deliberation. Jonathon Dancy gives an excellent account of the variety of considerations in his book, *Practical Shape: A Theory of Practical Reasoning* (2018).

The problem is that the states themselves are *relations* of believing and desiring certain things. But relations cannot be the contents of what we say or think. We think and say things which either are or express propositions. So psychological states cannot act as premises in reasoning. Therefore, they cannot be reasons. What can be said or thought (*a fortiori*, can be said or thought, in a piece of practical reasoning, on behalf of doing something), however, are propositions. But false propositions cannot favour anything – if it is false that the ice-cream is in the freezer, that proposition cannot be a reason to go to the freezer to get the ice-cream. Therefore, only true propositions can be reasons. Since one good sense of 'fact' is of being a true proposition, reasons are facts in that sense.

Secondly, there is a powerful and famous argument of Jonathan Dancy's (2000, p.103). He argues that it must be possible to be motivated to act by good reasons. For example, if the fact that my friend needs help is a good normative reason to help her, then it must be possible for me to recognise this and help her for that very reason. If it were not possible, then there could be no such thing as doing something for the right reasons. But the view that motivating reasons are desires rules this out, since no desire (or any other sort of mental state) could *be* a fact such as the fact that my friend needs help. Therefore, motivating reasons are not mental states but facts.

Thirdly, motivating reasons cannot be psychological states because of a neglected feature of the ontology of such states. States such as desirings and believings are relations aptly expressed with two-placed predicates like '__ believes that __' or '__ desires __'. So the view that motivating reasons are desires (say) is the view that these reasons are relations of desiring. This is most naturally considered an identity claim.[35] Now consider Sandra, who desires two things:

1. Sandra *desires* to eat the ice-cream
2. Sandra *desires* to drink the beer

Say Sandra's desire to eat the ice-cream was effective and she ate the ice-cream. Psychologism about motivating reasons says that her desire to eat the ice-cream is identical to her motivating reason. But because desiring is a relation, and Sandra stands in that relation both to eating the ice-cream (which she did) and to drinking the beer (which she did not), by the transitivity of identity, we have the claim that Sandra's desire to drink the beer is identical to her motivating reason to eat the ice-cream. But however plausible it sounds that her desire for ice-cream is what motivated her to eat it, it is clearly highly implausible that her desire to drink beer is her reason for eating ice-cream. This is not the right result.

One response would be to insist that it is Sandra's *token* desire that the identity claim concerns. Her reason is her desire *for the ice-cream*, which is clearly different to her desire *for the beer*. I have two responses. Firstly, trying to fix on token desires rather than the relation of desiring by fixing on the different objects of Sandra's desire will not help since all it does is says that Sandra's standing in the desiring relation to *this* is Sandra's reason, and her standing in the desiring relation to *that* is not. But *Sandra's standing in the desiring relation to something* is not a desire or a relation.

[35] The other possible reading is that it is predicative, however I will not explore that option here.

It is probably the state of affairs of Sandra's wanting something (to eat an ice-cream, say). This move therefore ends up identifying her reasons as states of affairs that somehow involve, but are not identical to, her desire. That move therefore collapses psychologism into an odd and unattractive form of externalism.

Secondly, although the claim that there are token, or particular, mental states is now rampant in philosophy,[36] in fact I think it is probably false. After all, if desire is a kind of relation, the claim that there are token or particular desires is the claim that there are token or particular relations. That is, there are relations which are not multiply instantiable or abstract; there are *concrete* relations. These are known as 'tropes' or 'property-instances' – concrete particular properties, distinct from the objects they are instantiated in, but spatially coincident with them, with causal powers all of their own (Maurin, 2016). I will not argue against the existence of tropes here. Rather, I will register two things: firstly, that I do not believe in them; secondly, that this way of avoiding my objection to psychologism involves very heavy-duty commitments in the metaphysics of properties. That should make anyone wary of taking this route. Therefore, unless we want (or already have) such commitments, we should not think of motivating reasons as mental states.

My fourth argument has a different sort of purpose. So far, I have argued against psychologism, and for the view that reasons are facts thought of as true propositions. This argument aims to show that the agent must *know* that *p* if the fact that *p* is their reason for doing something; therefore, *p* must be a fact.

Jennifer Hornsby has a neat argument for this claim (Hornsby, 2008a, p. 251). She constructs a Gettier case that looks like this. Edmund's normally reliable friend has told him that the ice is too thin at the middle of the lake for him to skate there, and so Edmund forms the belief *that the ice is too thin to skate in the middle*. Edmund's friend is reliable, so he is plausibly justified in his belief. And the belief is even true. However, Edmund's friend in fact has no view about whether the ice is too thin, and only told him that to make sure he stayed close by. It seems that Edmund's belief, although true, falls short of testimonial knowledge. This case is one in which the fact that the ice is too thin plays no role in explaining Edmund's staying close to the edge – he does not act *because* the ice is too thin, even though he believes that it is and does the appropriate things. There is no connection between what Edmund does and the reason there is for him to do it, and so we cannot say he acts for that reason. If this is all right, then Edmund must *know* that the ice is too thin in order for him to act for that reason. Hornsby's argument therefore shows that reasons must be actual *facts* because the beliefs that motivating reasons are the contents of must amount to knowledge.

These four arguments have aimed to show that motivating reasons cannot be psychological states and that they must instead be facts in the sense of true propositions. There is much to be said

[36] As a random(ish) sample, see Barlassina and Hayward (2019), Block and Fodor (1972), and Horgan (1984). Barlassina and Hayward even mysteriously talk of *tokening* states like pains, rather than merely of *token pains*, as if one does something called 'tokening' to one's pain. Note that, as Steward (1997, Chapter 4) points out, the token-type distinction is not the distinction between particulars and types of particulars since scarlet is a type of red, and yet scarlet is not a particular but a property. The token-type distinction is a logical distinction, not a metaphysical one. Nonetheless, since this goes unnoticed, talk of 'token states' is usually intended to be talk of *particulars*. For more, see Eric Marcus's (2009).

in response to these arguments and I am in no doubt that they are controversial. However, I will not enter into those controversies here. Nonetheless, I do want to say something about a standard objection to this view.

The view I have defended entails that all rationalising explanations are factive: if *S A*'s for the reason that *p*, then it must be a fact that *p*. A natural worry is that this imposes too high a bar on acting for reasons because sometimes agents seem to act on the basis of something that they take to be true, but which is actually false. Sally may have believed that a bear was chasing her, may have treated that proposition as a reason for her to run, and may have run. However, if there was no bear chasing her, then she did not run for the reason that a bear was chasing her. There was no such reason because there was no such fact.[37] Call this 'the problem of error': if rationalising explanations are factive, how are we to make sense of cases like Sally's?[38] This is, as Jimmy Lenman says, anti-psychologists' "biggest headache" (Lenman, 2009).

A full answer would take me far from my intended topic in this chapter. But I do want to try and dull the edges of the problem by questioning the challenge it is supposed to pose. I think that it is somewhat harder to make sense of the objection than it first seems, and that, seeing this, we can go some way to defusing the problem.

What *is* the problem of error? Is the problem that in a case of error like Sally's, we just cannot explain why she runs? If so, then there is a very easy answer: the explanation of why Sally is running is that she believes a bear is chasing her. This fact is not Sally's reason – it is not what motivates her – but it is a perfectly good explanation of why she is running.

Perhaps the problem of error is that if what Sally would say is her motivating reason is not a fact, then we cannot explain why Sally ran by appealing to what she had available from her own perspective, which is the point of rationalising explanations. Therefore, we would not be able to give a rationalising explanation of what Sally did. But this seems wrong. If we take Sally's perspective, since she thinks a bear is chasing her, and treats the proposition *that a bear is chasing me* as a reason to run, then we do have the materials available from within Sally's own perspective for understanding why she did what she did. The thing she believes is not a reason, but given the circumstances (Sally is camping in Bear County, there is a low growl from the bushes, rustling leaves, and the smell of wet fur), we can understand from Sally's own perspective why she ran.

Maybe the problem of error boils down to the following challenge: given all we know about Sally's situation and her motivations, there *just must* be something we can identify as her motivating reason. Perhaps this is true, but it is no longer obvious. We can explain why Sally ran, and we can make her running intelligible from her own perspective. Why do we *need* to also say that one of the things which explains why she ran was a motivating reason? It seems sufficient to say that she ran because she believed a bear was chasing her, and, from her perspective, there certainly seemed to be a bear. There is no problem if we cannot find a motivating reason in cases of error; nothing is

[37] This is different from saying that there was no such reason because there was a fact which was not a *good* reason, or a normative reason, to run. For discussion, see (Alvarez, 2010, pp. 142–143).

[38] For impressive discussions of this problem, see Alvarez (2010), Dancy (2000, 2008), Hornsby (2008a), McDowell (2013), Sandis (2013), and Stout (2009).

lost by accepting that, really, Sally had no motivating reason for doing what she did. It was just that, from her perspective, she *seemed to have a reason*. But that's the price of being in error.[39]

I think that this is a fair, if skeletal, response to the problem of error. However, it may not be necessary in all cases of error to say that a person acts for no reason at all. After all, given what I said in the previous paragraph, we are only forced to say that someone had no motivating reason when there are *no* candidates for being their motivating reason *other than* the false proposition. But often in cases of error, there *are* other candidates; people are rarely wrong about everything relevant to their practical situation. Imagine Bill who thinks he has two reasons to open the fridge: that the milk is in the fridge, and that the cheese is in the fridge. He opens the fridge, from his perspective, for both reasons. But he's out of milk, so it is false that the milk is in the fridge; therefore it is not a reason. But Bill still has a motivating reason to open the fridge: that the cheese is in the fridge. So although it is a case of error, Bill does not err in all respects, so he still acts for a reason (if not as many reasons as he thought). Sally may be in a similar position: although she does not act for the reason that she is being chased by a bear, it is plausible that her (apparent) *evidence* that she is being chased also counts as a reason to run. She knows there are bears in the woods, she smells wet fur, hears rustling in the bushes, takes this as evidence of a bear, and runs. Now she treats the false proposition about the bear as a reason, but it is not totally implausible that she (or someone in her position) also treats the (apparent) evidence of the bear as a reason to run. After all, even if Sally hasn't taken the evidence as conclusive and decided to withhold judgement about the presence of a bear, the evidence may still warrant running just in case. After all, the stakes are high: if one has putative evidence that there is a bear nearby, then one would be foolish not to do *something* to avoid the risk.

In this section, I have argued that reasons are facts. In particular, that they are true propositions, and that to act for the reason that *p*, one must know that *p*. Further, I have defended this view against the argument from error. That completes the defence of the first part of my view of rationalising explanations.

Section 3 - … and Reasons Explain Indirect Interrogatives (or Facts)

I have so far considered *explanans* in rationalisations. I will have much more to say about them in the next section. But I want to briefly consider the *explananda* in so-called 'action explanations'. The reason is primarily to gain some clarity over a very confusing issue. In fact, rather than saying what *explananda* are, my aim is to rule out some intuitive candidates and rule in some others.

[39] Jingbo Hu has suggested to me that the problem of error is that Sally is reasonable in running even if her belief is false. We may even say that Sally is *rational* in some suitably restricted (subjective?) sense. Of course, I agree Sally is reasonable, since it is reasonable and rational to act on one's apparent evidence, even when it does not amount to knowledge, in cases where there are high stakes. Therefore, we can admit all of this whilst denying that Sally actually had a reason to run

Actions are particulars; they are either events or processes (Alvarez & Hyman, 1998; Hornsby, 1980; Steward, 2012b). They have spatio-temporal locations, are unrepeatable, and they cause things to happen. They are things such as Sally's raising of her arm, Sally's running a mile, and Bill's opening of the fridge. Actions are different from what Hornsby terms 'things done', and which sometimes get called 'act-types' (Hornsby, 1980, Chapter 1). These are (unsurprisingly) the things we do. For example, we go to the shops, run from bears, and raise our arms. Things done can be done on different occasions by different people, and so are repeatable and multiply instantiable. They therefore bear the hallmarks of *properties*, universals, rather than particulars. So the things we do are sorts of properties, and our actions are our doings of those things. The relation between them is that an action, ϕ, is *S*'s doing of *A* if and only if ϕ has the property of *being an A-ing*. Of course, a single action ϕ always has many properties, *A-N*, so that whenever there is a person's action, they are doing all sorts of things: raising their arm; hailing a cab; contracting certain muscles, and so on.[40]

We might have expected typical action-explanations to be explanations of actions.[41] Recognising the distinction above, one might also wonder whether action-explanations may also explain things done. In fact, I think neither is the case.

Action-explanations often go like this: Sally ran because the bear was chasing her; Edmund skated by the edge because he thought the ice was too thin to skate in the middle; Bill opened the fridge because he wanted the milk. But these explanations have neither actions nor things done as their *explananda*. They do not give explanations of the event of (say) Sally's running, since they mention not an event but *what Sally did*. But they do not explain the thing Sally did, since what she did was run, and the property of *being a running* is not the sort of thing one can explain all by itself. It needs to be instantiated in an action of an agent's on some particular occasion. Anyway, even if we could explain *what* Sally did, that would not tell us *why she did it* (Sandis, 2012, p. 333).[42]

This suggests we must cast around for different candidates. Two theories from the literature on explanation suggest themselves. Firstly, we could follow Ruben and say that what we explain is *the fact that Sally ran* (Ruben, 1990). Secondly, we could follow Peter Achinstein and say that what we explain is *why Sally ran* (Achinstein, 1977).[43] The first is the proposition presupposed by the question 'Why did Sally run?', and the second is that question's indirect interrogative. The issue of which to prefer is very difficult, and to be frank I have not settled on a view. On the one hand, it can seem that the best way to distinguish some sorts of explanations is better accounted for by Peter

[40] I stated these views as assumptions of my thesis in my introduction.

[41] Hornsby describes this as the assumption that "We may move from knowing that we have an instance of 'action explanation' straight to thinking that we have an explanation of an action (event)" (Hornsby, 1997a, p. 142).

[42] Similarly, Hornsby points out that what we want to know is not why some event, ϕ, which is an *A*-ing, occurred, where its occurrence is seen as distanced from the fact that a particular person *A*-ed; we want to know why *this person A*-ed. See (Hornsby, 1997a, pp. 134; 139).

[43] Hornsby seems to suggest the first view in her 'Agency and Causal Explanation' (1997a). I think is also implied by much of what Donald Davidson and Elizabeth Anscombe say about the description-relativity of action-explanation (Anscombe, 1979; Davidson, 1970, pp. 147–148). Helen Steward defends the related idea that only facts about actions are 'up to us' at (Steward, 2012a, pp. 36–38). That said, the second view fits nicely with the question-theoretic discussion of explanations in Chapter 1.

Achinstein's view. For example, explaining *how* Sally ran is different from explaining *why* she did, even though both explanations have Sally's running as their topic. However, Ruben's view has the advantage of making explanation holistic in that both *explanans* and *explananda* are of the same ontological type: facts. This has the attractive feature that it allows us to explain the explainers, which one may think ensures the groundedness of explanations in a chain of further explanations.

However, I want to rescind from picking between these views. What is important is that we see that action-explanations explain neither actions nor things done. Rather, the objects of action-explanations are some sort of semantic item which takes what an agent does on some occasion as its topic, either as a fact or an interrogative question. Sometimes I will say that reasons explain facts about what agents do, and sometimes I will say that reasons explain why someone does something. This reflects my ambivalence on this issue, but ultimately has very little impact on my main arguments.

Section 4 – Rationalising Explanations Broad and Narrow (and How to Figure In Broad Ones)

I have been arguing that reasons for action are facts which favour doing things, and that when one acts for a reason, the reason explains why one acts as one does. I also argued that there are two requirements on acting for a reason: firstly, that one treat it *as* a reason for acting; secondly, that one knows one's reason. Finally, I argued that explanations of what agents do which appeal to the reasons which motivate them – the agent's own reasons – is what we home in on when we give rationalising action-explanations. This is because they make someone's doing something intelligible from the person's own perspective. In this section, I want to broaden our conception of rationalising explanations by showing that reasons are not the only important factors in rationalising explanations, and that this means there are really two sorts of such explanation.

Philosophers' focus on reasons has somewhat obscured the fact that other kinds of items have roles in rationalising action-explanations. As Hornsby says, "I might tell you why she refused the job, but leave you realising that there is more you could learn to help you see why accepting it was something she didn't do" (Hornsby, 1997a, p. 145). For example, say Sally decides to refuse a job offer of being a weapons developer for BAE Systems in favour of taking the job as the local vet for the reason *that imperialist warmongering is wrong and the local vets need help*. Even if we accept that this is the single motivating reason for which Sally made her choice, all by itself, it does not provide a full explanation of her choice. Many other items and conditions are relevant to Sally's decision and a full rationalising explanation of it, without which the decision would be unintelligible.

For instance, most obviously, Sally must have the rational capacity to know and respond to reasons, and to respond to them as reasons. Sally must also have some desires, values, beliefs, and knowledge, all of which are involved in various ways in her (short) deliberation about which job to take. What she values, believes, knows and desires affects which facts she takes to be relevant to her reasoning and which count as reasons from her perspective. Sally also has further intentions and life-plans which make sense of why she chose a job as a vet rather than as a weapons developer –

she intends to do good for animals rather than help arms dealers make money. Sally also has certain character traits that make these plans make sense – she is a kind and caring person, rather than a cold and callous one.

These features of Sally, as well as a host of others, help explain and make sense of her refusing the BAE Systems job and becoming a vet instead. However, they are not her reasons for doing so: they are not facts which she treats as reasons for her decision. They are also not merely explanatory reasons, since explanatory reasons are also facts, and the items I have listed are all properties. They are of the wrong ontological kind to be explanatory reasons. So we have a number of properties of Sally which do not explain why she took the decision she did, but are clearly *relevant* to any explanation of it. Moreover, they are relevant to the rationalising explanation of what Sally did. In the context of providing a rationalising explanation of Sally's decision, what knowing all this about her does is to provide a rich background against which her acting for that reason *makes sense*. In other words, knowledge of Sally's properties makes it rationally understandable for us, from Sally's point of view, why she acted for that reason.

To see this more clearly, imagine Sally's evil opposite, Jim, who has all the opposite values, desires, intentions and character traits, but who still makes the same choice as Sally. Let's also stipulate that his motivating reason is the same as Sally's. But once we have stipulated all this about Jim, we find his decision to refuse the BAE job and become a vet is incomprehensible. In order for it to make sense, we look around in Jim's life to find something that could be rationally relevant to his doing this – what is it that makes *this* reason something that *he* could treat as a reason? Sally's properties help us understand her decision; Jim's seem to positively obscure his. This shows that one's motivating reasons and one's other properties are tied together holistically. Davidson argued that "we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations, and the rest", the same is true of someone's doing something for a reason (Davidson, 1980d, p. 221).

Why is this? The reason is that rationalising explanations have a kind of dual character: when they give the reasons for which someone did something, they both home in on the person's *reason* and their *perspective*. And these properties of Sally – her beliefs and values; her desires and plans – *make up her perspective*. These properties constitute Sally's perspective on the facts, on their significance for deliberation, on which lives and projects are valuable, and which decisions are not to be made. So, Sally's reason for deciding as she does is only intelligible as an explanation because of her perspective. And that is why we find Jim's behaviour barely (if at all) intelligible. From the perspective of a cold, ruthless money-grabber, the fact *that imperialist warmongering is wrong and the local vets need help* just will not seem to favour declining the BAE job and becoming a vet. Jim's perspective is not one from which we can understand *this* as being *his* reason.

The point is that a person's desires, beliefs, knowledge, character traits, and more are relevant to a rationalising explanation of why they do things. They are relevant because they affect the agent's having of, and acting for, reasons by constituting their perspective on the world, values, and reasons. And since they constitute the agent's perspective on their reasons, they are rationally relevant to rationalising explanations of what agents do when they act for motivating reasons.

When an item is rationally relevant to an agent's acting for a reason but is not itself the reason for which they act, I will say that item *figures in* a rationalising explanation of why they act, or that it helps rationalise, or helps explain, why they did what they did. I will reserve the terms 'rationalises', 'rationally explains' and 'explains' for the reasons themselves.

Items which figure in rationalising explanations may do so in many different ways and there will be a long discussion of these ways in Chapter 6. But a few words now would be helpful. Desires seem to figure in rationalising action-explanations by presenting certain courses of action as good in some respect, and so highlighting certain reasons (or making there seem to be reasons to the agent) (Scanlon, 1998, p. 39). Some emotions seem to be like this as well. Indignation at injustice, for example, seems to sometimes reveal moral requirements. However, other emotions figure differently in rationalisations. Plausibly, I have a reason to avoid house spiders just because I am scared of them even though house spiders are not dangerous. So my fear *grounds* a reason rather than reveals an independent one. Something similar goes for pain. A pain in my knee provides me with a reason to avoid putting pressure on it. But the pain also provides me with knowledge of my reason, since one cannot be in pain and lack knowledge of the pain and the fact that it *hurts*. This is different again from virtue. If I am virtuously kind, then my kindness is a sensitivity to independent reasons, so that, when I act for those reasons, my kindness is rationally relevant to my acting kindly because it is the capacity to see certain facts as morally salient. Finally, there is the way that knowledge and belief figure in rationalisations. Belief that *p* and knowledge that *p* both put the subjects of these states in a position where, for all they know, they are able to act for the reason that *p*. So knowledge and belief can have a similar role in that they can both lead to someone's *A*-ing under the auspices that *p*. That is a similarity in the way they figure in rationalising explanations. There are differences, however, since someone who *merely* believes *p* will not be acting for the reason that *p*, whereas someone who knows *p* will. That is quite a significant difference in rationalising explanations invoking mere belief rather than knowledge. Further, someone who merely believes that their umbrella is in the hallway will fail to retrieve their umbrella by going into the hallway, whereas someone who knows that it's there will succeed in retrieving it (Hornsby, 1997b, pp. 116–117). I take it that this is partly due to a difference in the rational characters of belief and knowledge, stemming from their relation to justification and truth.

I have been trying to articulate a variety of ways that different properties that make up a person's perspective may figure in rationalising explanations. It should be clear that there is no single way that such properties must figure in explanations: some provide access to reasons; some merely *seem* to provide access to reasons; some, like pain, ground reasons and provide one knowledge of them; some, such as silly fears, simply ground reasons; and some, like the virtues, are *sensitivities* to independent moral reasons. Despite this diversity of *ways* that properties can figure in rationalising explanations, it is equally clear that they do all figure in such explanations without being reasons of any kind.

I say that these properties 'figure in' rationalising explanations. But if this is right, then we need a distinction between two kinds of rationalising explanation. This is because rationalising explanations, conceived narrowly as stating the reason for which someone did something, tend not

to make reference to these properties. If Sally is running because a bear is chasing her, the properties of Sally which are rationally relevant to her running for that reason do not literally figure in that explanation because the *explanans* does not refer to them – it just refers to the bear chasing her. If this were the only kind of rationalisation of Sally's running available to us, then we could not give the properties which make up her perspective a slot in a rationalising explanation.

This suggests that we must distinguish between the narrow sort of rationalising explanation and a broader, fuller, sort of rationalising explanation. The narrow sense is an explanation of why someone acts, where the explanation just gives the agent's own reason. The broad sense is an explanation of why someone acts which gives all the facts relevant to making the narrow explanation intelligible as a case of someone's doing something for a reason. So, although the narrow explanation of why Sally ran was that a bear was chasing her, the other rationally relevant materials contribute to a broader explanation: Sally knows a bear is chasing her; she knows how to get away from the bear; she fears the bear and wants to live; she does not want to shoot it. These facts which make reference to aspects of Sally's perspective constitute the broad rationalising explanation which makes the narrow explanation intelligible. And aspects of Sally's perspective figure in the broad explanation because the broad explanation makes essential reference to them.

This bears on how to characterise the division between 'motivating' reasons and 'merely explanatory' reasons. In Sections 1 and 2, the way I introduced the idea of rationalising action-explanations suggested that rationalising explanations are concerned exclusively with a person's motivating reasons. I gave an example of Jill, who is shy, who sits at the back of the class because she is shy, but whose own reason is that she won't get picked on at the back. I said that both reasons explain what Jill does, but that the rationalising explanation gives Jill's own reason; the fact that Jill is shy is merely explanatory.

Now we can see that this is true only when we construe the rationalising explanation narrowly. Since the fact that Jill is shy is not her reason for sitting at the back of the class, that fact does not narrowly rationalise her sitting where she does. It is true that this fact is a 'merely explanatory', rather than 'motivating', reason. However, when we construe the rationalising explanation of what Jill does broadly, or fully, this explanatory reason does figure in the full rationalisation of what she does. This is because the fact that she is shy is part of what explains *why* the fact Jill treats as her motivating reason – that she won't be picked on at the back of the class – is a motivating reason for her. Jill's shyness is rationally relevant to her acting for the reason she does because *being shy* makes up part of her perspective on the world. Being shy puts one in a state of emotional vulnerability, making social situations intimidating and off-putting. For someone who is not shy, the fact that they wouldn't be picked on at the back of the class will not stand out as a reason for them to sit there; for a student who wants to show off, it will seem to be a reason to *avoid* sitting at the back. So knowing that Jill is shy adds to the intelligibility of her sitting at the back of the class for that reason. Therefore, explanatory reasons in which aspects of a person's perspective figure often form a key part of the story when it comes to understanding why they do what they do. Similar things can be said about facts about what agents believe or want or love, and their traits and their foibles.

The distinction between narrow and broad rationalising explanations, is, I believe, crucial. As is the idea that properties may figure in broad rationalising explanations. In the final section, I will describe how these ideas help us state the Rationalising View of Habit properly, and how they determine what would count as an argument for it.

Section 5 – The Rationalising View of Habit

Habits are properties, sorts of behavioural tendencies which manifest in habit-bearer's doing, on occasion, what they do habitually. Since habits are properties, they cannot explain anything; explanation is a relation between facts (or between facts and interrogative questions). However, because habits manifest in habit-bearers' behaviour, the notion of habit is clearly an explanatory one. Therefore, we must say that habits *figure in* explanations of facts about why habit-bearers do what they do. Apart from the metaphysical picture of explanation I have framed this in terms of, this should not be controversial.

This allows us to frame the controversy between the Rationalising and the Non-Rationalising Views of Habit. On the Non-Rationalising View, although habits figure in explanations of what habit-bearers do, habits do not add anything to the intelligibility of the habit-bearer's acting for a reason.[44] The habit does not ground the reason for which they act; it does not shape their appreciation of their reasons; it does not present new reasons. My habit of going to *Gilmour's Café* in no way forms part of my perspective on things in a way which could be rationally relevant to a narrow rationalisation of why I go there (when I go there out of habit). That is, there is no broad rationalisation in which habits essentially figure.

In contrast, the Rationalising View says that habits figure in broad rationalising explanations of why habit-bearers do what they do when they act out of habit. That means that, like the other properties which figure in rationalising explanations but which are not themselves reasons, habits contribute to rationalising explanations by shaping the landscape of reasons which the habit-bearer possesses, or altering how they appreciate their reasons, or presenting some reasons as particularly salient, or some other way that a property might figure in such an explanation.

I wish to state the Rationalising View quite strongly, as the view that *whenever* a habit manifests it has a rational role and figures in a broad rationalisation explaining why the habit-bearer acted as they did. My aim in this thesis is to argue for this view. However, it is worth pointing out that a weaker view is available. The weaker view is that when habits manifest they *sometimes* figure

[44] At least not in the normal case. There are odd cases where someone may do something for the reason *that they habitually do it*. Perhaps they have been offered some money to do something they are in the habit of doing for that very reason. In that case, their habit figures in a broad rationalising explanation of what they do by contributing to the intelligibility of their acting for a reason, since the reason is a fact about their habit. But that is not what the dispute between the Rationalising and Non-Rationalising View is concerned with: they are concerned with ordinary cases of habitually going for a run, or picking one's nose, or interrupting people. Of these normal manifestations of habit, the Non-Rationalising View says that the habit contributes in a purely explanatory manner to the explanation of why habit-bearers act out of habit; the Rationalising View disagrees.

in broad rationalisations, but sometimes they may not. That is, there are some cases where habits have rational roles in broad rationalisations and other cases where they have some other explanatory role. Call this view the 'Weak Rationalising View of Habit' (I reserve 'the Rationalising View' for the stronger account). Insofar as the Non-Rationalising View represents total scepticism about habit's aptness for rationalisation, the Weak Rationalising View is inconsistent with it. Therefore, if one wanted to defend the stronger Rationalising View, arguing for the Weak View may be a good intermediary stepping stone. This is really my strategy. By the end of Chapter 3, we will have an argument for the Weak View from which my argument for the stronger view proceeds in Chapter 4.

I think that this clarifies the terrain of the debate. My discussion in this chapter also suggests a way of arguing for the Rationalising View of Habit. In the previous section, I argued that the properties of a person which constitute their perspective may figure in broad rationalising explanations of what they do. Now, the properties I listed were a somewhat rag-bag bunch. Some were paradigmatic mental states, like belief, desire, and pain. Some clearly have something to do with the mind, but whose status as genuine mental states is disputed since they depend on non-mental factors, for example, knowledge, which depends on truth.[45] And some of the properties were sorts of behavioural dispositions or tendencies, such as the virtue of kindness, and the trait of shyness. I will call all of these properties 'mental properties', very broadly construed, and for want of a better term.[46]

Now, an unprejudiced list of these properties should include habits. After all, even if one is not as impressed by habit as by virtue or knowledge, habits are clearly in the same rag-bag of mental properties. They are clearly not 'purely physical properties' available from a physicist's impersonal view of the world. Habits are integrated deeply with a network of properties which only come into view when we see human beings (and other animals) as *people* (or at least as *agents*). After all, habits are habits of *action*, whose bearers are *people*, and which are formed because those people have a *personal history* of acting, where they acted as they did, most often, for *reasons*, with *desires* and *intent*. The concept 'habit', then, is deeply integrated into our personal point of view, and a proper grasp of the concept requires having these connections and concepts available to one. For this reason, I think that we should quite clearly admit habits into the same category of properties as belief, knowledge, virtue, and pain.

This may look like a direct argument for the Rationalising View: habits are in the category of properties which can figure in broad rationalisations; therefore, they can figure in broad rationalisations. But the argument is not so simple. For one thing, the large category of mental properties is not *defined* as being a category of properties which can figure in broad rationalisations. It is just the case that these properties tend to be peculiarly apt to figure in broad rationalisations in virtue of their connections to the agent's perspective. The Non-Rationalising Theorist can admit habits into that category, and yet deny that they figure in broad rationalisations. This is because

[45] See Fricker (2009), Hyman (2006), and Williamson (1995) for the debate about whether knowledge is a state of mind.
[46] I want to suggest (very tentatively) that the term 'mental property' has a roughly similar extension to P. F. Strawson's term 'P-predicate' (Strawson, 2002, Chapter 3). At least, it is supposed to do similar yeoman's work.

properties in that category *can* figure in *non*-rationalising explanations of why someone does something. For example, think of Davidson's climber whose desire to let his friend die so unnerves him that he lets go of the rope (Davidson, 1980b, p. 79). The climber's desire figures in an explanation, but not a rationalising one. So the Non-Rationalising Theorist of Habits can say that habits may be mental properties, but the only explanations they ever figure in are non-rationalising ones.

But Davidson's example is instructive. What goes wrong for the climber is that his mental properties conspire in an explanation of why he lets go of the rope *where he does that unintentionally*. That is part of the reason we feel willing to say that his desire and fear jointly acted as a kind of sheer causal compulsion, and why we are disinclined to say that the climber *acts* in dropping his friend. But think of a case in which the climber *intentionally* killed his friend, and say that the explanation includes his desire to let him fall and his fear at his own desire. As it stands, we don't have enough pieces to form a full explanation, since it is not sufficient to explain why someone did something intentionally to just state that they had a desire to do so and a fear about their desire. When we introduce intention, we are no longer thinking of a case of causal compulsion which bypasses the agent's rationality, which means that the pieces we *do* have most likely form parts of a broad rationalisation of the climber's killing his friend. They form part of his perspective on the world such that, in dropping his friend intentionally, he acts for a reason. That is, when mental properties figure in explanations of what someone does *intentionally*, we seem to be compelled to think of the explanation they figure in as being a broad rationalisation which comes hand-in-hand with a narrow rationalisation.

Now, many philosophers deny that we can read back from the fact that someone acted intentionally to the fact that they acted for a reason (Alvarez, 2009; Anscombe, 2000, p. 23; Heuer, 2014, pp. 293–294). I will answer those arguments in the next chapter. But for the moment, if we grant that we can (as the consideration of Davidson's climber suggests) move neatly from intention to rationalisation, then the connection between acting intentionally, broad rationalisations, and mental properties suggests an argument for the Rationalising View of Habit. We need, though, to put the pieces together. Firstly, I have argued that:

1. Habits are mental properties (broadly construed);

I argued for this by appealing to the way that habits are embedded in a network of other mental properties which only appear when we think of their bearers as persons, or agents. Indeed, I think that even Gilbert Ryle would accept this. After all, he does include it in his discussion of the concept of mind. (1) is really the default view, and an argument against it would have to be very compelling indeed to make us doubt it.

Secondly, I have suggested that:

2. If mental properties figure in explanations of why someone does something *intentionally*, then they figure in broad rationalisations of why they do that thing;

This was suggested by my consideration of Davidson's climber, and it seems to me an incredibly plausible principle. Obviously, though, it is controversial, so I will argue for this in detail in the next chapter. But I think we can grant that it is motivated and at least plausible. If it is true, then, taken with (1), it opens the door for the Rationalising View. All we would need is one further premise:

3. Whenever habits figure in explanations by manifesting, they figure in explanations of why the habit bearer does something intentionally.

If I can secure that premise, then it follows that:

C) Therefore, whenever habits manifests, they figure in broad rationalisations

This is because habits, when they manifest, figure in explanations of something the habit-bearer does *intentionally*, and that, together with the fact that habits are mental properties, jointly entails the view that habits figure in broad rationalisations. So the Rationalising View is true.

That is a rough sketch of the argument in the next three chapters. I attempt to make the case for the Rationalising View by showing that there is a connection between acting intentionally and rationalising explanations, and then by showing a connection between acting habitually and acting intentionally.

Since my argument runs via some claims about intention, that is Chapter 3's focus. In it, I argue for two things: an Anscombean view of doing something intentionally; and the view that one acts intentionally if and only if one acts for a reason. The purpose of the Anscombean view is just to fix what I mean when I talk of someone's doing something intentionally. However, it is far more important to my overall aim that I defend the biconditional since that is ultimately what justifies Premise (2) above. Indeed, it will emerge at the end of Chapter 3 that we can argue from (2) to the Weak Rationalising View I mentioned earlier. Whilst it falls short of the Rationalising View proper, it is clearly a significant step. Then, I use it as a stepping stone to the Rationalising View by arguing, in Chapter 4, for Premise (3).

CHAPTER 3

INTENTION, PRACTICAL KNOWLEDGE, AND ACTING FOR A REASON

In the last chapter, I defended a number of views in order to state, and show how to argue for, the Rationalising View of Habit. In particular, I argued that rationalising explanations come in two kinds: narrow and broad. Narrow rationalisations explain why someone did something just by stating the agent's own reason for acting. Broad rationalisations explain why the agent acted by citing all sorts of facts that are rationally relevant to the agent's acting for the reason which narrowly explains what they did. And I argued that mental properties such as a person's beliefs, virtues, pains, and desires *figure in* a broad rationalisation when the broad rationalisation makes explanatorily essential reference to them. I argued, from this, that the way to state and think about the Rationalising View of Habit is as the view that, whenever someone acts habitually, their habit figures in the broad rationalisation of what they do. This means that the habit is rationally relevant to habit-bearer's acting for, or having of, motivating reasons.

At the end of the chapter, I suggested a way of arguing for the Rationalising View that runs via some claims about acting intentionally. The thought was that if mental properties figure in explanations of what someone did *intentionally*, then they figure in broad rationalisations of what they did. I motivated this by considering Donald Davidson's famous case of the climber. The thought, roughly, is that in Davidson's original deviant case, we do not see the climber's desire and fear as figuring in a rationalising explanation of his dropping his friend because he doesn't do it intentionally. If we imagine a case where he does intentionally drop his friend, those states are best seen, not as bypassing his rationality, but as partly forming his perspective on his reasons. This motivates the idea that where someone acts intentionally, the explanatory mental properties figure in a broad rationalisation of what they do.

Therefore, in this chapter, I focus on the notion of doing something intentionally, and on its connection with doing something for a reason. For, whilst the above line of thought makes a plausible connection between intention, reasons, and mental properties, much more must be said in order to argue for the claim I have just made. In particular, this is because many philosophers deny that we can read back from the fact that someone did something intentionally to the claim that they did it for a reason. If they are right, then however tempting my line of thought is, it would be wrong. So I will spend much of this chapter arguing that there is a strong enough link between intention and acting for reasons to justify my claims about the rationalising role of mental properties.

This chapter is split into two parts. In the first part, (Section 1-3) I defend a view of what it is to do something intentionally. The point of doing this is primarily so that when I make claims which involve the notion of intention, we know what I mean. For example, if I did not give an account of what it is to *A* intentionally, it would be much harder to assess (for example) claims of the form 'if *S A*-s intentionally, then *p*'. I think this is important, and the topic of intention is interesting in its own right. However, this part is less important for my overall project than the second half. Whilst I will sometimes fall back on the Anscombean view later on in the thesis, it does not (to my mind) form an integral part of my case for the Rationalising View of Habit.

Part two (Section 4-6) is more important for my overall argument. This is because, in it, I argue for the claim that someone does something intentionally if and only if they do it for a reason. The point of this is to show that there are very tight connections between intention and motivation by reasons, and that these connections are tight enough to licence the claim that if a mental property (such as belief, desire, or virtue) figures in an explanation of why someone does something intentionally, then the explanation it figures in is a broad rationalisation. Therefore, I will spend most of part two focussing on the biconditional, and I will end the chapter by linking the biconditional to the topic of mental properties and broad rationalisations. Indeed, in Section 6, I show that we already have the material to argue for the Weak Rationalising View.

The chapter is structured as follows. In Section 1, I motivate my Anscombean treatment of intention in terms of practical knowledge of what one is doing. Then I deal with two classic objections to that view. In Section 2, I answer a worry about whether knowing what one is doing is necessary for doing something intentionally. In Section 3, I respond to a worry about whether practical knowledge can have the content it is supposed to have. In Section 4, I defend the first half of my biconditional: if *S A*-s for a reason then *S A*-s intentionally. In Section 5, I defend the other half: if *A*-s intentionally then *S A*-s for a reason. I do this by considering and rejecting two types of apparent counterexamples, cases of acting just because one feels like it (Section 5.1) and cases of expressive or 'arational' actions (Section 5.2). In Section 6, I show that the biconditional's truth lends significant support to the idea that mental properties figuring in the explanation of why someone *A*-s intentionally thereby figure in broad rationalisations of their *A*-ing. I then give an argument for the Weak View which follows directly from this fact.

Section 1 – Motivating the Anscombean View

What is it to do something intentionally? I want to motivate a way of answering the question which is broadly Anscombean[47]: *S* is *A*-ing intentionally if and only if *S* has practical knowledge that

[47] I say 'broadly Anscombean' because the view is now as much associated with neo-Anscombeans as it is with Elisabeth Anscombe herself. For example, see Haddock (2011), Lavin (2013, 2015), McDowell (2011c, 2011b), Small (2012), and Thompson (2008). I am certainly not going to be attempting very much Anscombe exegesis here. This thesis is hard enough already. My own presentation of the motives for an Anscombean view of intention is indebted to (Small, 2012, pp. 137–139).

she is *A*-ing.[48] The view is that a special kind of knowledge of what one is doing is both necessary and sufficient for doing it intentionally: one must have non-observational, practical knowledge of what one is doing. The motivation for the view comes from the following sorts of considerations. Suppose I am standing on the hose, stopping the water flow, but am distracted by talking so that I do not know this. Then, it seems I am not blocking the water supply intentionally. And if, after an urge to dance has washed over me, I watch in horror as a vase I accidentally knocked smashes, I cannot be said to have intentionally smashed the vase. And if I did not see, but inferred from the loud crash that I had smashed the vase, I also cannot be said to have smashed it intentionally. And if I were told it had smashed, the same would go. It seems fairly intuitive, then, to think that these kinds of knowledge – what Anscombe calls 'observational knowledge' (Anscombe, 2000, pp. 13–15) – are not sufficient for making something one does intentional.

However, it also suggests that if one *finds out* that one is blocking the water supply, or has smashed a vase, then one cannot be doing, or have done, so intentionally. Therefore, it seems that *some* knowledge of what one is doing is necessary for one's doing it intentionally. But all of the observational ways of knowing are ways of discovering, or finding out, that one is doing something. If I hear the smash and infer that I have smashed the vase, then I make a discovery about the effects of my dancing. So not only is some knowledge of what one is doing necessary to be doing it intentionally, but the knowledge cannot be observational knowledge, since one cannot do something intentionally and yet *discover* that one is doing it.

Indeed, in the normal course of things, I do know what I am doing when I do something intentionally. If I am intentionally making an omelette, I know I am; If I am intentionally running for the bus, I know I am; and if I am intentionally taking aim at my enemy, I know I am. But this knowledge seems to be different from the kind discussed above. I do not work out that I am making an omelette as I go, as if I were watching someone else. And I do not know based on a funny sort of self-testimony ('Believe you me: buddy, you're making an omelette!'). And I do not see myself make an omelette and know in a non-inferential, perceptual, way that I am. I seem to know that I am making an omelette without any of those sources of evidence. Anscombe aptly calls this knowledge 'non-observational' in contrast to the observational kind based on perception, inference, or testimony (Anscombe, 2000, p. 13).

So, if someone, *S*, lacks knowledge that they are *A*-ing, they are not *A*-ing intentionally. But the knowledge they have cannot be merely observational knowledge. Therefore, for *S* to be *A*-ing intentionally, *S* must non-observationally know that they are *A*-ing. So non-observational knowledge of what one is doing is necessary for doing it intentionally.

It is fairly simple to see how this can be turned into a sufficiency condition as well. All we have to do is see that the content of this non-observational knowledge would, if expressed, be the content of an expression of intention. What I know non-observationally when I am intentionally

[48] As many have argued, *S*'s knowledge that she is *A*-ing must be first-personal. It must be knowledge that *she herself* is *A*-ing; in Casteñeda's coinage, that she* is *A*-ing (Castañeda, 1966). In more common tongue, *S* must know *that I am A-ing*. No other mode of presentation of *S* may be slotted into *S*'s thought. However, having acknowledged this, for ease of expression I will proceed to simply say that *S* must know that she is *A*-ing, or that *S* is *A*-ing. Please read these as expressing the relevant sort of first-person knowledge.

making an omelette is *that I am making an omelette* and this is what I would say if someone asked me what I was doing intentionally – 'Why, I'm making an omelette'. Things one could say about what one is doing based on observation would not express one's intention even if true (McDowell, 2011c). Therefore, non-observational knowledge of what one is doing is appropriate for fixing which thing one is doing intentionally (or, fixing the description under which one's action is intentional).[49] Call the non-observational knowledge that would be expressed in expressing one's intention 'practical knowledge'.

This becomes a sufficiency thesis because if what one knows practically is the content of one's intention in acting, then it is in virtue of that knowledge that doing something comes under the description 'execution of intention' (Anscombe, 2000, p. 88). The practical knowledge *that I am A-ing* is sufficient for its being the case that I am intentionally *A*-ing. So the necessary condition on intentionally *A*-ing is also sufficient – one is intentionally *A*-ing if and only if one practically knows one is *A*-ing. Now, this consideration hardly rises to the level of a proof of the sufficiency condition. Much more would be needed to argue for it. For my purposes, though, it is enough that the claim of sufficiency is motivated and somewhat plausible.

In the next two sections I will canvas two standard objections to this view. Firstly, I consider whether the practical knowledge is necessary (*Section 2*). Secondly, I consider whether it is sufficient (*Section 3*).

Section 2 – Counterexamples to the Necessary Condition

The main objection to the claim that practical knowledge is a necessary condition comes from the following remark of Davidson's: "A man may be making ten carbon copies as he writes, and this may be intentional; yet he may not know that he is; all he knows is that he is trying" (Davidson, 1980b, p. 50). The idea is that one can lack knowledge that one is making ten carbon copies, say, because the activity is difficult, or the success of the activity is uncertain from one's own point of view, and yet one can still (in some sense) succeed because one intended to. So we can say that one intentionally makes ten carbon copies. Why should the lack of knowledge of success matter to whether one is intentionally making ten carbon copies? If it does not, then knowing that one is *A*-

[49] Of course, there are well-known views of the content of putative mental states called 'intentions-in-action' which are much more complicated than this simple picture. For example, John Searle's (1979) view is that they involve self-reflexive contents which specify their own causal role. However, I will not argue against Searle on this. The main reason is that his complex contents are necessary (perhaps) only in the context of background assumptions about the nature of intentional action which I think are neither right nor relevant. For example, only if one thinks that the action theorist's task is to say how bodily movements are caused and causally guided by mental states will one be compelled to take Searle's arguments seriously. Equally, however, if that assumption is right, it may very well be that intentions-in-action are required for an account of the causation of intentional actions. However, we would need further assumptions to show that the mental states causing and causally sustaining an action, ϕ, which is someone's *A*-ing intentionally, must figure in an account of *what it is to A intentionally*. Compare: the brain is powered by the oxidation of glucose, so the oxidation of glucose causally sustains our thinking; however, the oxidation of glucose does not enter into an account of *what it is* to think.

ing is not necessary for *A*-ing intentionally (*a fortiori*, having practical knowledge that one is *A*-ing). Practical knowledge can fall short of what one does intentionally.

There are a number of plausible reactions to this. Firstly, we can take the approach recommended by Michael Thompson (2011) and Adrian Haddock (2011). They point out that one can *be making* ten carbon copies even if one never succeeds in marking the bottom page, just as one can *be making* a cake without ever succeeding. If the content of practical knowledge is in this progressive form, then one can know that one is making ten carbon copies without knowing that one has succeeded or is succeeding.[50] One can even doubt that one will succeed whilst knowing that one is doing something – I can be intentionally playing the Master of Puppets solo whilst doubting that I have the endurance to keep it up till the end.

I think this is a helpful start, but Davidson is really best read as saying that one can intentionally succeed in having marked the bottom page without knowing that one has marked it. It is not clear how the Thompson-Haddock response can guarantee that one *can have intentionally marked* the bottom page when one's practical knowledge is simply progressive rather than perfective.

But perhaps we should not guarantee that thought. Thompson (cryptically) suggests a useful line of thought. He says:

> "[T]he more ordinary case [in contrast to Davidson's] is like this: you write on the top sheet, trying to make a good impression to get through all the carbon, then look to see if your impression made it through to all of them. If it did, you stop. If it didn't, you remove the last properly impressed sheet and begin again. If necessary, you repeat. Even the man who has to go through five stages is all along, from the first feeble impression, making ten copies of the document, and he knows it, all along. [...] Well, for [Davidson's man], the making of the inscription is like the buying of a lottery ticket. You can say he made ten copies intentionally if you like, but it will not be an illustration of the topic of Anscombe's book, any more than lottery-winning is when you bought the ticket with that aim." (Thompson, 2011, p. 210)

Thompson's thought is that a part of acting intentionally is trying to ensure that one succeeds in achieving one's aim; one is trying to secure an aim by using means and methods at one's disposal. The reason that I did not intentionally win the lottery just because I bought a ticket with the intention of winning, and won, is that my winning was determined by processes outside of my control; it was simply left up to luck and nature. I did nothing to ensure my success. One can only win a lottery intentionally by rigging it – by intentionally messing around with the lottery machines and bribing some officials (not that this would be a genuine lottery anymore). One requirement on doing this is epistemic: I must make sure that everything is in place according to the plan. If I had done that in a comprehensive and conscientious way, then perhaps we can say that I intentionally won the 'lottery'.

[50] Incidentally, the form I introduced it in *Section 1* was progressive: 'I am making an omelette'. The progressive form of practical knowledge has not been introduced to save the view but is part of its presentation.

However, if I had tried to execute a plan, but clumsily, without proper precautions, where I did not *know* that the machine was rigged and the officials bribed, but where they luckily were, then I would once again have failed to intentionally win the lottery.[51]

The problem is that Davidson's carbon-copyist is not a conscientious one. He is like the person who botches his plan but comes up trumps. He does not keep himself in the epistemic position with respect to his goal that is required for ensuring success. So he may have successfully made ten carbon copies, and his intending to do so clearly played a role in that, but we should not say that he intentionally marked the bottom sheet – he left it up to luck. Of course, we cannot expect that agents always *guarantee* their success; a certain degree of luck must be allowed in a case of doing something intentionally. An archer may intentionally hit the bullseye even though they were lucky that a large gust of wind did not sweep their arrow away. But they hit the bullseye intentionally despite their luck because they took all the appropriate care and considerations into account, exercised their skill well, and took a good shot.[52]

This way of dealing with Davidson's case denies that we must accommodate the idea that one can have done something intentionally without knowing it; it is not an insight, but a mistake. And it suggests a strategy for dealing with the ever more complex Gettier-esque examples which may spring forth: in working out whether a case is one of practical knowledge falling short of what one does intentionally, we should stay alive to the fact that knowledge has an essential role in guiding action so that agents ensure their own success. After all, succeeding is usually the point of acting, and those who deny practical knowledge is necessary for intentional action may sometimes lose sight of that fact.

Section 3 – Practical Knowledge and Observational Knowledge

Another standard challenge to the Anscombean view is to question whether practical knowledge of what one is doing can be non-observational. The difficulty is in seeing how I can know *that I am making an omelette* or *that I am writing my name on the board* without resorting to perception. The Anscombean view presents perceptual knowledge as observational and so unable to determine what it is a person does intentionally. But this is in tension with the fact that one can, presumably, intentionally do things that one can only know one is doing by means of the senses – making an omelette; writing one's name on the board.[53]

---

[51] A similar clear case is of a careless driver who narrowly misses a pedestrian whilst looking at his phone, but who all the while intends not to hit anybody. He does not intentionally miss the pedestrian, surely, because he had no idea whether or not there was one in his path.

[52] Exactly *how* to accommodate these facts in an account will be difficult. I suspect that one may draw on the massive literature in reliabilist virtue epistemology to figure it out, something that Carlotta Pavese is beginning to do with great effect, for example in (Pavese, forthcoming-a, forthcoming-b, 2016b, 2016a).

[53] This challenge should not be confused with Davidson's. It also not the question of how mistakes can be possible on the Anscombean view. As I have presented it, the worry is a more circumscribed concern about how it is possible to have a certain kind of knowledge (non-observational, practical knowledge) of a certain realm of facts (facts about what happens in the environment).

This is a challenge to the claim that practical knowledge is sufficient for doing something intentionally because it is a worry about whether non-observational knowledge can have content which ranges over all of what we do intentionally – doesn't it need supplementation from perception, say? If practical knowledge is not allowed to take any of its content from perceptual or inferential sources, the Anscombean seems to be pushed to the interioristic view that all we do intentionally is try to act or will something to happen. Some have thought this view attractive, but I agree with most philosophers that it is best avoided.[54] My strategy will be to show that the objection's general line of thought, though appealing, is actually very hard to put into a specific form which is genuinely worrying. My hope is that attending to how the challenge must be stated will show that it cannot be stated in a sufficiently worrying way.

One tempting way of putting it is like this: 'the apparent content of practical knowledge in cases where someone intentionally writes one's name on the board is *that I am writing my name on the board*; but that fact is only knowable via perception; therefore, it is perceptual knowledge, not non-observational practical knowledge at all'.

But this argument is overly hasty. My name is 'Will', so as I write my name on the board, I will see those letters appear beneath my chalk in sequence. But does perception tell me that I'm writing *my name*? I could, after all, be writing 'willow' or 'willpower'. If I were writing those words, my perceptual experience would be the same as it is when I am writing my name. When I finish, does perception tell me that I have written my name, or that I have written the verb, '(to) will'? Perception is silent on these different options; at no point does perception have the resources to provide me the knowledge *that I am writing my name on the board*. Barry Stroud makes the same point when discussing how you know you are intentionally walking across the Golden Gate Bridge:

> "What could you be aware of while having [the experience of walking across the Golden Gate Bridge] that supports or grounds a claim to know that that is what you are doing intentionally? What could you be aware of in what is going on that would distinguish between your intentionally walking across the bridge from your intentionally walking only half-way across, or intentionally walking on the bridge for only five minutes, or intentionally walking from San Francisco to Sausalito by the shortest route, and so on?" (Stroud, 2013, p. 6)[55]

The idea is that it cannot be by perception that one knows these facts since perception is silent on which of these things is happening. So the putative content of the kinds of practical knowledge at issue cannot be simply 'gotten' or 'transferred' from perception as the tempting objection had it.

A more subtle statement of the challenge goes like this: 'If practical knowledge is non-observational, how can its contents have as constituents facts which one can only know through observation?' For example, although the previous objection failed, it is true that I can only know that

[54] See Hornsby (1980, Chapter 4) for arguments against such views of action.

[55] There is a nice discussion related to this point in Moran (2004, p. 53).

the board is in front of me, that the chalk is touching the board, and that I am holding the chalk by observation. So the putative practical knowledge *that I am writing my name on the board* involves claims about the world which, if they are known, can only be known observationally. Therefore, the worry goes, the knowledge cannot be non-observational because some of the parts of its contents are known observationally.

But I think this is wrong for two reasons. Firstly, it is not clear whether the *content* of putative practical knowledge *does* involve things that can only be known by observing. Although *that I am writing my name on the board* does involve a claim about the board, which part of that content is supposed to only be knowable by observation? That I am writing, that it's my name, or that I am writing on the board? To me at least, it is very unclear which part of the content is the target of the claim. Clearly, the putative practical knowledge presupposes observationally known things, such as that there is a blackboard, or (in Stroud's case) that I am on the Golden Gate Bridge. But the fact that one piece of knowledge presupposes some observational knowledge does not entail that the content of the first piece of knowledge *includes* the content of the second. Therefore, it is not obvious to me whether the claim that the putatively practical knowledge in question does have content which is observationally known.

But perhaps I am wrong about this. If so, this brings me to my second response, which is to reject an assumption the objection makes: that it is impossible to know that *p*, where *p* has knowable propositional constituents, where one's knowledge that *p* has one form, and one's knowledge of *p*'s constituents has another. This assumption drives this version of the objection, since it is supposed to be the inclusion of pieces of observational knowledge amongst putatively non-observational knowledge which blocks our ability to say that it really *is* non-observational.

To reject the assumption, consider certain claims of *a priori* knowledge. Say I know that *p* by *a posteriori* means, and I know that *q* also *a posteriori*. Well, since I know that if (*p*, *q*), then (*p* & *q*), I am now in a position to know that (*p* & *q*). This means that, if I do join the dots and infer (*p* & *q*) from (*p*, *q*), then I know that (*p* & *q*). But, importantly, I know that (*p* & *q*) *a priori* – I know it totally independently of experience, just by using the rules of deductive inference. It does not matter that the constituents of the conjunction are known *a posteriori*, or whether they are contingent, or whatever. The conjunction is known *a priori*, and the conjuncts, *a posteriori*. So it does not follow that knowledge of one form cannot have constituents that are instances of knowledge of another form. Therefore, it does not follow from the fact (if it is a fact) that putatively non-observational knowledge has contents gained observationally that it is therefore not non-observational.

Sometimes, when philosophers discuss the challenge posed by needing to reconcile practical knowledge with observational knowledge, it sounds as if what is at issue is that *the result* of what I do cannot be known to have happened without observation. For example, Richard Moran, in stating the problem, says "the question I am trying to press here is [...] how Anscombe can claim that my *knowledge* that I am doing something can be non-observational, when *what I do* includes, for example, the window I am opening or the words I take myself to write" (Moran, 2004, p. 52).

On one reading, this is just a version of the previous worry that I argued against, albeit not a particularly pressing formulation of it. It is not pressing because it does not even claim that any of

the content of putative practical knowledge is actually *known* observationally; it only claims that practical knowledge is often formulated in terms which refer to things such as windows and words, knowledge of which is typically (or often) observational. But why should it be concerning in the slightest if practical knowledge involves such concepts, if they are unproblematically in the conceptual repertoire of the agent?

But I think given what Moran says soon afterwards, another reading is more apposite. He says:

> "An agent will have such awareness of 'opening a window' *only when* an actual window is getting opened, and similarly for [Anscombe's] examples of pushing a boat out or writing with a pen. Hence if the pen has run out of ink and no actual writing is getting done, then the agent's non-observational awareness is *mistaken*. [...] But as we've seen, Anscombe rejects the idea that in the *successful* cases, where a boat *is* getting pushed out, or writing *is* being produced, all that is really known by the agent immediately is something like the feelings of pressure in one's hands, or something else which excludes the actual boat or the window or the pen." (Moran, 2004, pp. 52–53)

Moran's use of passive locutions, which I have underlined, indicates that he is worried that the Anscombean cannot say that practical knowledge of what one is doing (when one is doing it intentionally) is *non*-observational because knowledge of the *results* of what one is doing must be *ob*servational. After all, what I do is write some words, the result of which is that writing has gotten done; I push the boat, which results in the boat's having been pushed. Whilst Moran is here concerned to discuss the fact that practical knowledge cannot be limited to knowledge of pressure in and movements of the body, his way of putting his point belies a concern with knowledge of what one is doing that is generated by a question about knowledge of the results of what one does.

But it is not clear that we can form from this a genuine objection to the Anscombean view I have outlined, at least not if we stay clear-headed about the metaphysics of action. This is because my view is that *S*'s practical knowledge that she is *A*-ing is necessary and sufficient for *S* to be *A*-ing intentionally, where the schematic letter '*A*' stands in for a thing done, or an act-type, something like raising one's arm, or pushing a boat. So what is important, for me, is non-observational knowledge of what one is doing. But *what one is doing* is a different thing from the *result* of one's doing it. I raise my arm, the result is that my arm rises; I push the boat, the result is that the boat is pushed.[56] Now, it may be that I can only know that my arm is rising observationally, and it may be that my arm must be rising for me to raise my arm, and *it may* even be that I must know that my arm is rising to know that I am raising it. But all that granted, it is my knowledge *that I am raising*

[56] We can also distinguish between my actions and their results, for example, my raising of my arm and my arm's rising, where the first event causes the second. But I prefer to state my view about intention in terms of things one does, and so restrict my notion of a result to the one in the text. I first introduced the distinction between actions and results as an assumption of the thesis in my Introduction.

my arm which is at issue, and those conditions, if met, do not entail that my knowledge that I am raising my arm is observational. All we have stated are some preconditions on my knowing that I am raising my arm, none of which show that the *form* of that knowledge is observational.[57]

One might press me on my casual attitude to this. One may say: 'but if I must know that my arm is rising to know that I am raising my arm, then my knowledge that I am raising it is inferential, inferred from the fact that I intend to raise my arm and that my arm is going up.'

Why should this be? What justifies that move from the claim I accept, that one piece of knowledge depends on another, to the claim that the former is inferred from the latter? There is no general problem with thinking that there are cases where one can know that *p* only if one knows that *q*, but where *p* need not be justified by an inference from *q*. For example, I cannot know that *that over there is a pig*, where I know this by seeing it, without knowing *that I am seeing a pig*. But I need not infer from the fact that I am seeing a pig to the fact that this is a pig, as if my *seeing* the pig were independent evidence for the claim that it *is* a pig.[58] They are just two pieces of knowledge which need, for various reasons, to come together, where one naturally depends on the other.

I have tried to defend the claim that *S*'s practical knowledge that *S* is *A*-ing is sufficient for *S* to be *A*-ing intentionally against the argument that it cannot be sufficient since practical knowledge is non-observational and, for a large proportion of values of '*A*' where *S A*'s intentionally, *S* can only know that she is *A*-ing observationally. I have argued that, although there is a natural worry here, in fact, when we try to articulate it into a specific objection, it can be dissipated. Therefore, I think that my claim of sufficiency stands.

Section 4 – If *S A*-s for a Reason, then *S A*-s Intentionally

So far, in this chapter, I have defended my view of what it is for someone to do something intentionally in terms of their practical knowledge of what they are doing. This section begins the second half of the chapter in which I defend both halves of an infamous biconditional:

---

[57] The consideration about knowledge of results is more of a worry for other Anscombean pictures which have a certain rendering of Anscombe's catchphrase "I *do* what *happens*" at their centre (Anscombe, 2000, p. 52). On such views, what I do are *actions* (not what Hornsby calls 'things done') which are a kind of spatio-temporal event or process, and the catchphrase asserts an identity between actions and their results (Ford, 2014, pp. 16–18). So the event of my raising of my arm is identical to my arm's rising; the event of my pushing of the boat is identical to the boat's being pushed (Haddock, 2005). This view means that concern with practical knowledge of what one is *doing* is both concern with knowledge of a spatio-temporal particular and with what *happens* since what one does and what happens are the very same particular (perhaps 'differently described'). On this view, there is no space for genuine results of actions, since the doing and the happening are always identified. Therefore, when Moran seems to be concerned with what I think of as knowledge of results, for these Anscombeans, he is really concerned with knowledge of the *action*. Where knowledge of results is, for me, by-the-by, if one accepts the identity of actions and 'results', questions about how agents can non-observationally know what happens are *really* questions about how they can non-observationally know their own *actions*. But this means that Moran's line of thought really *is* troubling to those Anscombeans, since it threatens their ability to account for knowledge of one's own action. Indeed, for me, this may be one of the most serious reasons for rejecting any view which asserts an identity between action and passion.

[58] At least, to claim that we infer from facts about our seeing things to facts about the things we see is a substantive claim in epistemology and the philosophy of mind that needs defending. For an excellent discussion of this in connection with John McDowell's (1983, 1986) view of perceptual reasons, see Phillips (2018, Chapter 2).

1. If *S A*-s for a reason, then *S A*-s intentionally
2. If *S A*-s intentionally, then *S A*-s for a reason

As I have already said, this line of argument is really more important to the thesis as a whole than the previous one. It does the job of arguing that the connections between *A*-ing intentionally and *A*-ing for a reason are so close as to justify the claim that if a mental property (such as a belief, desire, hope, or virtue) figures in the explanation of why someone *A*-s intentionally, then the explanation in which it figures is a rationalising one. I have motivated that thesis already, however defending the biconditional allows us to make a proper argument for it. In this section, I argue for (1); in the next, I argue for (2). In Section 6, I connect the biconditional to the thesis about mental properties and show how we can derive an argument for the Weak Rationalising View of Habits.

So far as I can tell, everyone in the philosophy of action agrees that (1) is true.[59] In contrast, (2) is hotly disputed. I find this asymmetry somewhat mystifying, especially since, when one scours the literature for an argument for (1) there is little to be found.[60] A *locus classicus* for (1) is Davidson's 'Actions, Reasons, and Causes' (1963/1980), but all he says about it is this:

> "Because 'I wanted to turn on the light' and 'I turned on the light' are logically independent, the first can be used to give a reason why the second is true. Such a reason gives minimal information: it implies that the action was intentional [...]." (Davidson, 1980, p. 6)

Why does it imply that? Davidson never says.

Despite this, there is an argument I find compelling. Kieran Setiya suggests the following quick argument in connection with Anscombe's work: "[A]cting for a reason [...] essentially involves such knowledge: in acting for a reason, I know an explanation of what I am doing that cites that reason, and therefore know that I am doing it" (Setiya, 2018). Say Sally runs from a bear for the reason that it is chasing her. As I argued in Chapter 2, Sally must know this fact and treat it as a reason for her to run. Therefore, the fact explains why Sally runs only because her perspective on the fact is one that allows it to play a motivating role. So Sally knows an explanation of why she is running. But if she knows an explanation of why she is running, she must know that she is running. Now, what sort of knowledge does Sally have of the fact that she is running? It is deeply unlikely to be any kind of observational knowledge. Most obviously, Sally will have *practical knowledge* that she is running. Therefore, she is running intentionally. So if someone does something for a reason, they do it intentionally. The first half of the biconditional is true.

---

[59] For example, Alvarez (2009, p. 293), Audi (1986, pp. 511; 542), Heuer (2014, p. 291), Hornsby (1980, p. 80), Setiya (2003, 2007), to name but a few.

[60] Most of those cited in the previous note simply assume it. Al Mele (1992) does some work to defend it. Kieran Setiya is one of the few philosophers to have provided an extensive defence of (1) in his (2003, 2007). I will not discuss his complicated argument here, however I think that the quick I argument I draw from him in the text is more or less similar to his complicated argument.

Now, this is far too quick and dirty for anyone sceptical of (1). However, since there is such widespread agreement on its truth, and what is really controversial about the biconditional is (2), I do not think I am putting myself at an unfair advantage with the quick argument. So, in the next section, I will spend much more time defending the contentious part of the biconditional.

Section 5 – If *S A*-s Intentionally, Then *S A*-s for a Reason

Now I turn my attention to the second thesis, (2): if *S A*-s intentionally then *S A*-s for a reason. This is very controversial, and many philosophers deny it because they think there are counterexamples. However, this should be surprising since the thesis is initially quite plausible. After all, pre-theoretically, doing something intentionally is related to doing something *on purpose*. The idea of purposiveness, of having a purpose with which one does something, is itself quite similar to the idea that one does something for a reason. At least, perhaps it only makes sense to say that someone did something on purpose if they also did it for a reason. For example, if my purpose in waving at you is to get your attention, then for this to be my purpose, I need to have a reason to get your attention (for example, that I want to get you to come over here). Otherwise, it does not seem like we can describe my hand-waving as being purposeful, or having an end.

A similar, but more theoretically loaded, motivation for the thesis comes from consideration of 'the structure of intentional action' (Boyle & Lavin, 2010). Roughly, the idea is that intentional actions have a calculative structure where ones does one thing in order to do another, in order to achieve some terminal end-state such as having baked a loaf of banana bread or having poisoned some Nazis. But, the thought goes, this calculative means-end structure only makes sense if we see the first steps in the chain as being done (at least in part) for the reason that one wants to do (or is doing) the latter steps. If the normativity of reason were not in play, one would not be mixing the batter *in order to* bake the banana bread.[61] One would just be mixing some batter, and perhaps some banana bread would get made.

One need not buy this picture of the structure of action to see the attractiveness of the view that doing something intentionally requires doing something for a reason. To my mind, it is incredibly plausible so far as theses go. However, it tends to be rejected on the basis of seemingly powerful counterexamples. Therefore, my strategy will be to show that the counterexamples fail to show that the thesis is false. Since it is independently attractive, and the counterexamples fail, I argue that we should accept that (2) is true.

The two main counterexamples are very common, but I will focus on two significant discussions: the first is due to Maria Alvarez (2009) who discusses cases of acting 'just because one feels like it'; the other is due to Rosalind Hursthouse (1991), who discusses actions which express one's emotions.

[61] For very interesting objections to the calculative view of action, see Sylvan (forthcoming).

5.1 – 'I Just Felt Like It'

I'll discuss Alvarez's argument first. Alvarez thinks that there are cases where I do something "just because I feel like doing it" (Alvarez, 2009, p. 298). For example, walking down an empty corridor, I can just do a cartwheel for the sake of it. I don't do it for exercise, or to practice my technique, or to make anyone laugh, or (Alvarez says) for any other reason. "[T]here was no end or purpose in pursuit of which I did it, nor any fact that, in my eyes, made doing it seem good or valuable. [...] I just felt like doing it" (ibid). Alvarez calls the facts in virtue of which things seem good or valuable 'desirability characterisations' of the desired thing.

My first response is to register a complaint about the example: it is an odd choice, since there *are* desirability characterisations for cartwheeling – it *is* fun, it *is* a show of skill, and it *is* exercise. And any normal person who can cartwheel will be aware of this. It is very hard to imagine someone cartwheeling in a case where they are aware of none of these facts, or where they thought that all the facts characterise cartwheeling as undesirable. Can we imagine someone intentionally going to a football match, or listening to their favourite song, or having a beer with friends, without *any* sense of what might be good about it? I think not.

My point is that there are some cases which do not at all lend themselves to the construction of the kind of set-up Alvarez envisages. Alternative cases in the literature do lend themselves to this. Consider Warren Quinn's Radio-man who apparently intentionally turns on radios when he is unaware of any desirability characterisation (Quinn, 1993). However, the problem is that examples like that challenge us to consider whether these agents act intentionally at all, since they are cases in which we struggle to interpret the agent as seeing any *point* in doing what they are doing. In which case, they are not good fodder for Alvarez's argument.

Could there be other mundane cases, like walking aimlessly through the kitchen, which could play the role Alvarez needs the example for? Perhaps. But how are we supposed to tell? For anyone wanting to defend the second part of the biconditional, the question will be 'Why should we accept that characterisation of what's going on as intentional but reason-less?' It is not enough for Alvarez to just say that it is so. What we need is an *argument* to show that the characterisation of the kitchen-wandering as intentional but not done for a reason is correct. But it is not at all clear what that argument is supposed to be.

However, even if I grant the case as it is written, I do not think it is a counterexample to (2). This is because, perhaps naively, I think that *that I felt like doing it* seems a perfect candidate for being my reason for cartwheeling. After all, that is the fact which explains why I cartwheeled, and since there was a first-personally available felt urge to cartwheel, I knew the fact, *and* I know that it is the explanation of what I did. It bears many of the hallmarks of being my reason for acting. What is *wrong* with saying that my desire, or my felt urge, to cartwheel grounds my reason for doing so?

Alvarez countenances this objection and is unconvinced. She says that the fact that I desire to cartwheel does not make it desirable because "things do not seem desirable to me because I desire them but rather I desire them because they seem desirable to me" (Alvarez, 2009, p. 299).

The thought is that, in this case (the description of which I accept for the sake of argument), there is no 'desirability characterisation' of cartwheeling – no fact in virtue of which I desire to cartwheel, such as that it would be fun – and so I desire to cartwheel for no reason. If I desire to cartwheel for no reason, then Alvarez thinks that my desire is unintelligible. But if my desire is unintelligible – if it is not itself responsive to reasons – then it cannot ground a reason for me to cartwheel. So I cannot cartwheel for the reason *that I desire to cartwheel* or *that I felt like it*, because there are no such reasons.

There are several problems with this argument. Firstly, even if a desire is not based on a desirability characterisation, or at least not one which the agent is aware of as such, there are strong reasons for thinking that an agent's having a desire often *does* give them a reason to do what they want to do. One strong reason appears in Ruth Chang's (2004) discussion of Buridan's Ass-style cases. In those cases, there are no independent reasons to pick one option over the other because they are the same in all relevant respects (think bales of hay, tins of soup, and seats in the lecture theatre). If an agent in that situation just happens to have a strong desire for one option over the others, then that desire can break their indifference, tipping them over the edge into choice. *Ex hypothesi*, the desire cannot be grounded in independent desirability characterisations because the courses of action are relevantly the same. But it is true that the fact that a person (or ass) happens to have a desire for one option makes picking that option what they rationally should do. That is, their desire gives them a reason for action.

In fact, so described, Alvarez's case is somewhat like a Buridan's Ass scenario. I'm wandering around, with no particular desire to do anything, and therefore am somewhat indifferent to a number of courses of action. Then I am struck by a desire to cartwheel – I just feel like it. This desire breaks my indifference and provides me with a reason to cartwheel: that *that is what I feel like doing*.

A second strong response is developed by Edgar Phillips (2018, pp. 73–96) from an example of Stuart Hampshire's (1999). An astute collector of bronze sculptures is looking for a piece to add to his collection, and one day he sees two pieces at an auction. The first is much better made, more beautiful, rarer, and would add more notoriety to his collection. The second is fine, but in all objective terms inferior. However, he simply falls in love with it and buys it instead of the superior bronze. "The intensity of his desire is the reason he would give", rather than anything independently desirable about the inferior bronze (Hampshire, 1999). Phillips runs through, in great detail, various responses to this, and I am convinced by the case he makes that the only way to understand this fairly normal occurrence is to treat it as a case of acting for a genuinely idiosyncratic reason: he buys it because *it's the one he wants*. Again, this is relevantly similar to Alvarez's case – even though my cartwheeling isn't rationally merited by anything independent of my desire, once I have the desire, and I am taken with cartwheeling, that fact merits my cartwheeling. Why shouldn't we think of my feeling like cartwheeling as having a yen to do something which actually grounds a desirability characterisation rather than being explained by one?

Perhaps the reason Alvarez thinks this desire cannot figure in a rationalising explanation of my cartwheeling is that nothing good comes along with satisfying my desire to cartwheel if the only good thing to be said of it is that I desire it. Cartwheeling would satisfy the desire, but so what?

It is important to remember, though, that we have slid from 'I *felt* like it' to talk of desires. But talk of 'desire' can ride roughshod over some important distinctions. In particular, the distinction between two kinds of desire-satisfaction: the semantic and the psychological. A desire is semantically satisfied if the conditions of satisfaction laid down by its content are met. This is a thin, formal, sense of 'satisfaction'. The psychological conception of satisfaction is the conception of a mental state which "constitutes the last stage of a natural motivational cycle" (Phillips, 2018, p. 90). In this sense, a desire is satisfied when the person feels content that they have gotten or done what they wanted, and is contrasted with cases of frustrated desire where the agent is literally frustrated by not getting what they want. The first is a formal notion, being merely semantic; the second is a substantive psychological notion which implicates a variety of emotional and phenomenal states. And it is plausible that the two conceptions of desire-satisfaction each bring along a different kind of desire.

It seems that cases of doing something because one feels like it must be understood as involving desire and desire satisfaction in the substantive psychological sense. These desires are feelings of mini-compulsion, as when you feel the need to make a pun on what someone said, or see a bit of fluff on your partner's face that you have a desperate urge to remove. These feelings are a little unpleasant, and in their own small way they compel us to satisfy them. And if we don't, it feels uncomfortable and unsatisfying – 'I missed the moment, what a great gag!'; 'Argh, I just *need* to remove it!'. So there is something good about doing something just because you feel like it. The good is that you mollify the feeling, avoid the unsatisfactory feeling of not doing so, and feel the small satisfaction of doing what you wanted. But these goods are wholly dependent on the fact *that you feel like it*. The same, it seems to me, goes for my cartwheeling. In being struck by the small urge, these small pleasures and pains are implicated. We can call the urge a desire if we like, but we must keep in mind that it is a kind of desire which makes plain sense of the thought that in cartwheeling I act for the reason *that I feel like it*.

Helen Steward has suggested to me that this misses the point. She suggests that 'I just felt like it' is (at least sometimes) simply a form of words we sometimes use to indicate that we did what we did for no reason, and that it need not indicate that there was a psychologically real feeling or desire – *my feeling like cartwheeling* – to which I am referring back.[62] This thought can be bolstered by recognising that Alvarez had a number of verbal options when expressing her point, not all of which lead in the same direction. For example, we can give answers to 'Why?' questions such as these: 'for no particular reason'; 'no special reason'; 'hmm, dunno'; and 'what a stupid question – who needs a reason to cartwheel willy nilly?!'. All of these answers are as equally appropriate (depending on the conversational context) as 'I just felt like it'. And they all purport to express the fact that the agent cartwheeled for no reason. Alvarez could have really used any of these verbal forms to express her point: that I can sometimes cartwheel (or do something else) for no reason at all. In which case, my latching onto the apparent invocation of *a feeling* is a mistake – I have read too much into Alvarez's words.

[62] Steward's suggestion is reminiscent of (Anscombe, 2000, p. 91).

I find it hard to see this as not begging the question. Of course, I agree that languages have forms of words which people in fact use to express the thought that they did something for no reason. It is obvious that many people believe that this is possible, even when it comes to doing something intentionally. Some of those people are even philosophers who write papers on the topic. Perhaps this is some evidence in favour of the thought that it is possible. But recognising that someone may say 'I just felt like it' to express their belief that something they did intentionally they did for no reason does not settle the debate. We need to know whether what they say is *true*. To assess that, we need to marshal arguments and theoretical considerations.

One theoretical consideration is the general connection between someone's doing something intentionally and their doing that thing's being intelligible from their own perspective. For me to cartwheel intentionally is for me to cartwheel and practically know that I am. If I have practical knowledge of my cartwheeling, it should be intelligible to me that this is something that I am doing but might not have done; that this is something which I could stop doing, or whose course I could alter. I should therefore not be totally indifferent about what it is I'm doing intentionally, and how well it is going – I should see it as good, or desirable, or as something I have a reason to do. But in Alvarez's stipulated case, I don't care one jot about any of this. But then it is hard to see my cartwheeling as intentional, since it isn't even intelligible to *me* why I'm doing it.

One of Steward's charges was that my focus on the feelings associated with certain sorts of desires missed the mark because I read too much into Alvarez's words. However, the reason I focussed in on feelings associated with desire was that I think the form of words is actually fairly transparent: though it is of course often *used* to say that one had no reason by someone who believes that is possible, it seems that its home is in reporting a psychologically real and phenomenally present desire which implicates the small pleasures and pains that I have discussed. These desires are underappreciated in philosophy, and appreciating their existence sheds light on cases like Alvarez's precisely because without invoking them the cartwheeling seems unintelligible. If we have general reasons to think that there are demands on the first-person intelligibility of doing things intentionally, then in cases such as Alvarez's, we may look around for reasons which can meet the demands. As it happens, Alvarez's use of 'I just felt like it' is an important indicator of how to meet the demand: remember that very often, when we do something on a whim, there are these small pleasures and pains, anticipations and affects, which are involved in wanting to do something. If Alvarez had said (as Anscombe did) 'for no particular reason' (Anscombe, 2000, p. 25), or 'for no special reason', I would still have argued that the most plausible candidate for being the agent's reason to cartwheel was that they wanted to. The point is that the desires I have discussed play a role in making my intentionally cartwheeling intelligible from my own perspective, and this is true regardless of the form of words Alvarez might have used.

5.2 – Expressive Actions and Arational Actions

The second kind of apparent counterexample to claim (2), that if *S A*-s intentionally then *S A*-s for a reason, is a set of cases where one expresses oneself in the grip of an emotion a by shouting in anger, or kicking their printer in frustration, or covering their face in shame, or jumping for joy. A number of philosophers take themselves to be following Hursthouse in thinking that these are cases of doing something intentionally but for no reason at all. Hursthouse herself called these cases of 'arational action'. However, since that somewhat begs the question against me, I shall say that they are 'things done expressively', or 'expressive things we do'.[63]

Although it is common to think that things done expressively are intentional, that they occur because their agents are in the grip of an emotion, but that we cannot recover any reasons for which they do these things, philosophers have tended to simply advert to Hursthouse's discussion rather than develop the argument further.[64] What one finds in the literature is a lot of agreement with Hursthouse that such cases exist, and that they show that not everything done intentionally is done for a reason.

However, when we turn to Hursthouse's (1992) paper, we surprisingly find that this is not what she argues: she has been rather sorely misread. Not only does she not argue that the things done expressively are done for no reason, she actually says this explicitly.

Hursthouse first outlines 'the standard account' – that is, Davidson's account – that arational actions are counterexamples. The standard account is one according to which, if an agent acts intentionally, then they acted "for a reason in the sense that there is a true description of action of the form 'X did it (in order) to. . .' or 'X was trying to. . .' which will 'reveal the favorable light in which the agent saw what he did,' and hence involve, or imply, the ascription of a suitable *belief*" (Hursthouse, 1991, p. 59).[65] It is worth quoting at length what she then says at the bottom of the page:

> "To get quite clear about what is at issue, let us consider as an example, Jane, who, in a wave of hatred for Joan, tears at Joan's photo with her nails, and gouges holes in the eyes. I can agree that Jane does this because, hating Joan, she wants to scratch her face, and gouge out her eyes; I can agree that she would not have torn at the photo if she had not believed that it was a photo of Joan; and if someone wants to say, 'So those are the reasons for the action,' I do not want to quarrel, for these "reasons" do *not* form the appropriate desire-belief pair assumed by the standard account. On the standard account, if the explanatory desire in this case is the desire to scratch Joan's face, then the appropriate belief has to be something absurd, such as the belief that the photo of Joan *is* Joan, or that scratching the photo will be causally efficacious in defacing its original. And my disagreement is with adherents of the standard account, who must think that some *non*absurd candidates for

[63] This new terminology also encodes a shift from talking of *actions* to talking of *things done*, which better fits how I think about the metaphysics of action. However, this change makes no difference in what follows.

[64] See Döring (2003, 2007), Russell (2009, p. 188), Scarantino and Nielsen (2015), and Setiya (2007, pp. 52–53).

[65] The embedded quotation is from (McDowell, 1982).

> appropriate beliefs to ascribe to agents performing arational actions are available." (Hursthouse, 1991, pp. 59–60)

Hursthouse makes crystal clear that her argument is intended to show that *the standard view*'s ascription of *particular kinds of beliefs* to people who act in the grip of an emotion is absurd. But she accepts – rightly and readily – that there may be other accounts of how rationalising action-explanations work on which Jane *does* scratch the photo for a reason. Her point is that *Davidson's* conception of reasons forces ascriptions of absurd beliefs onto normal emotional agents, and that if *those* beliefs are required for rational action, then expressive actions are arational.

But my view is not Davidsonian, so my view is not one on which agents must have the absurd beliefs Hursthouse shows Davidsonians must attribute to emotional agents. In fact, on one reading of the passage above, Hursthouse provides a suggestion (which she says she will not quarrel with) of what Jane's reason may be: Jane's reason for tearing at the photo of Joan is *that it is a photo of Joan*. Now, Jane would not do this were she not in a fit of rage and wanted to deface Joan's image. But in the account I have defended, that means that Jane's rage and her desire figure in the broad rationalising explanation of why she tears at the photo, since they alter the landscape of Jane's reasons by turning the fact *that this is a photograph of Joan* into a reason to tear at it. Without her anger, that fact would not be a reason for Jane to tear the photograph. But given her anger, it is a reason to tear at that photo rather than any other.[66] In the same vein, if Jane screams 'I'll have you for this, Joan!' into thin air, then her reason for screaming this and not 'I'll have you for this, Billy!' is that it is *Joan* she is angry at, and 'Joan' is Joan's name. That is Jane's reason for shouting 'Joan' rather than 'Billy'.

Nothing Hursthouse argues in her paper touches, or is intended to touch, this view. These are not absurd means-end beliefs attributed to agents out of the necessity of a theory. These are reasons for doing the things which we do when we act in the grip of an emotion which can be imputed to us with both plausibility and explanatory power. But there are no arguments in the literature which go beyond Hursthouse's conclusions. So, the things we do in the grip of emotions pose no problem to the thought that if something is done intentionally then it is done for a reason.

We are now left without any reason to deny the contentious part of the biconditional, (2), for neither of the forms of counterexample that compel people to reject it pose much of an obstacle. And since they are the only reasons given for denying (2), and I argued that it is independently very plausible, I think we should accept it. Therefore, the whole biconditional is true: *S A*-s intentionally if and only if *S A*-s for a reason.

Section 6 – Intention, Reasons for Action, and Mental Properties

[66] There is a contrastive flavour to this explanation which I will address in more detail in Chapter 7.

It is time to step back a little and see what these results mean for my argument for the Rationalising View of Habit. Remember, the point of defending the biconditional formed by (1) and (2), is that it is an important ingredient in arguing for the claim that if a mental property figures in the explanation of why someone does something intentionally, then the explanation it figures in is a broad rationalisation. And that claim is the second premise of the argument for the Rationalising View of Habit that I sketched at the end of the previous chapter. So the biconditional really is quite important for me. However, it is not yet clear how we can get from the biconditional to the thesis about mental properties. After all, my treatment of the biconditional contained no general claims about mental properties as such. So, in this section, I want to say how the biconditional supports the thesis about mental properties. In the course of elucidating this, an argument for the Weak Rationalising View of Habit will emerge.

The important part of the biconditional is the second half, (2): that if *S A*-s intentionally, then *S A*-s for a reason. We can think of this conditional as 'taking us from intention to rationalisations' in the sense that, if we know that someone acted intentionally, then we know that there are two sorts of rationalisations of their doing so, one narrow and one broad. Take Sally, for example, who we know is intentionally running. The conditional allows us to take that information and cast around for Sally's reason for running and the aspects of her perspective that makes her running for that reason intelligible. Those aspects are the properties which make up her perspective on the world and what she has reason to do. Therefore, (2)'s truth means that if Sally is running intentionally, then there is an explanation in terms of her mental properties which makes the narrow rationalisation in terms of her reason intelligible. This means that (2) provides us with a strong connection between someone's doing something intentionally and their mental properties: if someone does something intentionally, then there must be a broad rationalising explanation of what they do in which their mental properties figure.

This is vital for me because it rules out the possibility that Sally may be running intentionally but without doing so for a reason. It rules out that she is running 'just because she felt like it' (on Alvarez's interpretation of that sort of case) or out of sheer fear, such that Sally's action is 'purely expressive' or arational. Ruling out these possibilities means that wherever we find someone doing something intentionally, we find them doing something for a reason, which in turn entails a broad rationalisation of what they do in terms of their mental properties. Since habits are a sort of mental property, this begins to open up a connection between acting intentionally and mental properties, such that we can say that if a mental property figures in an explanation of why someone does something intentionally, the explanation it figures in is a broad rationalisation.

But what I have said so far is not quite sufficient. It stops short of my stronger claim that if *any* mental property figures in the explanation of why someone does something intentionally, then the explanation is a broad rationalisation. For all I have said, it might be true that everything done intentionally has a broad rationalisation in terms of *some* of the agent's mental properties, but where there are limits on *which* properties these may be.

However, this does not seem very plausible. In the previous chapter, in introducing the idea of a mental property's figuring in a broad rationalisation I gave a very large number of examples. I

showed how classic mental states like belief, desires, emotions and pain may figure in rationalisations. But I also showed how properties such as *knowing that p*, *being kind*, and *being shy* may figure in such explanations. And the list was not exhaustive. Moods, perceptions, hallucinations, hopes, suppositions, skills, and fantasies can all also be found natural slots in broad rationalisations. They all, in various ways, contribute to making someone's acting for a motivating reason intelligible. Given the vast range of types of properties which can have some role in broad rationalisations, why would it be that some are left out, having no place in the story? Indeed, which properties not on that list should we leave out?

Of course, the Non-Rationalising Theorist of Habit will say we should leave *habits* off the list. As I pointed out at the end of Chapter 2, the Non-Rationalising Theorist need not deny habit's status as a type of mental property. But they will, here, express their discomfort at accepting that habits are mental properties apt to figure in broad rationalisations.

But the discomfort seems ill-placed when we realise that, according to the Non-Rationalising Theorist, habits seem to be the odd ones out. Why should that be? It is true that habits probably do not deserve the esteem we give to knowledge or virtue, and have nothing of the *obvious* rationalising force of belief and desire. But when every mental property but habit is conceded as being apt to play a role in broad rationalisations, we should reconsider our sceptical judgements about habit. What makes habit so special as to be left out?

There is another reason for doubting the Non-Rationalising Theorist's scepticism. Say that, on some occasion, I go to *Gilmour's Café* intentionally and out of habit. There are two important facts about this situation. Firstly, since I go to *Gilmour's* intentionally, (as per the biconditional) I go there for a reason. Therefore, there is a broad rationalisation of why I go there. Secondly, my habit manifests in my going to *Gilmour's*, so it figures in an explanation of why I do that. The Non-Rationalising Theorist must say that the explanation my habit figures in is different from the broad rationalisation. They must claim there is a narrow rationalisation, a broad rationalisation, and another, non-rationalising, explanation in which my habit figures. But is this at all plausible?

I do not think so. There are perhaps all sorts of non-rationalising explanations of why I go to *Gilmour's*, for example, in terms of fundamental physics, chemistry, and neurophysiology.[67] But

[67] In fact, I am sceptical that there are any such explanations. Explanations in physics do not answer questions like 'Why did I get coffee from *Gilmour's* yesterday?' That is not their business. Neither is it the business of biology, neurophysiology, or chemistry. It is partly the business of sociology to explain why I go to *Gilmour's*, where the explanation is given in terms of economic, geographical, and social terms. But one gets the sense that this explanation piggy-backs on existing explanations in terms of my own reasons for acting, and therefore on the broad rationalisation also. The sociological explanation puts the existing explanation into a broader perspective, showing how the rationalisations are not freak explanations, but exist because of wider facts about my place in a society. Certainly, the sociological explanation cannot *usurp* the rationalisations. If it did, sociology would be more like physics, explaining the motions of things we once thought moved because they wanted to, but which in fact only act because of inextricable forces of history. Some, such as Louis Althusser in his books *For Marx* (1969) and *Reading Capital* (2016), have defended such a view under the auspices of defending Karl Marx's historical materialism. See E. P. Thompson's classic *The Poverty of Theory* (1978) for a battery of arguments that show the Aluthusserian position to be neither Marx's own view nor, indeed, true. Maurice Merleau-Ponty puts the point I want to make (and which Thompson makes) well when he argues that people, when they appear as the subjects of historical investigation, are "not merely the economic subject or man as a factor of production, but more generally the living subject – man insofar as he is a certain productivity, insofar as he wants to give his life form, insofar as he loves, hates, and creates or does not create works of art, insofar as he has

do habits figure in *those* explanations? I think not. Those explanations cite facts about me which are patently at the wrong level to include my habits. Physics deals with atoms, quarks and fields; chemistry with molecules and bonds; neurophysiology with brains and nerves. None of these has space for habits, in part because they are concerned to explain my behaviour in terms of the behaviour of *my parts*, or the things *inside me*. But my habit of getting coffee from *Gilmour's* is not a property of one of my parts, or of anything inside me: it is a property of *me*. It only appears in an explanation which is at a level where *I* am included as a whole person with a subjective perspective on the world. Now, sociological explanations do include whole persons. However, mental properties of people only figure in sociological explanations because sociology is a science of human behaviour. Its subject matter is pre-packaged with psychological content: it is interested in giving broad, historical and cultural explanations of how people come to do, believe, love, and fight for things.[68] So, if habits appear in sociological explanations, it is only because they appear in lower-level explanations first.

So where the Non-Rationalising Theorist must say that my habit figures in an explanation of why I go to *Gilmour's* on this occasion *but does not* figure in the broad rationalisation, it turns out that there are no plausible alternative kinds of explanation my habit could figure in. So it is not right that, if I go to *Gilmour's* intentionally and out of habit, then my habit figures an explanation distinct from the broad rationalisation. It must figure in the rationalising explanation; there are no other kinds of explanation which we can slot habits into.

What we have struck upon is an argument for the Weak Rationalising View of Habit: it turns out that habits *can and do* figure in broad rationalising explanations of what someone does when they do something habitually, *at least* in cases where what they do is also intentional. This is only the Weak View, since it leaves open the possibility that habits sometimes manifest in someone's doing something non-intentionally or unintentionally.

Nonetheless, the Weak View is a substantial gain for anyone wanting to vindicate habit's aptness for figuring in rationalisations. It should be enough to silence the most vehement sceptical ideas, even if it does not go as far as the strong Rationalising View. In fact, though, we need not rest content with the Weak View. I think we can go one better. I think that we are now well-positioned to defend the full-fat Rationalising View that there are no cases where a habit manifests and does not have figure in a broad rationalisation and have a bearing on a narrow one. This is because I think that if one does something habitually, then one does it intentionally, which rules out any cases where

children or does not. Historical materialism is not an exclusively economic causality" (Merleau-Ponty, 2012, p. 175). The idea is that, even for a theory of historical explanation whose bread and butter are things like modes of production and productive forces, actual people, their relationships, and their psychologies must not go missing. If they do, we forfeit our right to call our explanations 'sociological' or 'historical'. (Indeed, Althusser and his followers recognised and embraced this result. But this move should make us ask what they think they even have a theory *of*). For more discussion of the way that historical explanations and action-explanations are nested together in the context of a Marxist conception of history, see Meiksins Wood (2016; 2016), Merleau-Ponty (2012, pp. 174–178; 458–523), and Miliband (2015a, 2015b).

[68] Of course, sociology is also interested in economic crises, the historical tendency for the rate of profit to fall, the uneven geographical development of technology, and other seemingly impersonal things. However, these are not *really* that impersonal. They are all intimately tied to people doing things together, to their working, buying, selling, and striking.

habits manifest and fail to figure in explanations of what someone does intentionally. And that is the view I will argue for in the next chapter, and defend from objections in Chapter 5.

CHAPTER 4

TALES OF ALICE AND BERT: A CASE FOR THE RATIONALISING VIEW

I have argued that habits, when they manifest in someone's doing something intentionally, figure in broad rationalisations. This is because they are mental properties, and if mental properties figure in the explanation of why someone does something intentionally, then they figure in a rationalising explanation of why they do it. The argument for this claim was, roughly, that if someone *A*-s intentionally they *A* for a reason, and so there is a broad rationalisation of why they *A*; if a mental property such as a habit figures in an explanation of why they *A* intentionally, then, since there aren't any *other* explanations they could figure in, they must figure in the broad rationalisation.

These considerations form the argument for what I called the Weak Rationalising View of Habit: the view that, habits figure in rationalisations when they manifest in someone's doing something intentionally, but that there may be cases where they do not figure in such rationalisations because what is done habitually is not done intentionally. The Weak View simply leaves open whether cases of doing something habitually and intentionally are the only cases. Indeed most philosophers will think it is obvious that we should leave that open. After all, don't I sometimes habitually bite my nails without knowing it, and don't some people habitually interrupt others with such little self-consciousness they cannot be said to interrupt intentionally?

If this were true, it would block my line of argument from the connection between habit and intention, and the connection between intention and rationalisation, to the Rationalising View of Habit. I cannot progress beyond the Weak View without arguing that everything done habitually is also done intentionally. And that is surely mad.

However, I happen to think that there are very strong reasons to think that the thesis is true. Everything done habitually *is* done intentionally, despite the apparently obvious counterexamples. It is the purpose of this chapter to argue for that claim which, in the argument sketch below, entails Premise (3):

1. Habits are mental properties (broadly construed);
2. If mental properties figure in explanations of why someone does something *intentionally*, then they figure in broad rationalisations of why they do that thing;

3. Whenever habits figure in explanations by manifesting, they figure in explanations of why the habit bearer does something intentionally.

C) Therefore, whenever habits manifests, they figure in broad rationalisations

In the next chapter I defend the claim from three powerful arguments. If it is true that everything done habitually is done intentionally, then we have completed the case for the Rationalising View of Habit: since everything done habitually is done intentionally, whenever habits manifest they figure in explanations of why the habit-bearer does something intentionally. And the last chapter showed that if any mental property figures in an explanation of why someone *A*-s intentionally, the explanation it figures in is a broad rationalisation. Therefore, habits, whenever they manifest, always figure in broad rationalisations.

My strategy for defending the view that if *S A*-s habitually then *S A*-s intentionally is fairly complicated, so I will spend all of Section 1 describing it. In doing so, I will tell a story which makes my abstract strategy concrete, and which I will be working with throughout the chapter. I have two different arguments, both of which employ the abstract strategy and work with the story. In Section 2, I provide the first, simpler, argument. In Section 3, I provide the second, more complex, one.

Section 1 – A Strategy, a Story, and a Question

My goal in this chapter is to argue that if someone does something habitually, then they do it intentionally. So, I will call the thesis '*Goal*':

For any habit-bearer, S, and act-type, A:
Goal: If *S A*-s out of habit, then *S A*-s intentionally.

My strategy for defending *Goal* involves exploring the criteria for *individuating* habits. The strategy starts from the thought that habits, like other kinds of powers (such as dispositions, tendencies and abilities), are individuated at least partly according to their manifestation-types. Call this the *Manifestation Thesis*:

For any habit, H:
Manifestation Thesis: *H* is partly individuated by some manifestation-type, *A*.

A manifestation-type is an act-type, or a thing done.[69] The idea behind the *Manifestation Thesis* is that in order to individuate habits, we need to know what they are habits of *doing*. For example, my habit of looking at my watch when I am late is distinguished from my habit of walking

[69] I reserve the term 'manifestation' for the action which is the doing of the thing done. I will have little use for talking about manifestations.

quickly when I am late because, although the types of context in which they are manifested are the same, they manifest in my doing different things.

The other factor which partly individuates habits is the context-type in which they are manifested, given by the schematic letter '*C*'. If I have the habit of running in the morning and someone else has the habit of running in the evening, then we have different habits even though what we are in the habit of doing is the same. The difference is found in the types of context in which we run. Therefore, roughly, if we know that *x* habitually *A*-s in *C*, and *y* habitually *A*-s in *C*, then we know that *x* and *y* have the same habit. I think that context-types are crucial for individuating habits, however my interest here is in how consideration of the *Manifestation Thesis* may support *Goal*. Therefore, I will often omit reference to context-types, saying things like 'the habit of *A*-ing', for ease of expression.

I think that there is an argumentative strategy from the *Manifestation Thesis* to my *Goal*. The strategy is to show that, for any habit, it is impossible to individuate it according to a manifestation-type unless we appeal to things the agent does intentionally when their habit fully manifests. What the agent does intentionally in fully manifesting their habit provides us with an answer to how the habit is individuated. If we can only pick out *H* as being a habit of *A*-ing by appealing to the fact that when *H* manifests the agent *A*-s intentionally, then we can only identify the habit by identifying its manifestations as the habit-bearer's doing something intentionally. I will argue that we can only individuate habits in this way. Therefore, if *S A*-s out of habit, then *S A*-s intentionally. *Goal* is true.

Since *Goal* is true, Premise (3) of the argument above follows quickly. This is because, whenever habits manifest, they figure in explanations of why the habit-bearer does what they do. One of those things must be what they do habitually, when their habit manifests. And, by *Goal*, what they do habitually they also do intentionally. So the habit figures in an explanation of why they do something intentionally. Being a mental property, it therefore figures in a broad rationalisation of why they do that thing. Therefore, there is a very simple route from *Goal* to Premise (3).

However, this is all rather abstract, dense, and hard to understand. Therefore, my argument will proceed by way of example. First, I'll start with a story about someone's habit. Then, I will set-up the question about how to individuate that person's habit. Then, I will describe how the abstract strategy I just described works in the concrete case. Then, I will pursue the strategy with two distinct arguments: firstly, I will provide a general argument for *Goal*; secondly, I will provide cases which are variations on the story, and show they provide counterfactual support for *Goal*. Let's begin with the story.

1.1 – The Story

Take Alice. Alice is a philosopher whose desk is normal except for a curious red button which is built into the table-top. Sometimes, when she is thinking very deeply about some problem, Alice has taken to pressing the red button. What she doesn't know is that when she presses the red button

an alarm goes off in another part of the building, irritating and interrupting Bert, who is hard at work dictating his latest novel to his secretary. So whenever Alice presses the red button, she annoys Bert. On the view of action individuation I assumed in my Introduction, according to which when one *A*-s by *B*-ing there is an event, ϕ, which is one's *A*-ing and one's *B*-ing, since Alice annoys Bert by pressing the red button, she does two things yet there is only one action.[70] We can stipulate that Alice pushes the button intentionally, but does not annoy Bert intentionally. This is because she has practical knowledge of her button-pressing but lacks all knowledge (practical and otherwise) of her Bert-annoying. And she interrupts Bert whenever she presses her button. Finally, say that Alice has intentionally pressed the red button as many times as she has unintentionally interrupted Bert.

1.2 – One Question; Three Answers

That's the set-up. Now I'll ask how to characterise and individuate Alice's habit. As I said, Alice has taken to pressing the red button when deep in thought, and after a while, she starts to form a habit. To try to describe this habit neutrally, Alice's habit is such that it manifests when she is thinking deeply about something, and whenever it manifests she both presses the red button and annoys Bert. We can put this by saying that her habit figures in an explanation of why she pushes the button and why she annoys Bert. But how should Alice's habit be individuated? What is it a habit of *doing*? Here are the options:

Alice has:

1. The habit of pressing the red button when thinking;
2. The habit of interrupting Bert when thinking;
3. The habit of pressing the red button *and* interrupting Bert when thinking;

The key thing to focus on for my purposes is not the differences between the things done that each characterisation mentions – there's nothing intrinsically interesting about interrupting Bert or pressing the red button. I could have written a story where Alice intentionally annoys Bert and in doing so always unintentionally presses a red button, or the story could have involved Alice doing any number of different things, were she does one thing intentionally and another thing unintentionally whenever her habit manifests. The important thing is that the story contains a clear separation between the characterisations of Alice's habit in terms of what she does intentionally and what she does unintentionally when her habit manifests. (1) represents something Alice does intentionally, (2) represents something she does unintentionally, and (3) represents a conjunction of the intentional and the unintentional things Alice does when her habit manifests.

[70] This view only affects my formulation of the story, not the conclusion of my arguments.

1.3 – The Strategy

I will argue for *Goal* by applying the abstract strategy described above in the following way. I will show that Alice's habit must be characterised as being (1), a habit of pressing the red button. If that is right, and the important difference between (1) and (2)-(3) is that (1) characterises the habit only in terms of what Alice does intentionally when her habit fully manifests, then what Alice's habit manifests in her doing is something she does intentionally: her habit manifests in her pressing the red button, and she presses the button intentionally. Therefore, we can only identify what she is in the habit of doing according to what she does intentionally when her habit manifests.

This means that, whenever Alice's habit manifests, it manifests in her doing something intentionally. Of course, this story is about Alice, but the abstract strategy is supposed to apply to anyone and any habit. What the story is designed to do is carefully isolate the things one does intentionally when one's habit manifests and the things one does not do intentionally. This division exists in all cases, however, in less artificial examples, it can be very hard to say exactly what someone does intentionally. Therefore, although the immediate conclusion of my arguments will be about a specific person and her specific habit, we can abstract from the particulars of the case and learn a lesson about habits in general: it is impossible to do something habitually and not do it intentionally. This is how my strategy gets us to *Goal*.

Importantly, the artificiality is only methodological. It allows us to see more clearly the way that habit and intention interact, which we can then apply across the board in non-artificial cases. Indeed, part of the point in Chapter 5 is to show that we can apply the results of considering Alice and Bert to normal cases where the relation between intention and habit is much murkier.

I have said that if we must accept (1), then we must accept that whatever Alice (or anyone) does habitually she (or one) does intentionally. And (1) characterises Alice's habit as a habit of pressing the red button. We must be careful to keep this apart from the claim that if Alice is in the habit of pressing the red button, then whenever her habit manifests she must intentionally press the red button. This is because it is false. I will say much more about why in Chapter 5, but for the moment note that it seems an open possibility that Alice might have the habit of pressing the red button, have that habit manifest, but simply miss the button with her finger. In such cases of 'habit misfires' or 'partial manifestations', Alice does *something* intentionally (she moves her hand toward the button), does it habitually, but she does not press the red button. That is, she fails to do what her habit is canonically identified as a habit of doing, but still does something habitually and intentionally. This is eminently plausible, and we should keep it as an open possibility.

Therefore, I want to be clear that in arguing for *Goal* – if *S A*-s habitually then *S A*-s intentionally – I am *not* committed to the view that everything done habitually is a habit-bearer's intentionally doing what their habit is *canonically identified* as a habit of doing. In Chapter 5, I will consider and reject an argument for this in more detail. For now, this is a warning not to conflate the two views. Until I am ready to discuss partial manifestations and habit misfires in more detail, I will sometimes suppress this complexity and simply talk about 'what a habit-bearer does intentionally when their habit manifests'. However, sometimes, I will say 'when their habit *fully* or

properly manifests' in order to remind the reader that my aim is to provide a link between the *Manifestation Thesis* and *Goal*. That is, the strategy is to show that we must give the *canonical* specification of a habit in terms of a manifestation-type which the habit-bearer does intentionally if their habit fully manifests in their doing the canonical type of thing.

A final word about the strategy. Although the story is designed so that the difference between options (1) and (2) is only captured by a difference in what Alice does intentionally, there may be other differences which turn out to be salient. And these other differences may turn out to play a role in ensuring that (1) is the correct characterisation, and thereby undermine my claim that it is *intention* which performs this job. The most promising alternative idea is that characterisations in terms of button-pressing and Bert-annoying might sustain different counterfactuals about what Alice would do in different situations, where the counterfactuals individuate her habit. It is possible they might thereby make the fact that Alice *in fact* presses the button intentionally redundant in the individuation of her habit according to (1).

My strategy, as I have described it, cannot rule this out from the off. However, the argument in Section 3 proceeds by way of counterfactual variation on the story of Alice and Bert. One of the lessons of that section is that the assessment of counterfactuals does indeed play a role in our judgements about the individuation of habits, but that our counterfactual assessments are themselves *guided* by our understanding of what Alice does intentionally when her habit manifests. It is because of what is done intentionally that we make the individuation-relevant counterfactual assessments that we do. I cannot argue for that here, since that is part of Section 3's role. So, in response to this particular worry about the strategy, I ask the reader to be patient and assess my execution of this point later in the chapter.

Section 2 – First Application of the Strategy: Do We Know Which Habits We Have?

My first approach to this strategy is to ask why anybody would believe option (3), that Alice has the habit of pressing the red button *and* interrupting Bert. Probably, (3) would look attractive to someone just because Alice's pressing the button always co-occurs with her annoying of Bert, and both always co-occur with her habit's manifesting. Rather than opt to characterise Alice's habit with any particular manifestation-type, some might lump for the principle that anything a habit-bearer does which always co-occurs with the manifestation of their habit is done habitually. So the right characterisation of habits in general will be conjunctive, regardless of whether the conjuncts only mention things habit-bearers do intentionally.

But this has extreme commitments. It entails that no-one has a habit of picking their nose or looking at their watch when they are late for work. Anyone who thinks they have a habit of picking their nose *really* has a habit of picking their nose, and contracting certain muscles, and disturbing air particles, and changing the local curvature of space-time, and making some propositions true (*x*'s arm is moving) and some false (*x*'s arm will not move from *the-time-just-before-x-moves-their-arm* onwards), and... so on. On the view represented by (3), or at least according to the only reason

to believe (3), it is not that the nose-picker has the habit of nose-picking which comes along with their habitually doing all this other stuff too. According to (3), they have a habit of which *picking their nose* is only one amongst a massive number of conjuncts, all of which make an equal contribution to characterising and individuating their habit. They just as much habitually change the local curve of space-time as they do habitually pick their nose.

But this is quite preposterous. It means that nobody knows what anybody's habits are, because nobody knows the conjunction of things a habit-bearer always does in manifesting their habit. But if no-one knows what anybody is in the habit of doing – if we only have minimal and deeply partial insight into anyone's habits – then I don't see how we can get on with habit-attribution and explanation at all. All of our habit-attributions and explanations would be false, unrevealing, and distort and hide the true characterisation of all habits. But this deep scepticism about such an ordinary psychological notion as habit is clearly a deep-scepticism-too-far. So we should reject the conjunctive characterisation of habits embodied in (3).

This means we should also reject option (2): that Alice has the habit of interrupting Bert. Whatever the reason one might have for thinking (2) *prima facie* plausible (I confess, I see none at all) it obviously suffers just the same problem. Should our habits really be characterised *only by reference to those things we do of which we are unaware*? That rules out the possibility that someone might do anything habitually and intentionally. But there are clearly such cases: I habitually play certain phrases on the guitar, but do so intentionally; one may habitually sit in one's favourite chair, and do so intentionally. Therefore, (2) is far from plausible.[71]

This leaves us with only option (1): that Alice has the habit of pressing the red button. If we are left with only (1), then we must identify what Alice is in the habit of doing with what she does intentionally when her habit (fully) manifests. That is, we should not include any mention of things she does unintentionally in the characterisation of what Alice is in the habit of doing. Therefore, when her habit manifests, she does all sorts of things – she (intentionally) presses the button, and she (unintentionally) annoys Bert, changes the space-time curvature, and contracts her arm's muscles. Only the thing she does intentionally is what she does habitually. If the abstract strategy I began with is correct, this entails *Goal*: if *S A*-s habitually, then *S A*-s intentionally.

Importantly, this fits with our habit-ascriptions and explanations in a way which (2) and (3) very much do not. A person can have a habit of picking their nose, and every action which is their intentionally picking their nose is also their unintentionally contracting certain hand muscles, but they do not thereby have the habit of contracting their hand muscles. I may have a habit of calling my mum every Saturday morning, but I do not have a habit of sending out micro-waves from my phone every Saturday. I may have the habit of pacing when nervous, but I do not have a habit of wearing out the carpet when nervous. These are very natural ways to identify our habits, and my

[71] Of course, a reader sceptical of *Goal* may say the same about option (1), that Alice has the habit of pressing the red button, if they think it obvious that one can act habitually and unintentionally or non-intentionally. The next chapter answers a variety of such objections, so I will not address the point here, preferring to plough on with my positive argument.

argument suggests that we can only avail ourselves of them if the canonical ways of identifying habits are guided by what a habit-bearer does intentionally when their habit fully manifests.

Section 3 – Second Application of the Strategy: An Argument From Variation in the Story

Now I will try to apply my strategy in a second way. This is somewhat more complicated, but hopefully it will be worth it.[72] The second approach to the strategy works by working through a variety of cases based on our story in order to decide how to characterise Alice's habit. How should we do this? I think a natural place to start is to see what would happen if Alice's situation were to change a little:

> (*Cut Wire*) One day Bert learns of the connection between Alice's red button and the irritating alarm, and during the night he severs the wire so that the alarm won't sound if the button is pressed. Alice wakes up, goes to work, sits at her desk, and begins writing.

When the philosophy gets very hard and she starts to think deeply, what will Alice do? She will most likely press the red button. After all, nothing has changed so far as Alice knows – her desk seems normal, she never knew what the button did, and doesn't know that it is no longer connected to an alarm. Indeed, nothing about Bert's cutting the wire even causally affects Alice's local situation. Her habit will manifest in her pressing the red button despite the fact that her doing so is not an interruption of Bert. This suggests that what Alice has a habit of doing is simply pressing the red button and not of interrupting Bert because her habit seems to manifest normally, albeit without her interrupting Bert. This is evidence against (2) and (3).

Now imagine a different case:

> (*Removed Button*) Bert learns of Alice's button, but instead of severing the wires to the alarm, he finds Alice's office and removes her red button. Alice goes to work in the morning and finds her curious button mysteriously missing.

What will happen? When she settles down to work, and reaches a difficult problem, she may reach for the button and press thin air, forgetting it had gone. Or she may remember and, annoyed, tap the table instead. Her habit would, as it were, 'attempt to manifest' and yet 'fail'. That is in the nature of habits. There is a force of habit which directs habit-bearers to do the things they are in the habit of doing. Habitual smokers who have quit find themselves rolling pieces of paper or lifting pencils to their mouths; someone who no longer wears glasses will still habitually try to push them up the bridge of their nose. It is very plausible that the way of individuating Alice's habit must be

[72] I have never discussed this material with Brendan Kelters, however I now think I have been channelling him ever since I began thinking about this line of argument. Therefore, I feel I owe him thanks.

sensitive to the fact that when she is thinking deeply she reaches for the removed red button, or that her hands are restless, or that she substitutes button-pressing for some other behaviour. Grasping the counterfactual depends, therefore, on seeing what Alice *wants*, or is *trying*, to do – she wants to press the button, and absentmindedly tries to do so. Since one cannot unintentionally try to *A*, Alice intentionally tries to press the button. So her habit seems to require a specification in terms of her pressing the red button. This evidence tells against (2). But, more, the evidence also shows that habit-ascriptions are sensitive to what Alice would *want and try* to do in counterfactual scenarios, therefore providing evidence *for* (1).

Notice that in (*Cut Wire*) Alice does not display any behavioural analogues to the restlessness in (*Removed Button*), and she does not attempt to annoy Bert in the way that she attempts to press the button when it has been removed. If Alice displays certain sorts of behaviours in (*Removed Button*), and those behaviours were reasons to opt for a specification of Alice's habit in terms of her pressing the red button, then the lack of analogous behaviours in (*Cut Wire*) is reason to opt for a specification which does not include her interrupting Bert. This would rule out (2) and (3), leaving just (1).

We also saw from (*Removed Button*) that our judgements about habit-ascriptions in counterfactual scenarios are sensitive to what agents are trying to do, and are therefore sensitive to (at least some of) what they are doing intentionally. Whilst Alice still presses the red button in (*Cut Wire*), this is not an attempt to interrupt Bert since she does not know what the button does. But if our habit-ascriptions are somewhat sensitive to what agents want or are trying to do, we should be wary of any characterisation in terms of Alice's annoying Bert. This is evidence against (2) and (3).

Here is a further case which brings out an interesting related feature:

> (*Complaining Bert*) Bert learns that Alice is accidentally annoying him and politely asks her to stop pressing the button. Alice is embarrassed, apologises, and promises to try and break her habit. Alice now knows that when she presses the button, she interrupts Bert, and she has set herself a policy of not doing those things. However, habits are somewhat resistant to policies, and she still finds herself pressing it absentmindedly.

The force of habit still has sway over Alice, regardless of her policy. But what is the force of habit directing her to *do*? In (*Cut Wire*) and (*Removed Button*) it is very doubtful that the force of habit directs her to annoy Bert, at least partly because the force of habit is a psychological phenomenon, and she has no psychological connection to Bert in those cases. But even here, where she does know of him and what her button does, it does not seem plausible that the force of her habit directs her to annoy Bert. This is because whatever the force of habit directs her to do now is most likely whatever it directed her to do before. Why would that change, just because Alice now knows that whenever she presses the button she annoys Bert? If I have a habit of biting my nails and then learn which muscles contract when I do so, it would be strange to think that the force of my habit now directs me to contract those muscles. Therefore, in (*Complaining Bert*), the force of

habit is directing Alice to press the button, not to annoy Bert. Plausibly, the force of habit is connected to habit-individuation such that, if the force of habit directs *S* to *A*, then *S* has a habit of *A*-ing. Therefore, this is reason to think that (1) better captures Alice's habit than (2) or (3), since (1) captures what the force of habit directs Alice to do.

Here is another sort of case:

> (*Interfering Cathy*) A friend of Alice's, Cathy, knows what the red button does and thinks that it would be funny to mess around with it. During the night, she adds a green button to Alice's desk right next to the red one and rewires everything so that pressing the green button now sounds the alarm in Bert's office and pressing the red button does nothing. Alice wanders into work and sees this mystifying new set-up. She gets to work, and when there is a tricky issue, she reaches for a button.

As the case stands, it is simply not obvious which button Alice will press; the counterfactual facts (counterfacts?) we have underdetermine what she will do and we cannot second guess our heroine. Alice may very well press the red button, ignoring the green one, and if she does, that would be evidence for (1) and against (2) and (3) for the same reason that (*Cut Wire*) was: Alice's behaviour is totally unchanged by the fact that her action is no longer one of annoying Bert, because the rewiring has no effect on Alice herself.

However, Alice may very well intentionally press the green button, eschewing the red.[73] But would that be evidence that her habit is one of interrupting Bert? No. At best, it would show that the best specification of her habit does not include reference to a particular colour of button, or a particular location, and so on. However, it is important to recognise that, despite the fact that if Alice has a habit of interrupting Bert then for her habit to (fully) manifest she should press the green button, her pressing of the green button does not count as evidence of that habit.

It is worth asking *why* Alice's pressing the green button would not be evidence for (2), say. I think we are unwilling to take it as evidence for that view because it would leave the relation between Alice's behaviour and the world totally mysterious. If (2) were true, then we would accept that habit-explanations require no knowledge at all of what one is doing when one acts habitually. Perhaps some will accept that. But it also requires accepting that when unknown facts change unbeknownst to the agent, the agent's behaviour is sensitive to that change such that their habit will manifest in a way which accommodates it. How could Alice (or her habit?) be sensitive to Cathy's re-wiring the buttons in (*Interfering Cathy*) if Alice herself has no knowledge of, or psychological grip on, those particular changes wrought by Cathy? If there is no explanation available, then (2) commits us to the view that habits can figure in explanations of why people act without there being *any* psychological mediation between the world and the habitual behaviour. Even worse, there is not

[73] It matters that she *intentionally* presses the green one. If she is aiming at the red button, misses it, and hits the green one accidentally, then this is most plausibly just a case of 'habit-misfire' as I called it earlier. If so, then (1) still seems the most plausible option. Throughout my discussion of any case where Alice presses the green button, I am only thinking of cases where she does so intentionally. This preserves a similarity between her pressing the red button and the green one: they are both intentionally done in the cases I discuss.

even any *causal* relation between Cathy's rewiring the buttons and Alice, such that her habit is causally (yet non-psychologically) sensitive to the meddling. Habits would manifest in ways that are sensitive to facts about the world which the habit-bearer is not even causally sensitive to. But that must be magic. The existence of this connection is deeply implausible, and any explanation involving it will be extremely strange. Therefore, so long as there is a more plausible explanation of Alice's pressing the green button which doesn't see it as explained in ways that totally bypass her psychology and her causal standing, we can avoid accepting the mysterious relation which someone who saw Alice's behaviour as evidence for (2) must posit.

Luckily, there is a very plausible alternative explanation of why Alice presses the green button (if she does so because her habit manifests):

Alice has:

4. The habit of pressing a button when thinking.

This preserves the psychological link between changes in the environment and Alice's behaviour that seems necessary for explaining why Alice presses the green button, whereas (2) severs it completely. It also preserves the causal connection between the new green button and Alice, in that Alice can see and interact with it. It is true that if she intentionally presses the green button, and presses it out of habit, then strictly (1) will also be false since no reference to colour or exact location will be required for the correct specification of Alice's habit. However, despite the differences between (1) and (4), they both contrast with (2) in the same way since Alice presses the buttons intentionally but does not annoy Bert intentionally. And it is that contrast which matters for my argument. We can therefore ignore the differences between (1) and (4), and continue to focus on contrasting (1) with (2)-(3).

Sceptical readers may worry that my strategy is not working. One may think that I have made a good case for thinking that Alice has a habit of pressing the red button, not that she must do it *intentionally*. Perhaps it looks as though my arguments are consistent with thinking that, although (1) is the best characterisation, Alice may still sometimes press the button out of habit but unintentionally. After all, Alice *might* press the button unintentionally, and can't we think of a way that she does this unintentionally and out of habit? I think the answer is 'no', and below are some cases to help us see why. This should undermine any concern that my arguments do not licence my claim that the individuation of habits depends on the link between habit and intention.

> (*Box and Lever*) Imagine that Cathy messes with Alice's buttons again, this time by covering them with a sealed box with a lever on top. If the lever is pulled, then an internal mechanism depresses the red button, such that (according to the view of action individuation I have adopted) any action which is a pulling of the lever is also a pressing of the button. Alice does not know that the box hides her red button, or that the lever is connected to a mechanism which depresses the button when pulled. So, if Alice pulls the lever, she will press the button, but unintentionally.

When Alice sees what has been done to her desk she wonders if she has been volunteered for some sort of thought experiment. But she dutifully gets to work and at some point starts to think pretty deeply again. What will she do?

As in (*Interfering Cathy*), the answer is simply not given by the description of the case. Maybe Alice will totally ignore the box's lever and display behaviours similar to those in (*Removed Button*) where Bert took Alice's button away: she may try to reach for the button, or fidget restlessly. She may want to press the button, or try and fail, since it's not available (she doesn't know where it's gone). These sorts of behaviours told in favour of (1) in (*Removed Button*), but for a different reason. In (*Removed Button*), Alice could not press the red button intentionally or unintentionally because it had been removed. In (*Box and Lever*), Alice can press the button by pulling the lever, but cannot do it intentionally because she doesn't know where it is. The fact that she is restless in (*Box and Lever*) therefore shows that the behaviours are not elicited by the mere *lack of presence* of the button, but also by the lack of *knowledge* of the button's presence. But if lack of knowledge is sufficient to produce fidgeting and absentminded-tryings, then this tells us not just that (1) is most apt, but also that the fact that (1) is apt is not brutely external to what Alice knows, wants, and intends. (1)'s aptness has something to do with the integration of button-pressing into Alice's psychology. To see what this means, let's alter the case:

> (*Box and Lever + Knowledge*) Things are just like in (*Box and Lever*), however Cathy has told Alice what the set-up is so Alice is in the know.

In this case, if Alice is sufficiently irritated by not being able to press her red button without having to pull a lever, *she will just remove the box*. In (*Box and Lever*), she manifests her frustrated behaviours because she cannot intentionally do what she is in the habit of doing; in (*Box and Lever + Knowledge*), her habit may motivate her to remove an obstacle to it in order that she can press her familiar button.

In (*Box and Lever*), if Alice manifests her irate, fidgety behaviours, we can infer that the force of her habit directs her to press the red button. However, since she doesn't know that she can press the button, the force of habit is directing her to do something which, if she did it from force of habit, she would do intentionally. Therefore, if she did know how to press her button, she may take steps to do so, as we can see from (*Box and Lever + Knowledge*). So the force of habit is directing her to do something, but she can only do that thing from the force of habit when she has a bit of knowledge that allows her to do it intentionally. So Alice can only manifest her habit from the force of habit when she presses the red button intentionally, which she can do in (*Box and Lever + Knowledge*) but not in (*Box and Lever*). This not only speaks in favour of option (1), but also indicates that the reason (1) is most apt is that Alice must press the button *intentionally* if she presses it out of habit. That is, (1)'s aptness does derive from its connection to intention as I have been suggesting.

However, as I said, in (*Box and Lever*), Alice may not ignore the lever – she may pull it. This is most pertinent to the worry I canvassed above: (1) may be correct, but that doesn't show that Alice

cannot press the button out of habit and unintentionally. If that were possible, then my strategy of arguing for *Goal* by arguing that option (1) is the best characterisation of Alice's habit would fail.

There are actually two ways for (*Box and Lever*) to falsify *Goal*. Firstly, if it turns out that Alice pulls the lever, pulls it out of habit, and presses the button out of habit. Secondly, if it turns out that Alice pulls the lever, does not pull the lever out of habit, and presses the button out of habit.

What to do about this? First things first. It is possible that Alice pulls the lever out of habit only if her habit is not properly specified by any act-type which mentions 'pushing' or 'buttons' or which excludes 'pulling' and 'levers'. If she pulls the lever out of habit, then the original characterisation of her habit must have been something like:

Alice has:

5. The habit of adjusting switches when thinking

This move is identical to the move from (1) to (4) – from *the habit of pressing the red button* to *the habit of pressing a button* – when we considered (*Interfering Cathy*) and the green button. In both cases, the connection between the individuation of Alice's habit and what Alice does intentionally when her habit manifests is maintained. This is because (1), (4), and (5) all specify Alice's habit in terms of something she does intentionally when her habit fully manifests. So if (5) is correct, (1) is wrong, but not in a way which is problematic for my purposes, since Alice still pulls the lever intentionally.

Also, importantly, if (5) is correct, then it does nothing to suggest that Alice thereby presses the red button out of habit when she pulls the lever, just as Alice's pressing the green button in (*Interfering Cathy*) did not to suggest that she really had the habit of annoying Bert.

How could we choose between (1) and (5)? The consideration from (*Complaining Bert*) and (*Box and Lever + Knowledge*) about the force of habit may help. If Alice feels the force of habit direct her to pull the lever, then that is evidence for (5) over (1) since it suggests the force of Alice's habit is indifferent to whether Alice adjusts levers or buttons. If the force of habit directed her to press the button, she would exhibit restless behaviours as I have discussed. But neither possibility indicates that Alice can do whatever she is in the habit of doing both unintentionally and out of habit.

This indicates what we should say if Alice pulls the lever, but not out of habit (if she just has a one-off urge, say). If her pulling the lever out of habit in (*Box and Lever*) was not evidence for her pressing the button out of habit, and her pressing the green button in (*Interfering Cathy*) was not evidence for her interrupting Bert out of habit, then Alice's casually pulling the lever cannot be evidence for her pressing the button out of habit. The theme connecting these cases is that we cannot make sense of how Alice or her habit could be sensitive to facts she does not know and is not causally sensitive to, but doing something habitually requires some sensitivity to one's circumstances. If this sensitivity is not mediated by some basic bits of Alice's psychology such as her experience or knowledge, then her behaviour's apparent sensitivity to those facts is simply mysterious. Just *how* could she press the button out of habit if it is totally obscured from her and she has to pull a lever to get at it? This is no less mysterious than if Cathy had put the button on a space station and wired a

contraption into Alice's head so that whenever Alice thinks deeply a signal is transmitted and the button automatically pressed, and then we said that whenever Alice is thinking deeply she presses the button out of habit. One case is wackier than the other, but their conclusions are equally mysterious, and we should avoid it at all costs. Therefore, we should say of (*Box and Lever*) that, if Alice unintentionally presses the button by intentionally pulling the lever (for whatever reason), she does not thereby press the button out of habit.

All this over-complicated working-through of cases shows us that we must individuate Alice's habit in terms of something such that, if she does it habitually, she does it intentionally. Although whenever her habit manifests she annoys Bert, the outcome of counterfactual variation in the cases shows that her habit only manifests in annoying Bert accidentally, and that its proper manifestation-type – the manifestation-type by which we identify it as that particular habit – is Alice's pressing the red button. Further, that is the proper manifestation-type *because* that is what Alice does intentionally whenever her habit fully manifests. And the counterfactual variations showed that if (1) is wrong, then whichever the right individuation is – for example, (4) or (5) – it will *still* always be true that whichever act-type the characterisation mentions must be something Alice does intentionally whenever she does it habitually. Which act-type is individuative of her habit is determined by these facts about what she does intentionally. But what goes for Alice goes for habit-bearers in general. Therefore, the argument puts us in a position to accept *Goal* by employing the strategy I outlined in Section 1.3. Therefore, for any subject, *S*, and any act-type, *A*: if *S A*-s habitually, *S A*-s intentionally.

Conclusion

I have argued that we can move from the *Manifestation Thesis*, that habits are partly individuated by manifestation-types, to the view that if *S A*-s habitually then *S A*-s intentionally. My argument has worked by providing an example of habitual action where what Alice does intentionally and unintentionally are sharply divided, and then shown that only the things she does intentionally when her habit fully manifests can be said to individuate her habit. And I have tried also to show that it is *because* she does these things intentionally that they are candidates for being done habitually, since Alice cannot press the button unintentionally and habitually. I have not given any single, definitive argument for the claim that if *S A*-s habitually then *S A*-s intentionally, but I think that is where the conceptual breadcrumbs have led us.

In the midst of the example-wrangling, it will have been easy to lose sight of the point of all this: defending the Rationalising View of Habit. Throughout the thesis, I have been working the following three premise argument-sketch:

1. Habits are mental properties (broadly construed);
2. If mental properties figure in explanations of why someone does something intentionally, then they figure in broad rationalisations of why they do that thing;

3. Whenever habits figure in explanations by manifesting, they figure in explanations of why the habit bearer does something intentionally

C) Therefore, whenever habits manifest, they figure in broad rationalisations.

I defended Premise (1) at the end of Chapter 2, and defended Premise (2) in Chapter 3 partly by defending the biconditional that *S A*-s intentionally if and only if *S A*-s for a reason. At the end of Chapter 3, I showed that, since there are at least some cases where habits manifest in a person's doing something intentionally, habits must at least sometimes figure in broad rationalisations. However, that argument leaves it open whether there are other sorts of cases where what one does habitually is done unintentionally. I called the resultant view the 'Weak Rationalising View'. In this chapter, though, I have argued that it is *not* open whether there are cases where someone does something habitually but unintentionally: everything done habitually is done intentionally.

I also showed that Premise (3) follows easily from this thesis. To repeat: if everything done habitually is done intentionally, then whenever a habit manifests it figures in an explanation of why the habit-bearer does what they do; one of the things they do is habitual; and, by *Goal*, that thing is also done intentionally; therefore the habit figures in an explanation of why the habit-bearer does something intentionally. That is how Premise (3) follows from *Goal*. And, from it, the Rationalising View follows suit.

This completes my positive case for the Rationalising View of Habit. I hope that it is at least a fairly compelling line of thought, and that I have given some strong reasons for believing it. However, parts of the argument are very controversial. In particular, I think the conclusion of this chapter will be seen as obviously false by many philosophers. Aren't there a wealth of examples of things we do habitually but unintentionally? Can't we bite our nails or shift our posture habitually and unintentionally? And what about habit-slips, where somebody's habit manifests in their going against their own set intentions? Despite my argument (which I am sure is also controversial) this conclusion may seem clearly wrong, and a number of powerful arguments against it immediately suggest themselves. Therefore, the purpose of the next chapter is to consider and reject three of these arguments aimed at showing that some things done habitually are not done intentionally. If I can show that these arguments fail, then the Rationalising View of Habit will be significantly stronger for it.

CHAPTER 5

SLIPS, EXPLANATORY CONTEXTS, AND SUB-INTENTIONAL ACTIONS

In the previous chapter, I ended my argument for the Rationalising View of Habit by arguing for the view that if *S A*-s habitually then *S A*-s intentionally. This is clearly controversial and somewhat surprising. I think that if any part of my argument looks the weakest it will be this. Therefore, in this chapter, I will consider three arguments against my claim, all of which stem from different kinds of quite powerful counterexample. In Section 1, I consider an argument from the possibility of action-slips brought about by an agent's habit. In Section 2, I consider the idea that some habits only fulfil their explanatory role when we see them as habits of doing things unintentionally. Finally, in Section 3, I consider the argument that if there are sub-intentional actions – actions which are not someone's doing of anything intentionally – then some habitual actions are sub-intentional. In each instance, I try to show that the arguments fail. Therefore, my claim that everything done habitually is done intentionally not only receives positive support from the previous chapter, but is also defensible against powerful objections. This in turn supports the Rationalising View of Habit.

Section 1 – The Argument From Slips

There is a powerful argument against the view that if *S A*-s habitually then *S A*s intentionally which I will call 'the argument from slips'. It has been made persuasively by Komarine Romdenh-Romluc (2013). This argument also runs best with an example:

> (*Doris Slips*) Doris has a habit of cycling a particular route to work. On Saturday, she organised to go for a picnic with her friends. Saturday comes around, and Doris hops on her bike to cycle toward the park. The first part of her park-ward journey is the same as the first part of her work-ward journey. Doris is feeling bothered by a family problem, so she is cycling absent-mindedly. She's absent-minded enough that instead of turning left at the junction toward the park, out of habit she turns right toward work. Doris continues cycling work-ward, taking the habitual corners,

and arrives at work. Only then does she realise what she's done – 'Rats!' she cries, and turns back, exasperated.

Romdenh-Romluc takes a story like this[74] as fodder for an argument:

1. Doris cycled to work habitually;
2. Doris did not cycle to work intentionally;
3. Therefore, not everything Doris did habitually she did intentionally.

If the conclusion is true, then Doris is a straightforward counterexample to the view that if *S A*-s habitually, then *S A*-s intentionally. And since cases like this are common, we have reason to think the counterexample is pervasive.

This argument has immense plausibility. Its premises are so plausible they look barely worth defending, although Romdenh-Romluc gives ample defence of Premise (2) (Romdenh-Romluc, 2013, pp. 5–8). She argues, rightly, that Doris's intention to cycle to the park is not sneakily replaced by a covert intention to go to work, since she is frustrated and annoyed when she realises her mistake. It would be hopelessly *ad hoc* to say that as soon as she got there she lost the intention to be there, and this explains her frustration. Doris certainly didn't change her mind, and she does not seem to have conflicting intentions pulling her in different directions – she is set on seeing her friends, not on working. So, Romdenh-Romluc says: "[T]he appeal to habit [...] is supposed to explain why I begin cycling work-wards, even though I intend to go for a picnic. In other words, the explanation holds that my habit produces my action *in the absence of any relevant intentions*. [...] Thus, there are cases of habitual action that are not brought about by the agent's intentions" (Romdenh-Romluc, 2013, p. 7).

Officially, Romdenh-Romluc is arguing against the view that intentions are mental states with plan-like contents, that intentional actions are caused by them, and that all actions are intentional. This is one way of putting the so-called 'Standard Story of Action'.[75] That is not my view. The only view of intention I have defended is that to do something intentionally is to have practical knowledge that one is doing it. But whilst some of Romdenh-Romluc's terminology does not neatly fit my view, her case for Premise (2) seems to carry over to my view of practical knowledge. Doris is cycling work-wards, but does not seem to have practical knowledge that she is. So she is not cycling to work intentionally; she is surprised and annoyed when she gets there. So I am therefore happy to accept Premise (2).

Therefore, I must reject Premise (1). How is that plausible? What could possibly stop us from accepting it? Although Premise (1) may seem obvious, it is not an unquestionable feature of the story as described. Neither should it be; since that is precisely what I want to reject it would be tantamount to begging the question to stipulate its truth. This means we need to find what it is about the case which makes it seem so obvious that Doris cycles to work out of habit. I suspect it is these two facts:

[74] Romdenh-Romluc's own story is in the first-person. Further, in it, her character never gets to work – she realises what she has done half-way, and turns around.

[75] See Bratman (2000), Mele (2017), Searle (1979), and Smith (2012).

a) Doris has a habit of cycling that route to work;
b) Doris's habit does manifest in her behaviour, helping explain why she does what she does, where one thing it helps explain is that fact that she cycled to work;

These two facts of the case speak in favour of saying that Doris cycled to work out of habit; that her habit manifested in her cycling to work. But how? They do not straightforwardly entail it, because it is false that wherever a habit manifests and explains why a habit-bearer does something that the habit-bearer thereby acts habitually. For example, that a smoker's habit manifests is part of the explanation of why they try to quit, but they do not try quit out of habit. Therefore, we need some bridge principle taking us from (a) and (b) to Premise (1). In the following sections, I will consider and reject two potential bridge premises that might be invoked.

1.1 – Support for Premise (1): The Common Factor Thesis

One way of moving from (a) and (b) to Premise (1) is by accepting the following thesis:

For any agent S with habit of A-ing, H:
The Common Factor Thesis: *H* manifests only if *S* successfully *A*-s.

The thesis's name derives from the fact that it says that all manifestations of a given habit must be of the same kind; they must all be *A*-ings, say. That is, *A* is a common factor in all manifestations of *H*.[76]

According to (a), Doris has *H*, and according to (b), *H* manifested. Therefore, if the *Common Factor Thesis* is true, then Doris must have cycled to work out of habit since that is what her habit is characterised as a habit of doing. The *Common Factor Thesis* acts as a bridge which takes us from undisputed facts of (*Doris Slips*) to Premise (1) of the argument from slips.

Of course, it is only worth considering this route if the *Common Factor Thesis* is independently attractive. In fact, it has recently been defended by Christos Douskos (2019b). Douskos argues that, since (according to the *Manifestation Thesis*) habits are partly individuated by what they are habits of doing, it is impossible for a habit to manifest without its bearer doing just what they are in the habit of doing. "[T]he way one acts provides the habit's identity, and hence, a given habit can be manifested in only one way" (Douskos, 2019b, p. 992). This simple argument is very intuitive. One's habit of smoking roll-ups can't be manifested by smoking straights, or by rolling them and

[76] The name is also intended to alert the reader how similar the claim is to widely discussed views in the philosophy of perception and epistemology. For example, it is commonly held that a subject's empirical evidence in a case of seeing that *p* is the same as their evidence in a case of hallucinating that *p* on the grounds that the two cases are subjectively indiscriminable. Such a view is known as 'the highest common factor view of perceptual evidence', because it says that one's evidence is the strongest evidence a subject can have in both good and bad epistemic cases. See McDowell (1983, 2008) for criticism, and Wright (2002, 2008) for defence, of the view.

giving them away; one's habit of listening to music too loudly can't be manifested by listening to it too quietly; one's habit of going to work by a set route can't be manifested in just going half-way, or by going a different route.

This plausible argument for the *Common Factor Thesis* bolsters the case for Premise (1) of the argument from slips because, if the thesis is true, and Doris's habit manifested, then we know that it manifested 'all the way'. And since her habit is of cycling to work that particular route, we know she cycled to work habitually. Premise (1) is secured.

However, I think that the *Common Factor Thesis* is false. By arguing against it, I am therefore defusing this way of getting from the facts of (*Doris Slips*) to Premise (1). This will leave Premise (1) with substantially less support since one natural way of defending it is blocked.

To make this case, let me pick an example:

> (*Eric Pull-cord*) Whenever Eric goes to the bathroom, he turns on the light by pulling the cord. After doing this for a while, there is the usual sort of evidence that he has this habit: he need not deliberate about whether to do it, he experiences a force of habit, and so on.

Turning on the light is the act-type which putatively characterises Eric's habit given the evidence in the case. By the *Common Factor Thesis*, whenever Eric's habit manifests, he turns the light on. But reflection on this case calls the *Common Factor Thesis* into question. This is because, on some occasion, Eric might walk into the bathroom and pull the cord, yet the light doesn't come on because the bulb has blown. If we abide by the *Common Factor Thesis*, there are two options.

Firstly, if we rightly characterised Eric's habit in the first place, we should say that he did nothing habitually because he failed to turn the light on, meaning he did not do what he is in the habit of doing. But there is pressure to doubt this. When we think about why Eric pulls the cord, it is immensely plausible that he does so as a matter of habit – the very same habit that operates when the bulb is working. Why would it be that the bulb's being broken makes it the case that his habit does not figure in an explanation of what he does, or that he does nothing out of habit? The point is analogous to my point in (*Cut Wire*) and other similar cases from the last chapter: the changes to the environment are causally downstream of habit-bearers – in the wire's being cut or the lightbulb's being blown – and so have no impact on the explanation of why they do what they do. So the first option is implausible.

The second option is to accept that we were wrong to characterise Eric's habit as we did, since it is clear that it can manifest even when he fails to turn on the light. This fact shows us that it cannot be essential to his acting habitually that he actually turns on the light. We therefore take it as evidence that we must revise our judgements about what Eric is in the habit of doing. If we do this, then by the *Common Factor Thesis* we must find a common factor between the case where he turns the light on and the case where he fails because the bulb is broken. The most plausible common factor is Eric's *pulling the cord*. On this approach, the fact that he acts habitually even when he does not turn the light on is evidence that what he is in the habit of doing is really something less than

turning the light on – what he does habitually is *pull the cord*. Turning on the light is something like an *effect* of his doing what he does habitually.

But this is an unstable resting place. What goes for the characterisation in terms of turning the light on goes for the characterisation in terms of pulling the cord. Say one day Eric walks into the bathroom and reaches for the cord, but it has snapped off (somebody pulled too hard yesterday!) and he grabs at the air. The question of why Eric does this arises. But now we have set a precedent on what to say about such cases. By the *Common Factor Thesis*, we must deny that Eric's habit was really ever one of turning the light on *or* pulling the cord, but instead perhaps *moving his arm toward the cord*. But, under pressure from the theory, we have now rescinded from any natural characterisations of Eric's habit.

This conclusion is worse than it may seem, but it will take some terminology to explain why. Moving one's arm is a teleologically basic act-type: one does it without taking any means to do so.[77]

---

[77] This is an exceptionally simple way of characterising teleological basicness. It is similar to the thesis Jennifer Hornsby's original discussion begins with, albeit hers is stated in terms of *descriptions* of an action rather than in terms of things done. However, Hornsby's discussion ends with a very different view which involves what an agent knows how to do (Hornsby, 1980, Chapter 6). I think that for my purposes the simpler formulation is sufficient, since I am not giving an account of teleological basicness. Rather, I aim to use it in a formulation of a problem for the *Common Factor Thesis*. To that extent, refinements on the formulation can be ignored since it is not the detail that matters but the general idea.

There are those who have tried to argue against the very idea of teleological basicness. Michael Thompson (2008) and Doug Lavin (2013) have argued that there are no teleologically basic intentional actions because all intentional actions have stages which are themselves intentional actions. "The difficulty is to find a describable part of [an action] A, A*, which is something the agent did intentionally in order to do A, but which does not itself resolve into further sub-actions that the agent did intentionally in order to do A*" (Lavin, 2013, p. 276). But I am not sure this sort of challenge is to the point. Hornsby's notion of teleological basicness (which my formulation is a dim cousin of) is the notion of a relation between things a person can do, not between *actions*. It may very well be that every intentional action, ϕ, which is *S*'s intentionally doing something, *A*, that is basic for *S* has some parts, for example, ξ and ψ, which are also intentional actions (which means also that ξ and ψ themselves have intentional actions as parts). But that does not mean that when *S A*-s she takes some means to *A*, since means are *things we can do*, act-types, *not* actions. Actions are not means to anything, so the idea that all actions have actions as parts does not entail that all actions which are *A*-ings involve the agent's taking some means, *B*, to *A*.

Thompson and Lavin's 'initial segment argument' might sound like it can get a grip on Hornsby's and my notion of teleological basicness when stated like this: if I move my arm from location L1 to L3 via L2, then moving my arm from L1 to L3 cannot be teleologically basic since I do it by moving my arm from L1 to L2, and then from L2 to L3. Since space is infinitely divisible, whenever I move my arm from one point to another, I do that by moving my arm through an infinite number of other points. This argument does not mention actions, since moving my arm is an act-type.

But the argument fails, since it works by inserting adverbial phrases into characterisations of the act-type. It is true that to move my arm from one point to another, it must pass through other points, but the thing I do when I move my arm is just *move my arm*. There is no act-type which is *move my arm from L1 to L2*, or *move my arm from L1 to L3*. Adverbial phrases like 'from here to there' introduce information about the trajectory of my arm, not about what I do. To think otherwise would be to include all the prepositional and adverbial information about one's doing something in a characterisation of what one does. So if someone jumps for joy on a trampoline at 10am, they do not really *jump*. What they *really* do is *jump for joy at 10am on the trampoline from x-spatio-temporal-co-ordinates to y–spatio-temporal-co-ordinates*. But this means that nobody ever does the same thing as anyone else, since not all the same adverbial information applies to everyone. This is quite a radical and implausible view. Better to say that the adverbial information is left out of characterising *what* we do, and left to characterise *when*, *where*, and *how we do it*. Therefore, although my arm must go through L2 to get from L1 to L3, *moving from L1 to L2* is not my means to *move from L1 to L3*. Rather, I *move my arm* and it goes through all those points. That is, stripped of adverbial information as it should be, what I do to get from L1 to L2 is *move my arm* and what I get from L2 to L3 is *move my arm*, and that's also what I do to get from

The act-types *pulling the cord* and *turning on the light* are therefore teleologically non-basic: one does other things as means of doing them.[78] Eric moves his arm in order to pull the cord, and he must pull the cord in order to turn on the light. The teleologically basic thing – moving his arm – does not admit of the sort of failures I have been discussing, since one does not do basic things by doing something else.

Why does doing something non-basic always admit of such failures? Because if one *A*-s by *B*-ing, then *A*-ing and *B*-ing must be connected by some fact about the world. If I poison the inhabitants by pumping poisonous water into the house by operating the pump by moving my arms, then each of these things I do is mediated by some fact which may not have held. My poisoning the inhabitants is made possible by the fact that the water is poisonous; without that fact holding, I would fail to poison them. My pumping the water into the house is made possible by the fact that the plumbing connections between the pump and their house are in good working order; without that fact holding, I would fail to pump water into their house. And so it goes until we reach the basic thing I do: I move my arms. Now, certain facts make this possible: that I am an agent; that I have arms; and that I can move my arms. But these are necessary conditions on moving my arms, and once they are met, I can just do it. There is no possibility of failure deriving from a faulty connection between my *taking a means* to moving my arms and my moving my arms: there is no means to take. My moving my arms is the first link in the chain, so is immune to the sorts of mistakes possible in my doing non-basic things. This means that the *Common Factor Thesis* forces us to say that habits can only be characterised by non-basic act-types because they are the only things we can do which do not admit of these sorts of failures.

There are three things I want to say about this argument and its extreme conclusion. The first thing is to just emphasise how bad this seems. If every non-basic act-type admits the possibility of the sort of failure I have discussed, then the procedure we followed under pressure from the *Common Factor Thesis* pushes us to accept that we can only characterise habits in terms of basic act-types. For any habit putatively characterised in non-basic terms, there will be some possible mistake which will force us to drop that characterisation if we hold the *Common Factor Thesis*. So, in the end, there will be no habits of turning on the light, of going on holiday to the Isle of Wight, of brushing one's teeth, or of interrupting people. There will only be habits of moving one's body thusly or suchly.[79] And that just seems intolerable – surely it must be possible for Eric to have the habit of turning on the light despite the fact that he sometimes fails to do so successfully.

L1 to L3. So I take no means to move my arm from L1 to L3 – I just move my arm all the way. So we are left with no argument against teleological basicness, therefore I will use the notion freely.

[78] All references to basicness in this section refer to teleological basicness, so I will suppress the 'teleological' terminology.

[79] This is one classic way of delimiting what is basic, even if it is not always put in terms of teleological basicness (Sandis, 2010). There are reasons to think the classic way is wrong: for example, Hornsby thinks that speaking is more teleologically basic than making mouth movements (Hornsby, 1980, p. 79). And Annette Baier (1971, 1972) has rather a sophisticated view of basicness on which quite a lot more than bodily movements can be basic. I am very sympathetic to these extensions of teleological basicness to things done where what is done involves 'extra-bodily' or 'not-purely-physical' things. Nonetheless, the things we do the descriptions of which are primarily couched in terms of bodily movements do have a special status when it comes to teleological basicness. This is simply because it's very hard to do anything without moving one's body. The finer points about basicness do not matter here, though. My argument works

The second thing I want to say is to make a claim of guilt by association by noticing the parallel between the argument I have just sketched and a number of rightly maligned 'arguments from illusion and hallucination'. For example, one argument starts with the naive thought that perceptual phenomenal character is determined by ordinary perceived objects. Then it notes a certain kind of 'mistake' or 'misfire', perceptual hallucinations, where there are no such ordinary perceived objects. At that point, we have a choice: we either deny that there is any phenomenal character to hallucinations, or we say that phenomenal character is not determined by ordinary perceived objects. [80] Since it seems to most so obvious that hallucinations have phenomenal character, we must take the second horn. So perceptual phenomenal character is not determined by ordinary objects, and is instead determined by non-ordinary objects like sense-data or relations to representational contents. Another, similar, argument starts with the naive thought that perception is a capacity which can provide indefeasible reasons for knowing facts about the world. Then it notes the existence of illusions, in which one's perceptual experience leads one to believe falsely. Then, the choice: either we deny the existence of misleading experiences; or say that perception is not really a capacity which can provide indefeasible reasons for knowledge.[81] Since there are misleading experiences, we must deny the attractive naïve view of perceptual knowledge.

There is a clear structural analogy between these arguments in the philosophy of perception and the argument from the *Common Factor Thesis* to the extremely limited view of habits. And these arguments, though they are powerful and compelling, are widely thought to fail for a number of important reasons. Jonathon Dancy (1995) has complained of their methodological focus on bad cases; John McDowell (1983, 1986, 2011a) argues that the idea that perceptions and illusions share a common factor in the evidence they provide us leaves us totally unable to account for perceptual knowledge. Finally, M.G.F. Martin (2004, 2006) has argued that naïve realism about perceptual experience is attractive enough that we should fight tooth-and-nail to earn the right not to posit common phenomenal factors between hallucinations and perceptions. I think all three approaches are broadly appropriate.[82] A methodological focus on bad cases generally obscures the phenomena we are trying to understand by blinding us to what happens when things go as they should. And there really *is* reason to fight to maintain a conception of perceptual experience and knowledge which is not intolerably interioristic. If that is right – that we should, in general, resist this form of argument when an attractive and natural characterisation of some important feature is at stake –

---

by showing that no habits can be characterised in non-basic ways. Whatever the class of basic things done exactly amounts to, this conclusion should be unacceptable because there are clearly habits of doing things which are not basic for the habit-bearer.

[80] Bill Fish (2009) takes the first option. Most philosophers take the second, representationalist, option (Byrne, 2009; Schellenberg, 2018; Siegel, 2010). We might think of M. G. F. Martin's (2004, 2006) disjunctivism as attempting to provide a middle-ground solution.

[81] To my mind, Charles Travis (2004) is best read as taking the first option. Most philosophers of perception take the second.

[82] To be clear, the *approaches* to the arguments from illusion are correct, however it may be that the positions which motivate the approaches are ultimately untenable. I am not recommending dogmatism about controversial aspects of perception or action. But I am recommending that we keep in mind what an attractive picture of the phenomenon we want to account for might be. Arguments from illusion almost invariably push us away from attractive pictures, and that is a reason to be sceptical of them and subject them to serious scrutiny.

then we should also resist the *Common Factor Thesis*, since that is what drives us towards the intolerably limited conception of habits.

So far, I have tried to show how unattractive the limited view of habits we get from the *Common Factor Thesis* is, and that it suffers from guilt by association with arguments from illusion and hallucination. However, I am aware that these are not decisive arguments; the complaints may fall on deaf ears. Therefore, here is the third thing I want to say: I think there is a strong argument for rejecting the limited conception of habits.

Firstly, remember that forming a habit of *A*-ing requires having *A*-ed repeatedly. Naively, this is easy to accommodate: to form a habit of running ten miles, I must have run ten miles before. But on the restricted view I am considering, one's habit must be a habit of making thus-and-such bodily movements. But which bodily movements do I repeat every time I run ten miles such that I can form a habit of making those movements? Running ten miles on different routes involve very different bodily movements – think of the difference between running a route with lots of uphill parts and one which is flat. Running the same route on different occasions is similar – if it is wet I may avoid puddles; if it is hot and dry I may run slowly, dragging my feet. There is no set of bodily movements which, in every case of running ten miles I have repeated such that we can say that I have a habit of making those movements. This carries over to a vast array of habits. So the restrictive view either cannot account for the fact that habit-formation requires repetition, or it must deny that repetition is strictly necessary. Both views seem rather hopeless.

I think together, these three responses hep make a very strong case against the extremely restrictive view of habit and therefore against the *Common Factor Thesis* which saddled us with it. So if the *Common Factor Thesis* is false then it cannot act as a bridge principle taking us from the facts of (*Doris Slips*) to Premise (1) of the argument from slips. That is, we cannot infer from the fact that (a) Doris has the habit of cycling that route to work and (b) that her habit manifested, to the fact that she went to work out of habit in (*Doris Slips*). This is because it is possible for Doris's habit to manifest in her doing less than going to work, even if that is what her habit is canonically identified as a habit of doing. So Premise (1) of the argument from slips, which says that she went to work out of habit, is left without support.

1.2 – Support for Premise (1): The Argument from Best Explanation

I was considering the *Common Factor Thesis* because it seemed to provide a bridge between facts about Doris's situation when she unintentionally cycles to work and Premise (1) of the argument from slips, which said that Doris cycled to work habitually. I argued that some bridge principle was needed to take us from the facts of the case to Premise (1) and that the *Common Factor Thesis* both did that work and had independent support expressed by Douskos. However, I have argued that it is false. Therefore, it cannot be invoked to support Premise (1). So, the argument from slips has a crucial, intuitive, but unsupported premise.

Romdenh-Romluc might opt for a different approach to defending Premise (1). The option I want to consider is the idea that the best explanation of why Doris cycles to work by this route is that her habit manifested. And if that is the best explanation, then we should accept that she went to work out of habit. This embodies no commitment to the *Common Factor Thesis*, and preserves the feeling that it is just obvious that the facts about (*Doris Slips*) stand in favour or Premise (1).

My response is twofold. Firstly, I will show how accepting this is consistent with my denial of Premise (1), meaning this argument fails. Secondly, an argument *against* Premise (1) will emerge from my first point.

So, I want to agree that the fact that Doris's habit manifested is the best explanation of why she cycled to work by that route. When she reached the cross-roads, a little absent-minded, she turned right out of habit and continued to take a route which, on reflection, she would prefer not to have taken because it led her to work rather than to the park. Now, she took these turns intentionally – she was in control of her bike and herself; she knew what she was doing when she took the turns (*I'm turning right*; *I'm cycling on this road*; *I'm trying to avoid those pedestrians*). And after some cycling, she ended up at her workplace. So the best explanation of why Doris ended up at work is that her habit manifested.

But all of this is consistent with my claim that Doris did not cycle this route to work habitually. Sometimes, *S*'s habit of *A*-ing by *B*-ing may manifest, *S* may *A* because their habit manifested, however *S fails to A out of habit* and only *B*-s out of habit. I will call cases like this 'deviant cases' for reasons which will become clear shortly. I want to argue as follows: there are uncontroversial deviant cases; we can formulate a sufficient condition on deviance; (*Doris Slips*) meets the sufficient condition; therefore (*Doris Slips*) is deviant.

Here is a paradigm deviant case:

> (*Fred's Holiday*) Fred has a habit of going on holiday to Devon, and habitually books his holiday via bookingz.com. Think of the form of Fred's habit as a habit of *A*-ing by *B*-ing. One year, Fred is intending to break his habit of going to Devon, intending to go somewhere else instead. But he still habitually goes to book via bookingz.com. He sees a cottage advertised as being in Cornwall and snaps it up. He spends a nice week there, but, unbeknownst to Fred, the cottage is actually in Devon.

This means that Fred's habit of *A*-ing by *B*-ing has manifested in him *B*-ing habitually, and this explains why he *A*-s. But we should not say that he *A*-s out of habit. He *A*-s, but it's a sort of accident that he does; the force of habit does not move him to *A*. Fred's case is like that of someone who is ordered to sit down in a language they don't understand, and who sits down because the order sounded threatening. In one sense, they sat down because of the order, but they did not sit down *because they were ordered to sit down*. The order had an impact on what they did, and it has the right impact, but not in the right *way* because of the hearer's lack of understanding. The same is true of Fred: his habit has an impact on what he did, and it had the impact appropriate to the habit (the 'right' impact), but not in the right way. The connection between his habit and his behaviour is

one of genuine explanation, but of a deviant kind. This is why I called these cases 'deviant'. This is my first point: the existence of deviant cases means that we cannot infer that someone *A*-ed out of habit just because the fact that they *A*-ed is best explained by the fact that their habit of *A*-ing manifested.

That should be enough to block Romdenh-Romluc's best-explanation argument for Premise (1). But by considering (*Fred's Holiday*), we can see our way to an argument against that premise. This is my second point. Notice, (*Fred's Holiday*) is deviant because there is a lack of knowledge (as is the case of the order). And what Fred doesn't know is that, in this circumstance, what he is doing when he is *B*-ing is a way of *A*-ing. Lacking that knowledge seems to be sufficient to generate the deviancy, for if Fred had known the cottage was in Devon, he would not have booked it. So we know that when a habit-bearer lacks knowledge that what they are doing is a way of *A*-ing, *that is sufficient to generate a deviant case where S's habit figures in an explanation of why S A-s, but S does not A out of habit*.

This is just the position Doris is in in (*Doris Slips*). She has a habit of *A*-ing by *B*-ing, just like Fred. She takes the turns intentionally, and therefore knows, when she is taking the turns, that she is taking them. But she is not intentionally cycling to work, since she does not know that she is cycling to work. Therefore, she meets the sufficient condition for being in a deviant case. But if (*Doris Slips*) is deviant, then Doris does not cycle to work habitually on that occasion, just as Fred does not, on that occasion, go on holiday to Devon habitually. So Premise (1) of the argument from slips is *false*. This is good news for me: not only do I *have to* deny it; I have an *argument* for thinking that we must.

One might object that it is incoherent of me to accept that Doris does not go to work intentionally *and* that she cycles the route intentionally. After all, if Doris knows that cycling this way is a way to cycle to work and not to the park (which presumably she does since this is her usual route to work), then how could she fail to know that she is going to work? How is my view that Doris takes the turns and the route intentionally consistent with *denying* that Doris goes to work intentionally, when she knows that this route is the route to work?

Two things. Firstly, it is a sad fact that subjects can fail to know things that they are in a perfect epistemic position to know. We can simply fail to connect some things we know together in the right way. Perhaps if someone knows that *p*, that *if p then q*, but fails to infer that *q* then we should hold them epistemically responsible. But it is *possible*; we have to allow that sometimes people just fail to make the requisite connections.

Secondly, the sort of knowledge of what one is doing that is relevant to doing it intentionally is *practical knowledge*, which is non-observational and non-inferential. But the objection is asking how Doris could go to work unintentionally given that she has the resources available to her *to infer that that's where she is headed*. But the objection therefore misses the mark. No amount of inferential knowledge of where Doris is going can make her going there intentional, since inferential knowledge is not practical knowledge.

Romdenh-Romluc might try a different objection. I have said that, in the story, Doris turns right (in the direction of work, away from the park) intentionally. Romdenh-Romluc might ask how

that can be consistent with having a standing intention to go to the park, which means having a standing intention to turn left. For, perhaps, if Doris intentionally turns right, then she intentionally turns away from the left. But she has a standing intention to turn left. So why the sudden conflict?

Firstly, I don't think Doris does intentionally turn away from the left just because she intentionally turns right. This is for the same reasons as given above: she may simply fail to make the inference, and anyway, the inference wouldn't be enough to make it something she did intentionally. Secondly, it is not clear to me why we should avoid positing some degree of conflict in the case. Surely that's part of the interest of habit-slips – they involve an agent doing things contrary to their standing intentions because of their habits. That is a conflict, and it is generated *by the habit*. So if there is reason to think that things done habitually are done intentionally, we also have reason to posit conflict in intention.

I think this is enough to ward off the argument from slips. Whilst Premise (1) looks eminently plausible, I tried to show that it is not true because (*Doris Slips*) is a deviant case: Doris's habit figures in an explanation of why she goes to work, but it does not do so by manifesting, so she does not go out of habit. I have also tried to show that Premise (1) actually has little support that goes beyond its intuitiveness. Finally, I have argued that the consequences of denying it are not disastrous. Therefore, I think we should reject it. In which case the argument from slips fails, and my view that if *S A*-s habitually then *S A*-s intentionally is protected.

Section 2 – Interruptions and Feminine Comportment

In Chapter 4, I argued that if *S A*-s habitually then *S A*-s intentionally. I did this by arguing that we cannot even identify habits unless we identify things that habit-bearers do intentionally when their habits properly manifest. But pressure can be put on this by thinking about habits whose role in explanations seems precisely to require that they explain why someone *A*-s, but where they do not *A* intentionally. Can't brash and un-self-conscious people have a habit of rudely interrupting others without paying any consideration at all to whether others are speaking, so that we should not credit them with intentionally interrupting people? What about some pervasive habits of members of social groups which they have in virtue of norms governing those groups? For example, a bit of anecdotal sociology[83] tells us that (some) women in Western society sit, and throw, and run in ways which we regard as distinctively feminine. We can assume that some of these ways of moving are habitual, and it seems very plausible that the correct description is that they have habits of sitting, throwing, and running in a feminine way.[84] But (at least some) women do not intentionally sit in a

[83] See Young (1980) for a phenomenological argument for this claim. See Preston (1996) and Ruggeri (2019) for a phenomenological response. The sociological evidence for Young's claim seems fairly compelling (Ehl et al., 2005; Evaldsson, 2003; Fredrickson & Harrison, 2005; Langendorfer & Roberton, 2002; Runion et al., 2003). However, see Downey (2010) for a very interesting critical cross-cultural take on the empirical research and its theoretical basis.

[84] A careful reader may recognise this from my Introduction as a *habit of style*, where I explicitly said I would leave them aside because I do not know what agentive styles *are*. For me to argue by employing an example of a habit of style may therefore seem like piracy. "The reader should take notice that my Jolly Roger is now unfurled" (Frankfurt, 1969, p. 835).

feminine way, even though they may intentionally sit in the way that they do. The challenge for me is to say what is wrong with these habit-ascriptions when their explanatory role seems to require that the relevant agents do not do the habitual thing intentionally.

I will first lay out some characters, Gene and Hannah, who each have a habit which is characterised correctly exclusively by either the left *or* right columns:

	Thin and Intentional	**Thick and Unintentional**
Gene	The habit of speaking over others	The habit of rudely interrupting
Hannah	The habit of sitting with her knees crossed *thusly* (with legs crossed at the knees, say)	The habit of sitting femininely

Gene's and Hannah's habits can either be best individuated in a thin way, which says what they do only in terms which are not laden with socio-cultural evaluative concepts, or a thick way, which does encode such concepts. We can think of the doing the things in the thin columns as *counting as* doing the things in the thick columns in virtue of doing them in certain sorts of social contexts. The '__ counts as __' relation does not generate an identity relation, since one can sit thusly in a different context of norms and not be sitting femininely, and there may be cultures where speaking over others is not rude. Therefore, these are different things done, even though doing the thin things counts as doing the thick things in certain contexts.

Importantly, thickness is not the only difference between the characterisations in the two columns. Gene intentionally speaks when others are speaking but is too un-self-conscious and self-centred to know that what he does counts as rude, so he is not rude intentionally. Hannah intentionally sits thusly (with a certain arrangement of her legs), but it does not cross her mind that she might thereby be sitting in a feminine way. So she does not intentionally sit in a feminine way. So the cases are structured so that only the thin thing is done intentionally by both Gene and Hannah.

The charge laid to my account is that the best characterisations of Gene's and Hannah's habits are found in the thick and unintentional column. The reason is that only the thick characterisations are apt to serve in explanations of the right kind. For example, Hannah's habit, we can assume, is induced by growing up in a culture with certain sorts of Western sexist norms. It is those norms of femininity which play the most important role in explaining why Hannah sits how she does, and it is those norms she inadvertently perpetuates by so sitting. Given the explanatory nexus that Hannah's habit fits into, the thought is that her habit must be characterised in the thick way. The thin characterisation seems as out of place in the sociological explanation as kinematic descriptions of a dancer's movements seem out of place in an explanation of why they dance so beautifully. The idea is that, given the explanatory nexus, it just should not *matter* whether Hannah sits femininely *intentionally*. Indeed, it may even be thought that to explain how it is possible for Hannah's behaviour to have been shaped by these norms without her knowledge, it is positively required that we go for the thick unintentional description.

The same goes for Gene, whose behaviour is shaped by his privileged youth and obliviousness to what others are trying to say. As with Hannah, given the sorts of explanations Gene's habit figures in, and given his lack of awareness of his rudeness, it may seem a requirement that we opt for the thicker characterisation.

This argument is appealing, but it doesn't seem quite right. Consider a modification on Hannah's situation:

> (*Gender Revolution*) Hannah has been in a coma for a while. In that time, there has been an overhaul in her society's gender norms such that sitting *thusly* now counts as sitting masculinely rather than femininely. Now, to sit femininely, one must sit *suchly* (with legs spread wide). One day Hannah wakes up, knowing nothing of the gender revolution. The coma itself has no effect on her habit. Hannah gets out of bed to sit at a table to eat some breakfast.

How will Hannah sit, when she sits at her table? The thick description predicts that Hannah will sit *suchly* because sitting suchly is now the way of sitting which counts as sitting femininely. And *that* is what she is in the habit of doing. Therefore, if her habit manifests, it should probably manifest in her organising her legs in a way which is very unfamiliar to her.

But that would be very strange. After all, the gender revolution has had no effect on Hannah. She knows nothing of it, has no psychological grip on it, and it has not even had a causal bearing on her. If she sat suchly because that is what now counts as sitting femininely, it would be simply mysterious what connection there could be between Hannah and the overhaul of gender norms such that Hannah's habit and behaviour is sensitive to the revolution.[85]

To avoid this whilst retaining a thick characterisation, one might think that Hannah will sit thusly (with legs crossed at the knees, say) as she always did, but say that the change in the social context means that her habit is now a habit of sitting masculinely. But whatever this may have going for it, the option fails. Having a habit of *A*-ing requires one have *A*-ed in the past, and when Hannah sits thusly after waking up, although she will be sitting masculinely, she will never have sat masculinely before. So she cannot be in the habit of doing so.

So we have three facts to respect: Hannah's habit manifests when she sits; Hannah will sit thusly (not suchly); and she does not have a habit of sitting masculinely. It seems that the only way to accommodate these facts is to accept that Hannah's habit was always a habit of sitting thusly, with legs crossed at the knees. We should opt for the thin and intentional characterisation.

The same goes for Gene. In a culture where speaking over others is not rude, Gene will still habitually speak over others. But he won't be being rude. So either his habit has disappeared, or

[85] This is similar to the points made in Chapter 4 when I discussed cases like (*Cut Wire*) and (*Interfering Cathy*). The point seems quite general: habits cannot be sensitive to changes which the habit-bearers are not causally or psychologically sensitive to. In that respect, there is a disanalogy between habits and mental contents on the externalist conception of them, for example in Burge (1979).

stopped operating, or has been replaced. But why think any of this? Better to think Gene has a habit of speaking over others, not of being rude.

So in fact, although it looked like the explanatory contexts of our habit-attributions supported picking out Gene's and Hannah's habits in terms of things they do not do intentionally when their habit properly manifests, this isn't right. The explanatory role of habit-attributions surpasses the ones the argument appealed to, and the view has odd implications for when habit-bearer's move between systems of norms.

So I don't think that we should opt for the thick and unintentional descriptions in these cases. It is important to stress, though, that this is *not* because they are *thick* characterisations (characterisations in terms of things people do which figure on the right-hand-side of the *__counts as__* relation). The problem is that neither Gene nor Hannah do the thick things *intentionally* when their habit fully manifests. Gene's brother, Ian, could very well have a habit of rudely interrupting people so that, when his habit manifests, he rudely interrupts them intentionally. What would happen to Ian in the possible culture where speaking over others doesn't count as rudely interrupting them? Well, he would become frustrated and try to find some other way of rudely interrupting them. And if there is no mechanism for being rude in the culture, then his habit might partially manifest in frustrated attempts. Perhaps he will shout, stamp his feet, and do all sorts of other things just to get the locals to falter in speech and feel offended. All of this is consistent with the view defended in this thesis, and so there is no general block on characterising habits in terms of thick, norm-governed manifestation-types. The problem only arises when we want to do that in cases where the habit-bearer does not do those things intentionally when their habit fully manifests.

Section 3 – Sub-intentional Actions?

There is a final argument against the view that if *S A*-s habitually then *S A*-s intentionally which I want to consider. The argument starts with the observation that there seem to be some actions which are not the agent's doing anything intentionally – they are *sub-intentional*. Typical examples include absentmindedly biting one's nails, minutely shifting one's posture, fiddling with one's hair, and moving one's tongue around the mouth. Call these putative actions 'sub-intentional actions'. The argument then proceeds like this:

1. There are sub-intentional actions;
2. If there are sub-intentional actions, then, sometimes when *S A*-s habitually the action, ϕ, which is *S*'s *A*-ing is sub-intentional;
3. If *S A*-s habitually and ϕ, which is *S*'s *A*-ing, is sub-intentional, then *S* does not *A* intentionally;

C. Therefore, some things done habitually are not done intentionally.

I think this is a very powerful argument. Premise (3) just follows from the definition of 'sub-intentional action', so the crux of the argument is premises (1) and (2). The idea behind Premise (2) is that there is so much overlap between typical examples of habitual action and sub-intentional action that, if sub-intentional actions exist, some of them are probably also manifestations of habit. Perhaps we do not typically move our tongues or shift posture habitually, but we do habitually fiddle with jewellery, tap the table, crack our knuckles, and bite our nails. If there are sub-intentional actions, at least many of them will be our doings of these things. And if so, then it seems immensely plausible that these cases will overlap: the doings of some of the things we do habitually will be sub-intentional. Therefore, in those cases, the things we do habitually will not be done intentionally.

Strictly speaking, I think I am within my rights to deny Premise (2) on the grounds of my arguments in Chapter 4. I could stubbornly maintain that because everything done habitually is done intentionally, whilst there could be sub-intentional actions, none of them can be habitual actions. However, I would prefer not to take that route. For it *does* seem very plausible that *if* there are sub-intentional actions then some of them are also manifestations of habit. I do feel the force of the consideration in the previous paragraph.

Instead, I wish to cast doubt on Premise (1), the claim that there *are* sub-intentional actions. I will not aim to argue that it is false (although I think it may be). Rather, I want to show that the arguments for it are substantially less compelling than they seem, and that in fact it is very hard to show that there are sub-intentional actions. Therefore, I aim to force a stalemate: if one wants to use this argument against me, one needs to make a much stronger case for Premise (1).

First, I will present the case for the existence of sub-intentional actions. Secondly, I will attempt to show why this case fails.

3.1 – The Case for Sub-Intentional Actions

Although the vast majority of philosophers of action deny that there are sub-intentional actions, Brian O'Shaughnessy (1980) and Helen Steward (2009, 2012a) have both defended their existence. They argue for this by first delimiting a class of phenomena which I call the 'Target Phenomena': examples of agents fiddling with jewellery, or shifting in their seat. The Target Phenomena are simply those events which stand as plausible candidates for being sub-intentional actions. O'Shaughnessy and Steward then mount arguments in favour of thinking that the Target Phenomena are not the agent's doing anything intentionally, and that they are nonetheless actions. Call arguments against the intentionality of the Target Phenomena 'negative arguments' and arguments for their agential status 'positive arguments'. O'Shaughnessy and Steward utilise the same negative argument, and there are two positive arguments: one unique to Steward, and one they both make. I will present all three arguments in turn.

The negative argument aims to show that the Target Phenomena are not intentional under any description.[86] There are two ways one could do this. One could defend a particular view of intentional action, and then argue that the Target Phenomena do not meet its requirements. Or, one could argue that the Target Phenomena do not meet a minimal necessary condition which all views of intentional action agree on. O'Shaughnessy and Steward opt for the latter. They argue that the Target Phenomena do not meet the *Awareness Condition*:

> *Awareness Condition*: An action is intentional under some description only if the agent is aware of the action.[87]

The *Awareness Condition* is very weak. It does not require practical knowledge (as I argued is required in Chapter 3) or non-practical knowledge; it does not require that the awareness be justified belief, mere belief, perceptual experience, or attention; and it does not even connect the description under which the action is intentional to the content of the agent's awareness. As such, it is so weak that I think everybody in the literature – regardless of the diversity of views – would agree on it. Therefore, if one can show that the Target Phenomena do not meet the *Awareness Condition*, one has succeeded in making the negative part of the case for sub-intentional actions.

The negative argument appeals to the plausible thought that, for an awful lot of the bodily behaviours we engage in throughout the day, we are simply unaware of them. O'Shaughnessy invites us to attend to our tongue, or feet, or hands as we read, and notice that we can make the discovery that we are moving them. I can simply *find out* that I am moving my tongue, or fiddling with my jewellery, or tapping my feet. But if I can make that discovery, then I cannot have known beforehand, since discovering that *p* at some time requires being ignorant of *p* until that time – "new knowledge replaces ignorance of the now-known" (O'Shaughnessy, 1980, p. 61). But if I can, by means of turning my attention to it, discover that I am fiddling with my jewellery, then I cannot have been aware that I was doing so. O'Shaughnessy concludes from this that the agent was unaware of their action *simpliciter*, "under any and all descriptions" (O'Shaughnessy, 1980, p. 61). Steward argues in the same way when she says that "[absentmindedly scratching one's head and fiddling with one's jewellery] often occur below the level of our conscious notice, and when they do, it seems impossible to characterise the events by means of which they are produced as ϕ-ings of any kind such that '*S* ϕ-ed intentionally' is true – not [...] without doing quite considerable violence to the concept of intention" (Steward, 2009, p. 298).

---

[86] In this section I sometimes talk about actions 'being intentional under some description', where this means that there is an action, ϕ, of which it is true that ϕ is *S*'s *A*-ing, and *S* practically knows that ϕ is her *A*-ing. I take it that this is roughly synonymous with '*S A*-ed/is *A*-ing intentionally'. The reason I have switched to talking in this way is that we are now discussing agents' awareness of their actions, and asking whether there are any actions which the agent has no awareness of at all. We are therefore talking about the *content* of a person's awareness of a *particular* (an action). Since content is a semantic notion, it makes sense to talk about the *description* under which an agent is aware of things; and since our focus is on actions rather than things done, the thing the agent is aware of as falling under the relevant description is an *action*.

[87] See O'Shaughnessy (1980, p. 62).

The first positive argument for the claim that Target Phenomena are actions is that when we become aware of a Target Phenomenon, we are willing to self-ascribe it in a way that we are not so willing with mere happenings in and around one's body. Steward insists that "[w]hen I fiddle with my jewellery, it seems to me, it is *me* who is fiddling with it, even if I am not aware that I am doing so" (Steward, 2009, p. 300). Now, if I trip and break a vase, it is me that breaks the vase, and I am often willing and able to put my hand up and admit it, thereby self-ascribing the vase-breaking. However, Steward thinks we should distinguish between these two self-ascriptions: "I am *active* in the fiddling though not in the tripping. [...] It is *me* who is moving my body in these cases" (Steward, 2009, p. 300). The first is an ascription of agency, the second of mere involvement in the causing of an event. Similarly, O'Shaughnessy says that "I become aware of *an activity of moving* that is performed by me" (O'Shaughnessy, 1980, p. 60).

The argument is that, if what I notice when I notice that I am moving my own body or fiddling with my jewellery is just that, *my* moving of my body and jewellery, then what I notice myself doing is *acting*. Implicit in the argument is the idea that it is sufficient for something's being an action that it is traceable back to an agent in this way, such that self-ascriptions of ownership and agency are appropriate. It certainly seems to be a part of the phenomenology of these cases that one can trace the behaviours back to oneself – they do not feel like a belly-rumble or an alien force taking over one's body – and so they are appropriately self-ascribable as one's moving one's own body. Therefore, they are actions.

The second positive argument is Steward's, and in it she attempts to show that the Target Phenomena are actions because they are exercises of a basic sort of bodily control. Steward says, for example, that "an action [is] an exercise of bodily control" where "for an animal to have control over its body is merely for it to be able [...] both to bring about some particular movement of its body, and to be able not to bring it about" (Steward, 2009, p. 308). And it seems plausible that the Target Phenomena are cases of an agent's control over their own body.

There are interesting questions about *why* we should think that Target Phenomena exhibit control, and about *what* control *is*. Steward has wrestled with both questions, but the issue has proved a thorny one.[88] However, I do not have space to discuss those problems here, and propose to simply grant that the Target Phenomena exhibit control, and therefore that they are actions. I agree with Steward that, however we spell out the notion of control, when I fiddle with my hair or shift my posture I am exercising control over myself.

In fact, I also propose to grant, though more cautiously, the first positive argument. The spirit of the argument seems to me correct, even though I have serious misgivings about the letter.

[88] Steward's first attempt was a purely modal condition on control, such that actions are exercises of control because, throughout their duration, the agent has the capacity to stop, alter, reverse, or change the trajectory of their body's motion (Steward, 2012a, p. 52). Douskos (2013) argued that this makes blinking and breathing, in the normal course of things, actions. Steward conceded this point, accepting that we need a way of thinking of actions as exercises of control from their onset (Steward, 2013, p. 698). In her more recent work, Steward (2017) attempts to articulate the idea that agency is a power of a whole organism to cause changes in its parts and the rest of the world, where this power is the result of a high-level of integration and co-ordination of the organism's parts and sub-systems. One might then try to develop a notion of control that appeals to this co-ordination, and then argue that the Target Phenomena exhibit that co-ordination.

The first misgiving is that there are delusions of control, where schizophrenic patients self-ascribe the movements of others they see acting (Martin, 2013). They seem to miss-self-ascribe on the basis of a feeling that it was them who controlled the other's body. This means that our willingness to claim ownership of things as actions of ours is not really sufficient for their *being* actions of ours. The second misgiving is just that it is unclear *why* these self-ascriptions of agency should matter to something's being an action. Neither O'Shaughnessy nor Steward gives a reason for thinking this. However, there seems to be something right about taking these self-ascriptions as a mark of agency in the normal case where I find myself fiddling with my jewellery. If asked 'Who's making that annoying rustling sound?' I would say 'Sorry, it was *me* – I was fiddling with my jewellery'. I think that we should take such cases at face value, even if the argument for that leaves something to the imagination.

Therefore, I am happy to accept that the Target Phenomena are actions, and that the reasons for this are expressed (if somewhat inchoately) in O'Shaughnessy and Steward's two positive arguments. This means that my argument against the claim that there are sub-intentional actions will not deny that the Target Phenomena are *actions*, but that they are *sub-intentional*. That is, I aim to put serious pressure on the idea that these actions do not meet the *Awareness Condition*, and on the negative argument which purports to show this conclusion.

3.2 – Rejecting the Negative Argument

I will now argue against the negative argument, which I think fails in at least three ways: firstly, I think the argument is either fallacious or embodies a less than definitive strategy; secondly, the view that some actions do not meet the *Awareness Condition* has extreme commitments; thirdly, the argument assumes that conscious attention is the kind of self-awareness relevant to an action's being intentional, but it is not. I will take these in turn.

The first problem is that, even if O'Shaughnessy and Steward are right that I can find myself scratching my head or tapping the table, this does not show that I was not aware of the action which was my doing of those things under any description at all. Perhaps I learn *that I am fiddling with my hair* when I turn my attention to my hand upon being asked 'What are you doing with your hand?'[89] Does that mean I was not aware *that I am moving my hand*, or *that I am touching my hair*, or *that I am fiddling with something*? No. It is perfectly consistent with knowing all of these facts that, when I turn my attention to my hand, I learn *that I am fiddling with my hair*. So we cannot infer from the fact that I can learn something about what I am doing by attending to my action to the claim that I was unaware of my action *simpliciter*, or under *any* description. If that is how the argument goes, it is fallacious. However, sticking to the letter of at least O'Shaughnessy's argument, this is just what we have (O'Shaughnessy, 1980, p. 61).

[89] I say 'perhaps', because in fact O'Shaughnessy gives us no reason in the text to think that we can make this discovery – he simply assumes that we can. Now, this is obviously a plausible assumption, but in this context it steers uncomfortably close to begging the question. But I will grant it for the sake of argument.

I do not want to accuse O'Shaughnessy and Steward of a baldly fallacious argument, though. It is clear that there is a more plausible argumentative strategy nascent in their discussions: isolate an action which is a Target because it plausibly does not meet the *Awareness Condition*, then show that it does not. But it is actually very difficult to see how this this strategy could be borne out successfully – it seems to me to have a structural problem. Examples of Target Phenomena mostly involve descriptions of an agent's action as being an *A*-ing, where the agent is unaware that they are *A*-ing, and where they may later learn that they are or were *A*-ing. But because the cases are described by isolating descriptions of the action which we already assume the agent lacks awareness of the action as falling under, the cases simply leave it open whether or not there are any *other* descriptions under which the agent *is* aware of the action.

To see the dialectical difficulty, consider an analogous argumentative strategy. First, isolate a case where *S* is contracting her muscles, and is not aware of doing so. Second, say that *S*'s contracting of her muscles is an action of hers, ϕ. Now, try and show that nothing *S* does, *A-N*, where ϕ is her doing of *A-N*, is something she does intentionally. This is a hopeless task – *S* may be doing all sorts of things intentionally. Isolating ϕ by means of a description of it under which *S* is not aware of ϕ doesn't help us at all in finding out whether ϕ is *S*'s doing *something* intentionally.

Perhaps this is where the notion of Target Phenomena plays more of a role. ϕ would only count as a Target Phenomenon if it were antecedently plausible that nothing *S* does, *A-N*, where ϕ is the doing of *A-N*, is intentional or figures in a description under which she is aware of ϕ. So we do not isolate ϕ with a single description like '*S* contracted her muscles' – that is not enough to make ϕ a Target. We must be able to isolate ϕ by means of its being plausible that the *majority* of what *S* does – *A-N* – where ϕ is the doing of *A-N*, is not in *S*'s awareness. So the muscle-contraction must be something like a tongue-movement or a posture-shift, where we already know that most of what one does when one does those things is not done intentionally and one is not aware of doing. A part of what guides our selection for membership in the class of Target Phenomena is therefore an antecedent appreciation of the sorts of activities which have nothing to do with our projects or desires, which we are not thinking about, and which we care little for. When an action is this far from our projects and cares, it becomes much more plausible that we are not aware of it because it plays so little a role in our psychology. Hence why Target Phenomena are mundane, insignificant bodily movements.

But even if we have identified ϕ like this, there is a gap between what we can be antecedently sure *S* is not aware of ϕ as and the descriptions under which *S may* be aware of ϕ. Even if we grant that *S* is not aware *that I am shifting posture* or *that I am contracting some muscles*, how can we be so sure that she is not aware *that I am moving*? *S*'s awareness of ϕ need only be so thin for it to meet the *Awareness Condition*.[90] The structural problem for O'Shaughnessy's and Steward's strategy is that there is *always* a gap between the ways we isolate Target actions and the thinnest possible descriptions under which an agent may be aware of the isolated action. But there is no way to bridge that gap.

[90] Ultimately, it was this consideration which drove O'Shaughnessy to reject the view that there are sub-intentional actions (O'Shaughnessy, 2008).

Perhaps O'Shaughnessy and Steward may stipulate a case where there is an action ϕ of agent *S*'s of which it is true *that ϕ is S's A-ing, B-ing… N-ing*, and that *S* is not aware of ϕ as being an *A-N*-ing. The main problem here is that it is doubtful whether we *can* just stipulate that *S* lacks all awareness of their action. Firstly, it looks awfully close to question-begging. I am disputing the claim that there are such cases, so how can one be stipulated? Secondly, sub-intentional actions are supposed to be ubiquitous features of agents' lives. If we cannot find them in the wild, this casts doubt on their reality.

Now I want to turn to my second response to the negative argument. I think that there are underappreciated, but extreme, commitments of the view that some actions do not meet the *Awareness Condition*. Say *S* is fiddling with her hair, and the action which is her fiddling is sub-intentional. Must we imagine that *S* has lost proprioceptive and kinaesthetic information about her body and its movements? It seems so, since her action is a moving of her body, and as such an awareness of this event should be provided by proprioceptive and kinaesthetic channels. The same goes for nail-bitings and toe-tappings. The only way for *S* to lack *all* awareness of these events is to lack the bodily senses, at least momentarily. But there is a problem with this: those who lack the body senses cannot easily move themselves – it takes a lot of effort and control after a long period of practice (Wong, 2015, 2018). But putative sub-intentional actions are not supposed to be the results of lots of effort and attention; far from it. Therefore, from the fact that *S* does easily fiddle with her hair, we can infer that she has her body senses intact. This means that she is proprioceptively and kinaesthetically aware of her action, so it meets the *Awareness Condition*. So it is in fact very hard to believe that any actions do not meet the *Awareness Condition*, simply because of the operation of proprioception and its connection to bodily action.

I now want to focus on my third and final problem for O'Shaughnessy and Steward. Both give a central role to conscious attention or conscious noticing. For example, Steward claims that Target actions "often occur below the level of *conscious notice*" and, when they do, characterising them as intentional does "considerable violence to the concept of intention" (Steward, 2009, p. 298). But if something's occurring below the level of conscious notice makes it inappropriate to call it intentional, then there is an implicit acceptance of the idea that *attentive awareness* of one's action is necessary for its being intentional. Therefore, although O'Shaughnessy and Steward aim to show that the Target Phenomena do not meet the *Awareness Condition*, they seem to actually be working with the more controversial *Attention Condition*:[91]

> *Attention Condition*: An action, ϕ, is intentional under some description only if ϕ figures in the agent's conscious attention.

The *Attention Condition* is extremely unattractive. Firstly, it is notoriously difficult to specify the content of conscious attention. If we think of conscious attention on the model of a spot-light, according to which something figures in conscious attention only if it is the object of focal attention,

[91] I do not claim that O'Shaughnessy and Steward endorse the *Attention Condition*, only that their argument implicitly presupposes it.

then a whole host of things that we intuitively do intentionally are straightforwardly excluded by the *Attention Condition*. For example, if I am walking through a park, deep in conversation with a friend, and my attention is directed at what I am saying rather than my walking, then the condition yields the claim that I am not intentionally walking. But that is unacceptable. On more plausible views which think of conscious attention as structuring the phenomenal (visual, auditory, etc.) field, things in the periphery of one's vision, say, figure in one's conscious attention even if they are not the focal objects (Arvidson, 2003, 2013; Watzl, 2017). But on that view, it would be hard to show that the Target Phenomena do not figure in conscious attention, since, even though they may not be the *focus* of attention, they are plausibly somewhere in the periphery, for example, in one's proprioceptive experience. And now we are back with the difficulty of showing that agents are not aware of moving their body, or some equally minimal thing they do in acting. So we cannot employ the notion of conscious attention in this context without facing the considerable difficulties of spelling it out and applying it.

The *Attention Condition* also excludes an important class of things done intentionally from being intentional: our acts of attending, themselves. I can, for example, intentionally look at, or attend to, a painting in a gallery. But if consciously attending to one's action is necessary for its being one's doing something intentionally, then to intentionally attend to a painting, I must attend to my attending. But given attention is a limited resource, it is not clear whether it is possible to attend to one's attending to a picture. Even if it is possible, I take it that we very rarely *do* attend to our own attendings. But it is not rare that we intentionally attend. On the suggested view, then, this disparity forces us to accept that we rarely intentionally attend to things. So when I just perceptually attend to a painting in the normal way – that is, not in the narcissistic way that I would if were to attend to my own attending – I never do so intentionally. But that is intolerable. Non-narcissistic intentional attendings must be admitted. If so, then the *Attention Condition* must be rejected.

O'Shaughnessy and Steward may avoid these issues by leaning on the fact that both employ the concept of *noticing* or *noticing that* something is true. Noticing is a more epistemic notion than conscious attention and so may not incur the worries outlined above. Noticing that *p* entails coming to learn that *p*, but this has little connection to phenomenal consciousness or attention. Then the negative argument can be placed in the key of noticing or noticing-that, rather than the key of attention. But this gives us implicit acceptance of:

> *Noticing Condition*: An action, ϕ, is intentional under some description only if the agent notices ϕ, or notices that *p*, where *p* is some fact about ϕ, knowledge of which would count as awareness of ϕ.

But even if O'Shaughnessy and Steward could show that some actions do not meet the *Noticing Condition*, this is not sufficient for showing that they do not meet the *Awareness Condition*: we are aware of plenty of things we don't notice. After all, I am aware of the colour of my pot plant, but do not notice it; typists are aware of the letters they type but do not notice them; when I intentionally raise my arm, I am aware of doing so but do not notice my action. Given

O'Shaughnessy's and Steward's strategy was to show that the Target Phenomena do not meet a very weak condition on intentional action, their argument fails since it operates with a stronger condition. Worse, not only is the *Noticing Condition* stronger than the *Awareness Condition*, it is also false, as the typing and arm-raising examples show.

I have been trying to cast doubt on the view that there are sub-intentional actions, and I have done so primarily by arguing that O'Shaughnessy's and Steward's negative argument for thinking the relevant actions are *sub*-intentional suffers significant problems. Interpreted uncharitably, it is fallacious. But interpreted charitably, it seems to suffer a serious structural problem: there is no way of identifying a Target action which then allows us to show that there is no description under which its agent is aware of it. The negative argument has a dialectical impasse built into it, and so cannot do the work it is being asked to do. Worse, I argued that when we consider proprioception, it seems that it is actually quite hard to see how one could fail to at least be proprioceptively aware of the action. Finally, the appeal to attention and noticing means that Steward's and O'Shaughnessy's argument fails to properly target the idea that actions fail to meet the *Awareness Condition*.

The conclusion of this is that we presently have little reason to think that there are sub-intentional actions, meaning that we should not accept Premise (1) in the argument from sub-intentional actions. Of course, I have not argued that it is *false* (although some of the things I have said suggest ways to do that). But blocking the acceptability of the premise is sufficient for defusing the power of the objection for my purposes.

Conclusion

In this chapter, I have canvassed three powerful sorts of arguments against the view defended in Chapter 4 that everything done habitually is done intentionally. The importance of that claim was that it took us from some other premises, defended in Chapter 2 and Chapter 3, to the Rationalising View of Habit. Therefore, the objections I have been considering are of vital importance, for they threatened to block my argument for the Rationalising View. However, I have tried to show that each of them fails to show that there are cases of doing something habitually but not intentionally.

In the case of slips, I argued we should not say that agents like Doris get to work, say, out of habit, even if their habit figures in an explanation of why they get there. Secondly, I argued that although there may seem to be some explanatory contexts where we need to ascribe habits of doing things where habit-bearers do them unintentionally, I have argued that this is not what those explanatory contexts require. Finally, I put pressure on the powerful idea that there are sub-intentional actions. Of course, much of this is controversial. But I hope to have made a strong enough case that these objections do not work. This, therefore, concludes my argument for the Rationalising View of Habit. I argued for the key premise connecting habit and intention in the previous chapter, and defended it against objections here.

In the next two chapters, I turn my attention the specific form of contribution to broad rationalisations that habits make. After all, the Rationalising View simply says that habits, when they manifest, figure in broad rationalisations. But since we have seen that pain, knowledge, fear and virtue all figure very differently in rationalisations, the Rationalising View leaves us in the dark about what sort of role habits have. How *precisely* do they bear on someone's having of, or acting for, reasons? The next two chapters develop an answer to this which has the habit-bearer's familiarity with their habitual course of action at its heart.

CHAPTER 6

MODELS OF THE RATIONALISING VIEW

I have argued for the Rationalising View of Habit. According to it, when habits manifest, they figure in broad rationalising explanations of why the habit-bearer acts as they do. Therefore, when I go to *Gilmour's Café* out of habit, I go there for reasons, and my habit figures in an explanation which makes it intelligible why I go there for those reasons. This means that there is a difference between the rationalising explanations of why I go there and why Alice goes there when she goes for the first time. Habits are 'rational difference-makers'.

Put like this, the view I have defended is incredibly general. In fact, if true, it tells us nothing about *how* habits figure in explanations of why habit-bearers do what they do. This is because there are so many different ways that different sorts of mental properties figure in broad rationalisations. Desires present courses of action as good, pain grounds reasons for action and provides epistemic access to those reasons. Belief and knowledge have different but related roles, and the emotions have generally very complicated relations to rationalisations. Sometimes they reveal independent reasons (as fear of a bear, or indignation at injustice do), and sometimes they ground them (as fear of harmless spiders does). If we accept the Rationalising View of Habit, what should we say of habit's rational role? Do they ground reasons? Do they provide us with sensitivities to independent reasons? Or do they have a distinctive rationalising role? The Rationalising View of Habit is not at all informative on this front.

There are different things we may want to say about habits, here. All of them will be versions of the Rationalising View since they are all ways of making good on the idea that habits figure in rationalising explanations. Therefore, I will call the variety of views 'models of the Rationalising View'. In this chapter, I want to consider and reject a number of models of habit's rational role. Some have been defended in the literature, and others are worth considering because they shed light on the similarities and differences between habits and other, related, mental properties like virtues and character traits. In arguing that these models fail, I draw out a number of constraints on any plausible model of the Rationalising View. The constraints, in the end, are very demanding. But I argue that they can be met by what I will call 'the Familiarity View'. The Familiarity View is the view that habits are tendencies to act for a reason which is idiosyncratic to a habit-bearer because it depends on their own history of action: the reason is *that A-ing in C is familiar to me*. Therefore,

when a person does something habitually, a tendency to do something for that reason is manifested. This is how habits figure in broad rationalisations. Further, I claim that (at least typically) the fact that one is familiar with doing what one usually does is revealed to one in experience by the feeling of familiarity.

In this chapter, my argument for the Familiarity View is accumulative. First, I argue against other possible models, deriving some constraints on a plausible theory. Then I show that the Familiarity View meets these constraints, and I argue that it is independently motivated. This is something like an argument by elimination. It is not decisive, but I do think it gives us good reason to believe the Familiarity View is on the right track. In the next, final, chapter, my argument is more direct. In it, I argue that the Familiarity View can help us explain important features of habit, in particular, by giving sense to the idea of 'the force of habit'. Together, I think they provide a very strong case for this particular model of habit's role in broad rationalisations.

This chapter is structured as follows. In Section 1, I consider and reject three models that either have been suggested or may suggest themselves. From my arguments, I derive five constraints of adequacy on models of the Rationalising View. At the end of the section, I motivate the thought that something like the Familiarity View may meet the constraints. In Section 2, I consider in detail the twin notions of being familiar and seeming familiar so that we have a good sense of what the central concepts employed in articulating the Familiarity View are. In Section 3, using these materials, I develop the Familiarity View and show that it is both independently plausible and meets the constraints. Finally, in Section 4, I address an objection drawn from Wittgenstein that the feeling of familiarity cannot play the role I give it.

Section 1 – Models of the Rationalising View

1.1 – Velleman's View

David Velleman's remarks in *Practical Reflection* (1989) suggest one way to model the Rationalising View of Habit. His remarks occur in the course of his defence of his very general views of agency. They are complicated, but it will be worth explaining them first, since they will help us understand what he says about habit.

Velleman defends a controversial and grand vision of human, reflective, rational agency. The cornerstone of his view is that, just as belief has the constitutive aim of truth, such that nothing counts as a belief unless it aims at truth, action also has a constitutive aim. Velleman's view is that the constitutive aim of action is self-knowledge: nothing counts as an action unless its aim is to improve one's knowledge of who one is and what one is doing. If that is right, then Velleman thinks we should orientate all our central agential notions around the axis of self-knowledge. To this end, he identifies intentions as beliefs about what one will do, which cause us to do what we believe we will, thereby maintaining and guaranteeing one's self-knowledge. He similarly analyses practical deliberation as theoretical deliberation about what one is most likely to do, so that reasoning towards

an intention is just reasoning towards a prediction about one's own behaviour. Finally, he thinks that reasons for action are derived from one's own nature, so that we only have reason to do what we believe we will do or which will produce the most self-knowledge. All of this is set against the background a Frankfurtian conception of agency as a matter of getting one's own desires, values, and actions to 'line up'. The further one is from perfect alignment, the further one is from genuine action.

In discussing how habits fit into these wider views, Velleman begins by saying that "[H]abitual action is usually accompanied by habitual knowledge. For example, I not only have a bedtime routine but also know that I have one" (Velleman, 1989, p. 70). Later on, after defending his view of reasons as things belief in which would allow someone to gain practical self-knowledge by acting, he says:

> "[I]f one knows that one has a habit of doing something, then one has a potential explanation for doing it, and so by doing the thing, one can take advantage of the potential self-understanding contained in one's knowledge of the habit. By my definition, then, having the habit of doing something qualifies as a reason for doing it. [...] My reason for wearing jeans today is that jeans are what I habitually wear to work." (Velleman, 1989, pp. 202–203)

It should be clear that Velleman is proposing a version of the Rationalising View of Habit. He thinks that habits can either simply *be* reasons, or *ground* facts which are reasons and which directly refer to habits (such as *that I habitually wear jeans*). Since I have already argued that reasons are facts, I'll consider Velleman's view in that form:

> *Velleman's View*: Habits figure in broad rationalising explanations of why someone does something habitually in virtue of the person taking the fact *that I have this habit* as a reason to do what they are in the habit of doing.

Can this be the correct way of modelling the Rationalising View? I am doubtful. For one thing, *Velleman's View* is only really intelligible if one already accepts Velleman's grand picture of action. After all, without that picture, it is hard to see how the fact that I have a habit of eating biscuits could speak in favour of eating biscuits. Now, plenty of things *do* stand in favour of biscuit-munching, but it is very hard to see how the fact about my *habit* can do so. That is, unless one thinks that one has reason to do whatever one knows one will do because action has the constitutive aim of self-knowledge. But that is a very controversial view with a potentially high-price.[92]

What this draws out is that a model of the Rationalising View of Habit must model habits as bearing on reasons for action which can actually seem to favour doing something from the agent's own perspective. Facts about our habits do not (unless we agree with Velleman) seem to favour

[92] See Bratman (1991), Hornsby (2008b), Langton (2004), and Setiya (2007) for a wide array of criticisms to Velleman's conception of agency.

doing what we are in the habit of doing. So a broad rationalisation where a habit manifests must imply a narrow rationalisation in terms of something which, from the agent's own point of view, actually seemed to favour doing what they are in the habit of doing. This is constraint (1):

1. The model must be able to make sense of things done habitually in terms of facts which, from the habit-bearer's perspective, actually *do* seem to speak in favour of acting;

There are other reasons to doubt *Velleman's View*, independently of its inextricability from his controversial meta-normative theory of reasons and agency. One is that it is very rare that, when people act out of habit, they take the fact that they have that habit as a reason to act. This is not my earlier claim that unless Velleman's meta-normative view is right then it would be unintelligible for them to do so; it is a simple psychological claim. When I bite my nails, I don't think of my habit as providing me with reason to do so. For one thing, many habit-bearers do not know what they are in the habit of doing. Unlike beliefs and experiences, habits do not confer any special first-person epistemic authority on the habit-bearer regarding knowledge of their habit. In fact, the situation seems somewhat reversed. Our families, friends and loved ones often know our habits much better than we do ourselves (Pollard, 2011). Therefore, it is psychologically implausible that everything done habitually is done for the reason that one has the habit, since many habit-bearers don't even *know* they have their habit. This gives us our second constraint:

2. The model must be psychologically and phenomenologically plausible;

Therefore, the correct model of the Rationalising View of Habit should not entail that everyone knows which habits they have, or that people think of their habits as giving them reasons to act, and so on. So *Velleman's View* is false.

1.2 – Virtues

One way of trying to model the Rationalising View of Habit is to look to mental properties which are related in some respect, and whose contributions to broad rationalisations is better understood. To this end, two sorts of properties stand out: virtues and character traits. In this section, I will assess whether we can model habits on virtues, and in the next (Section 1.3) I try modelling them on character traits. I will find both wanting,

Let me start with virtue. It is typically said by those following Aristotle that virtues are sorts of rational propensities which play a role in explaining why the virtuous agent does the right thing (Annas, 2011; Hampson, 2020; Hursthouse, 1988). But, as John McDowell puts it, "a kind person need not himself classify the behaviour he sees to be called for, on one of the relevant occasions, as kind" (McDowell, 1979, p. 51). Instead, the virtue of kindness plays a role in rationalising the agent's doing something kind because "[a] kind person has a reliable sensitivity to a certain sort of

requirement that situations impose on behaviour" (McDowell, 1979, p. 51). That is, a kind person is able, by having that virtue, to take features of their situation as morally salient and act accordingly. If they were not kind, or were horrid, they would not be reliably privy to those facts (or at least to their moral salience).[93] Therefore, a virtuous person knows something about the world's normative layout that a vicious person does not.

This is partly explained by the non-idiosyncrasy of (at least many) moral reasons. For a reason to be idiosyncratic, it must depend somehow on the features of the person for whom it is a reason. For example, if loving someone generates special obligations to them, then one has idiosyncratic reasons to do things for the beloved which others do not have (Phillips, 2018). However, most standard moral reasons which virtue theorists are interested in – reasons to do something kind or protect another – are not idiosyncratic: they bind people no matter who they are. This is important, because if moral facts were all idiosyncratic to virtuous people, we could not explain the epistemic failure of vicious people since there would *be* no reasons for a vicious person to be ignorant *of*, not themselves being virtuous. This part of the point of saying that virtues are sensitivities to 'objective' or 'independent' reasons.[94]

Aristotelians think that virtues like kindness are acquired by doing kind things, and thus they have often used talk of *habituation*. The thought is that virtuous sensitivity requires practice before one can grasp reasons in the right way, and this practice involves repetition of the right act-types in the right circumstances (in the right way, at the right time, with the right panache). As such, the plausible picture of the virtues I have described is also often coupled with a commitment to the view that repetition of an act-type in relevantly similar circumstances can endow the agent with the capacity to grasp morally relevant facts *as* the reasons they are.

The thought that there is a specifically *moral* sort of habituation which bootstraps in a sensitivity to reasons such that, when a virtuous agent responds to those reasons, we think that their virtue figures in a broad rationalising explanation of why they acted is already widely accepted (though I admit, not in those terms) and plausible. Therefore, the relation between virtue and habit may invite a comparison according to which we model the Rationalising View on the virtues. One could hold that having a habit, like having a virtue, provides one with access to reasons which are already there:

---

[93] If a non-virtuous person, or a learner on the way to becoming virtuous, were to somehow appreciate these facts' salience and act accordingly, the reasons in their possession would match the virtuous person's. But learning, of two people who had the same reasons and did the same thing, that one was virtuous and one not clearly makes a difference to how we understand the rationality of what they did. For one of them came to grasp their reasons by employing a long-standing rational sensitivity to what is required of them; the other had a moment of clarity amidst a moral fog. And this is a difference in the broad rationalising explanations of why they each do what they do, since rationalisation is a kind of explanation which appeals to how agents stand within the space of reasons, and (as I argued in Chapter 2) which capacities and propensities agents manifest in grasping reasons are relevant to that standing. That is the point of talking of an item's figuring in a broad rationalising explanation. For a very helpful discussion of related issues in Aristotle's ethics, see Hampson (2020).

[94] I will generally avoid use of 'objective reasons' and 'independent reasons'. Objectivity is a slippery notion and 'independent' invites the question 'independent from what?', which I wish to skirt for the moment.

Virtue Model: Habits are sensitivities to non-idiosyncratic reasons, and to their status *as* reasons, such that, when habit-bearers act out of habit, they act for those reasons.

This model seems to meet the two constraints from Section 1.1. Firstly, it does not seem to say anything psychologically or phenomenologically untoward. For example, just as a kind person need not think of themselves or their action as kind, the habit-bearer need not think of themselves as having a habit, or think of what they are doing as habitual, in order to act habitually. Secondly, since the habit provides habit-bearers with a sensitivity to reasons *as* reasons, the habit-bearer's doing the habitual thing is a candidate for rationality.

However, I think that the *Virtue Model* is not right. It gets two main things wrong. Firstly, the *Virtue Model* says that habits are *sensitivities*, which are kinds of capacity. Sensitivities are abilities to pick up information of a certain kind, or to differentially respond to different stimuli, whether they are colours and shapes in the case of perceptual sensitivities, or moral facts in the case of the virtues. But in Chapter One, I argued that habits are not capacities because capacities figure in explanations of how things are possible, whereas habits figure in explanations of why people do certain things. So identifying habits with sensitivities runs afoul of metaphysical considerations from Chapter One. And however one wants to model the Rationalising View, it must at least be consistent with habits' metaphysics, with what habits *are*.

This immediately suggests a third constraint. However, I think there is a further point. Whilst it is a minimal adequacy condition that a model be consistent with what habits are, it would also be attractive if the model could connect habit's rationalising role with its metaphysics. After all, the fact that habits are tendencies is part of what gives them their distinctive explanatory profile, and that profile is also what we are trying to model. Therefore, our third constraint should come in two parts:

3. The model should: (a) minimally, be consistent with the metaphysics of habit; (b) ideally, go some way towards explaining how the metaphysics of habits connects with their role in rationalising explanations;

Here is the second problem for the *Virtue Model*. Remember that we can think of a vicious person as being epistemically deficient with respect to the virtuous person because the reasons which virtues are sensitivities to are not idiosyncratic – they bind the non-virtuous just as much as they bind the virtuous. However, it does not seem that someone who lacks my habit of tapping the table after writing a sentence, or who does not habitually misspell 'phenomenology' is at an epistemic fault with respect to how things are, normatively speaking, independently of them. It's not that Alice, who has never been to *Gilmour's Cafe*, is ignorant of some facts which I am rightly aware of as

reasons to go there.[95] She is not epistemically culpable, and I am not epistemically excellent. Indeed, we can imagine that I have a habit of drinking coffee at *Gilmour's Café* but, for some reason, think it's a dump and the coffee's rubbish, and that Alice, going there for the first time, has testimonial knowledge that it is cosy and that the coffee is lovely. So, with respect to the independent normative layout, Alice is in a much better epistemic position than I am. On the *Virtue Model*, this possibility is excluded, since we would have to treat Alice as the epistemically deficient one. But that is the wrong result.

Therefore, the problem with the *Virtue Model* is that it treats habits as sensitivities to non-idiosyncratic reasons, reasons which could be reasons for anyone, regardless of their history of acting. Therefore, we must think of the reason for which a habit-bearer acts as being somewhat *idiosyncratic*. This would allow that someone without the features on which those reasons depend cannot be said to be at an epistemic fault for not knowing or acting on the habit-bearer's reason, since *they do not have such a reason*. I would like to put the fourth constraint which arises out of this discussion like this:

4. When a habit-bearer *A*-s out of habit, their *A*-ing is (at least partly) explained by an idiosyncratic reason, a fact whose status as a reason for the habit-bearer to *A* depends on some feature of the habit-bearer.[96]

1.3 – Character Traits

The idea that habits figure in broad rationalising explanations of why habit-bearers do things by bearing some relation to idiosyncratic reasons – reasons that only those habit-bearers have – is promising, yet it might seem also a little mysterious. How might propensities or tendencies such as habits be related to idiosyncratic reasons such that the rationalising role of habit is elucidated by appeal to such reasons?

In answer, we may turn to character traits, since we have actually already seen one way that character traits can be related to idiosyncratic reasons. Irascible people, such as Jane from Chapter 2, get angry very often, and when they are angry, they will be motivated to shout or throw things as angry people are wont to do. When Jane gets angry and tears a photograph of Joan in spiteful fury at having been mildly slighted by her, one reason she had for doing this is *that this is a photograph*

[95] Of course, Alice probably *will* be ignorant of *some* reasons to go there plausibly because she lacks first-hand knowledge of the place. The point is that it does not seem to be in virtue of Alice's ignorance of non-idiosyncratic reasons in general that there is a difference between the broad rationalisations of my going there and hers.

[96] I am aware that this requires some delicacy, since Humeans typically think that all reasons are idiosyncratic in this sense, since a subject's having a reason depends on the particular configuration of their motivational set (Shemmer, 2007; Williams, 1979). A Humean may worry, therefore, that (4) is redundant since there *are* only idiosyncratic reasons. I want to be conciliatory about this: if Humeanism is true, then (4) is redundant because it's trivially entailed by Humeanism; if non-Humeanism is true, then (4) is true and non-redundant, since non-Humeans need reminding that there are idiosyncratic reasons. This is sufficient warrant for stating (4) as a constraint, albeit one that Humeans think all rationalising explanations meet anyway.

of Joan. That is a reason to tear this photo and rather than another. But it is only a reason to tear the photograph given the fact that Jane is angry at Joan; otherwise the fact that this is a photo of Joan is not a reason to tear it. And Jane is angry, in this case, because of her irascible nature. Joan's slight against Jane was minor, and only someone as committed to fury as Jane would get upset at it. So, if Jane were not irascible, she wouldn't have gotten angry. Jane's possession of the fact *that this is a photograph of Joan* as a reason to tear it up is dependent on her being angry, which is in turn dependent on her irascibility. Therefore, although the fact would be the case regardless of Jane's anger, the fact is only a *reason* for Jane because of her anger, and therefore because of her irascibility. This is clearly a kind of idiosyncratic reason, and since the idiosyncratic feature on which it depends is a sort of propensity – a character trait – we can call it a *propensity-dependent reason*.

Character traits like irascibility, cheerfulness and anxiousness, as it were, 'turn certain facts into reasons for action' by making one angry, happy, or anxious, which thereby gives certain facts about one's situation the status of reasons. This was part of my point when I discussed the shy student, Jill, in Chapter 2. Jill's shyness polarizes situations for her, making some otherwise fairly innocuous activities (such as answering questions) and seating places (such as the ones at the front) repellent. Instead, her shyness makes the back of the room seem attractive. Certain facts become reasons for Jill in virtue of the affective salience they have for her as someone who feels shy, where Jill feels shy in certain circumstances partly because of her *being a shy person*.[97]

The connection between habits and character traits, like that between habits and virtues, seems to run quite deeply. Therefore, we might try to explain the idiosyncrasy of a habit-bearer's reasons in the same terms:

> *Character Trait Model*: Habits are propensities which turn some independently existing facts into idiosyncratic reasons for action such that, when one acts out of habit, one acts for those reasons.

This does not seem to suffer the epistemic problem that the *Virtue Model* suffered, since a person has the relevant reason to do what they are in the habit of doing only if they have that habit. An onlooker without the habit would not be epistemically deficient for not knowing *they themselves* have a reason to do something – they don't have that reason.

However, the problem with the *Character Trait Model* as a model of the Rationalising View of Habit is that it is unclear which already existing facts could be 'made into reasons' by an agent's having a habit. Think again of my habitually drinking coffee at *Gilmour's Café* and Alice's first trip. This model amounts to claiming that there are some facts which we can both be aware of, which are not reasons for Alice to drink coffee there, but which, since I have my habit, are reasons for me to go there *and* which I am aware of *as reasons*. But it is hard to see which facts they could possibly be. Both Alice and I take facts about the nice coffee, atmosphere, comfort, convenience and hospitality

[97] Of course, we can be anxious on an occasion without being an anxious sort of person, and we can be angry without being irascible. In those cases, we still have these reasons. They just aren't idiosyncratic in the same *way* as the reasons of anxiety-ridden and irascible people.

as reasons to go there. Which facts that are not reasons for Alice to go there become reasons for me to do so? Does my habit mean I start to take facts which seem irrelevant to Alice as reasons to drink there? Am I suddenly attracted by the colour of the parquet flooring, by the number of bricks in the wall, or the hair colour of the baristas? None of this seems at all plausible. And it is hard to see how any of these facts have anything to do with my habit; *why* would it be the case that a habit turns already existing facts into reasons for me to do things?

This suggests a fifth constraint on a plausible model of the Rationalising View of Habit:

5. The fact which is one's idiosyncratic reason is a fact whose truth depends on features of the habit-bearer.[98]

The idea is simply that these are facts such as *that I really want to have it*; or *that I love her*; or *that Laura knows how to do that*; or *that Jane has done this before*. We can cast such facts in indexical or non-indexical form, but what matters is that they are facts about a particular person whose truth depends on how things are with them. According to this constraint, then, Alice does not have my reason to go to *Gilmour's Café* partly because my reason is a reason for me and not for her (constraint 4), and partly because my reason is a fact which is true of me but not true of her (constraint 5).

1.4 – Five Demanding Constraints

We now have five constraints on a plausible model of the Rationalising View of Habit:

1. The model must be able to make sense of things done habitually in terms of facts which, from the habit-bearer's perspective, actually *do* seem to speak in favour of acting;
2. The model must be psychologically and phenomenologically plausible;
3. The model should: (a) minimally, be consistent with the metaphysics of habit; (b) ideally, go some way towards explaining how the metaphysics of habits connects with their role in rationalising explanations;
4. When a habit-bearer *A*-s out of habit, their *A*-ing is (at least partly) explained by an idiosyncratic reason, a fact whose status as a reason for the habit-bearer to *A* depends on some feature of the habit-bearer;
5. The fact which is one's idiosyncratic reason for *A*-ing, when one *A*-s out of habit, is a fact whose truth depends on features of the habit-bearer.

These constraints are actually quite demanding. It is not so easy to see what model might satisfy all of them. What we are looking for is a plausible candidate reason for doing what one is in the habit of doing, whose truth and status as a reason somehow depends on the habit-bearer. But

[98] One might therefore call it an 'idiosyncratic fact', but I see no reason for the terminology.

what could that be? Which propositions are made true by features of a person which might be relevant to their habit? And which of those facts could possibly be a reason for action? These are difficult questions, and it can be hard to move beyond facts such as *that this is just what I do*. But this Wittgensteinianism is no comfort for a defender of the Rationalising View, since it is notoriously difficult to draw reason from regularity.

However, one fact strikes me as distinctively suited to play the required role: *that doing the habitual is familiar*. For one thing, when one forms a habit of going to *Gilmour's Café*, for example, one also becomes familiar with doing so. Alice is not familiar with it because she has never been before. However, I am, so *that going to the Gilmour's Café is familiar to me* depends on my history of action and repetition. So it seems to respect (5). And, intuitively, the fact that doing something is familiar gives doing it some appeal – the familiar is importantly linked to safety, normalcy, ease, and comfort, and doing the unfamiliar can feel strange and unhomely. So it seems to respect (1) and (4). Doing something because it is familiar does not entail knowing that one has a habit, and, on the face of it, familiarity might figure in a plausible psychological and phenomenological story about habit. So it meets (2) also. That leaves only (4) not intuitively motivated. However, since there are appealing connections between familiarity and habit, it is an option worth exploring.

This is my gambit, at any rate. In the next section, I will try to say more about familiarity and the feeling of familiarity, so that we can get a better grip on what it is I am invoking. Then, in Section 3, I will show how, given the nature of familiarity, there is an independent motivation for what I called earlier 'the Familiarity View'. I will then show that the view meets the five constraints. This does not amount to a definitive argument for the Familiarity View as a model of the Rationalising View, but I think it does make the case quite strong. Finally, I will consider an objection to the view in Section 4.

Section 2 – Being and Seeming Familiar

In discussing familiarity, we must first make a crucial distinction. The distinction is between something's *being familiar* and something's *seeming familiar*. Something *is* familiar only if one has engaged in or with it before. A friend's face is familiar because I have seen it so often, whereas a strangers is not because it is new to me; London is familiar because I have lived there for many years, whereas São Paulo is totally unfamiliar to me. That is not to say that engagement with *x* is sufficient for familiarity with *x* – things may be much more complicated than that. But it is at least true that engaging with things is *how* one becomes familiar with them.

Familiarity also seems to be gradable: I am more familiar with playing jazz than funk, but more familiar playing blues than jazz; I am more familiar with London than Athens, but am somewhat familiar with Athens since I have spent some time there. So the degree to which one is familiar with something depends on the degree to which one has engaged with it before. But one can engage with things in different ways: someone can become somewhat familiar with Naples by reading Elena Ferrente's wonderful novels (2015), but their familiarity will be both far less extensive

than a Napolese person's and have a different source.[99] But, for most things we are familiar with, there seem to be proprietary ways of becoming familiar with them. Familiarity with a city comes from walking around it and eating in its diners; familiarity with a dance comes from dancing it; familiarity with a person comes from talking with them.

We can also be familiar with both particular objects and with types of thing. So, whilst a stranger's face is unfamiliar to me, it is familiar to me *as a face*. That is, I am familiar with that type of thing, even though I am not familiar with that particular thing. In contrast, I am familiar with my friend, an individual, and with a variety of their properties which others share. The distinction between being familiar with an individual and being familiar with a type requires careful tracking when we talk about what we are familiar with. Being familiar with London seems a case of familiarity with an individual (of a sort); being familiar with playing the blues is familiarity with a type of activity, or a thing one can do, which is not an individual.

That is the first side of the crucial distinction. On the other side, things can also *seem*, *look*, or *feel* familiar. The sense that something is familiar – what I will often call 'the feeling of familiarity' – is notably connected to our perceptual recognitional capacities. We can get a better idea of the feeling by first thinking about cases where it is especially prominent. For example, in the street I might notice someone smiling and waving at me without immediately recognising them as a close friend. Seconds later it might dawn on me who it is. Matthew Ratcliffe puts it like this: "[T]he whole experiential structure changes and takes on an air of familiarity as the face's significance is registered; 'It's him!' Without that sudden reorientation, perhaps he would remain unfamiliar, unrecognised" (Ratcliffe, 2004, p. 39). These cases where the feeling of familiarity washes over one are not especially rare. We can have these experiences when we return home after a long time away, or when one hears a song on the radio which one *knows* but can't place. These cases where the feeling of familiarity washes over one are useful ways of isolating the phenomenon. However, the experience is much more omnipresent than even these common experiences. Mostly, the feeling of familiarity does not wash over one, attracting significant focal attention. It usually sits in the background of experience making that to which we are repeatedly exposed seem normal, and making things which deviate from the norm come to seem alien or strange.

This suggests that the feeling of familiarity has an affective profile. When we feel something as familiar it typically feels normal, homely, comfortable, and the unfamiliar can feel alien and strange. This fact is brought out nicely (if melodramatically) by Marcel Proust, whose character, the child, Marcel, had just gotten used to his new room, when:

> "Someone had [...] the happy idea of giving me, to distract me on evenings when I seemed abnormally wretched, a magic lantern, which used to be set on top of my lamp while we waited for dinner-time to come; and, after the fashion of the master-builders and glass-painters of Gothic days, it substituted for the opaqueness of my walls an impalpable iridescence, supernatural phenomena of many colours, in

[99] This suggests that being familiar with things admits of distinctions somewhat like Russell's (1911) distinction between knowledge by description and knowledge by acquaintance.

> which legends were depicted as on a shifting and transitory window. But my sorrows were only increased thereby, because this mere change of lighting was enough to destroy the familiar impression I had of my room, thanks to which, save for the torture of going to bed, it had become quite endurable. Now I no longer recognised it, and felt uneasy in it, as in a room in some hotel or chalet, in a place where I had just arrived by train for the first time." (Proust, 1984, pp. 10–11)

One of the things this shows is that when we have become familiar with a room, say, we form a sort of emotional attachment to it which can be destroyed by even slight changes. Places can start to feel alien, where before they were felt to be places of safety.

This is not to say that the feeling of familiarity, contrasted with the feeling of unfamiliarity, is always positively valenced. The familiar can feel stale, dull, and monotonous; the unfamiliar can feel fresh and exciting. Here is Yuriko Saito:

> "Because we take most things for granted in our everyday dealing with them, thus paying very little attention, wearing an artistic lens often renders the familiar things strange, and we experience them as if we have never experienced them before. Such experiences are refreshing, enlightening, and exciting." (Saito, 2017, p. 17)

To say that the feeling of familiarity has an affective profile, then, is not to say that it typically has a positive valence, or that it feels good. Sometimes feeling a place or activity to be familiar has positive features; sometimes unfamiliarity does too. It strikes me that, when we think about the natural ways of explicating the affective character of experiential familiarity – comfort, monotony, homeliness – there is no simple mapping of these descriptions onto positive and negative valences. Instead, it seems that properties like comfort, homeliness, and monotony are sorts of evaluative properties, and that the best way of capturing the idea that the feeling of familiarity has an affective profile is just in terms of the sorts of evaluative properties it can serve to reveal, imply, and be implied by in different sorts of situations.[100] Exactly how to capture this is a difficult question, and one that I will not deal with here.

I have called the experiential phenomenon I have isolated 'the *feeling* of familiarity', and this may make one think that I must have in mind something like an identifiable *sensation* akin to a tickle or an itch. For my part, the terminology is not intended to indicate anything other than that the phenomenon is *experiential*. I am as happy to talk about the experience of familiarity, things looking or sounding familiar, or the sense that something is familiar.

However, some do think of the feeling of familiarity on the model of sensations. For example, it is common to describe the feeling of familiarity as being "a warm glow" (Dokic, 2010, p. 41).[101] But

[100] This way of thinking about things has been influenced by Johnston (2001b), Poellner (2016), and a number of very helpful conversations with Komarine Romdenh-Romluc who has worked hard to get me to see the possibility of this sort of view.

[101] Talk of 'the warm glow of familiarity' is perhaps most common in psychology (Corneille et al., 2005; De Vries et al., 2010; Monin, 2003). Apparently, it dates back to Titchener (1918).

I do not think this is useful. Taken literally, it is false: people and places, just in looking familiar, do not emanate a special sort of light; and we do not, ourselves, go warm whenever we recognise things. The familiar does not glow, and the feeling is not like blushing. Taken less literally, talk of a warm glow seems mostly to indicate that the experience is affective. However, not all affective experiences are sensations.

In fact, the sensation model of the feeling of familiarity seems very implausible on reflection. One reason is that, as Bruce Mangan points out, "while we know without question *that* we feel it, just *how* it feels is a good deal more obscure" (Mangan, 2001, p. 3). This does not seem to be the case with sensations like tickles and pains. They are actually quite easy to locate in one's mental life and describe (for example, 'it feels like a feather brushing lightly down my ribcage'; 'it hurts sharply in a central location on my thigh, and then the pain emanates from there through my leg, becoming duller farther away from the sharp pain'). In contrast, the feeling of familiarity is diffuse, hard to describe, and hard to locate in one's stream of consciousness, or in any place on the body or in the world. If I see a friend walking down my street and recognise them as familiar, exactly *where* in my visual field is the familiarity located, and exactly where do I feel it? If these questions seem impossible to answer, I think the reason is that they don't *have* answers. My friend looks familiar, and that is a feature of my experience of him. But there is no place where the feeling occurs and no place where the familiarity is presented as being. This stands in stark contrast to the thought that the experience is anything like a sensation.

Instead, I think that the experience of familiarity is somewhat like 'aspect perception'. Aspect perception occurs, roughly, when one perceives an object *as* being of a certain kind. One perceives it under a certain aspect or guise. Put this broadly, plausibly all perception is aspectual. Even if one stands so close to a white wall that things look homogenously white, one sees the wall as being white. But what is interesting in this context is that the aspect that one perceives things as being under is not tied-down to the things one actually perceives: aspectual seeing allows for Gestalt switches. For example, there is the famous line drawing that can be seen either as a duck or a rabbit, and the Necker Cube which can be seen as either facing one way of the other. In all these cases perceivers can switch from seeing the very same item as being one way and then another. What is seen – the actual drawing, or object – stays the same.

The first reason the feeling of familiarity is similar to aspectual perception is that it allows for cases like Gestalt switches. That is what Ratcliffe's case, above, shows. There is no difference in what is seen, when one comes to recognise one's friend. But there is a difference in how the friend now looks. The second reason is that experiencing something as being familiar brings the item under a specific guise – the aspect of *familiarity*.

The third reason is more involved. Although when one perceives something under an aspect the aspect contributes to the phenomenal character of the perceptual experience, it does not contribute in the same way as the items in the presented scene do. When I see a rabbit, the rabbit itself contributes to my experience by being the thing presented, in all its determinate fluffiness. It is the thing I see. But I may mistakenly see the rabbit as a curled up cat. After suffering the illusion a while, I may come to understand that I am seeing it wrongly, and undergo an aspectual shift toward

seeing it as a rabbit.[102] Therefore, what is actually seen does not fully determine the nature of the experience. The aspect under which things are presented (or represented) therefore shapes how the rabbit, presented in perception, looks to me. But it cannot contribute in the same way as the rabbit, since then it would have to be one of the things seen. But if the aspect were itself perceived, we could ask under what aspect *it* appears to me.[103] But that would lead to a regress. Therefore, the aspect under which I see the rabbit cannot contribute in the same way as the rabbit does.

A suggestive way of putting this is that the rabbit's contribution is being the presented element, the matter of perception. The aspect, then, is a certain sort of form which organises what and how the presented element is taken to be. The aspect is not, then, presented in perception. There is a dual character, therefore, in the way that aspectual perceptual experiences get their phenomenal character: their character is mutually determined both by the nature of the thing perceived and the nature of the aspect or guise under which it is perceived. If the contribution of the rabbit is presentational, call the contribution of the aspect 'non-presentational'.

The feeling of familiarity, or things seeming or looking familiar, has the same dual character. On the one hand, there is the city, or the face which is perceived and contributes presentationally to the phenomenal character of the perceptual experience. On the other hand, when they look familiar, the familiarity does not contribute presentationally. It is not among the manifold of perceived items, contributing the matter of perception. Instead, it is a sort of form, or aspect, which shapes how the perceived things seem to one. When things look or feel familiar, that feature of the phenomenal character of one's experience is determined non-presentationally by bringing the perceived thing under the aspect of familiarity.

However, I want to stop short of claiming that the feeling of familiarity, or something's looking or sounding familiar, *is* a case of aspect perception. There are two reasons for this. For one thing, more work would have to be done to show that something's being familiar can really enter into the content of perception.[104] For another thing, things can be felt to be familiar even when they are not perceived. Ideas can feel familiar, for example, when you are reminded of something you used to think, or when you realise that something you currently think is reminiscent of something you recently read. Examples of an 'intellectual' feeling of familiarity are not cases of *perceiving* something under the aspect of the familiar, and so the feeling of familiarity is not limited to perception. Nonetheless, the intellectual cases share something of the structure I described in the perceptual case. One has the thought that *p* so that one's thinking about this proposition occupies

[102] There is a fantastic discussion of corrigible illusions in Siewert (2005).

[103] This is why the term 'aspect perception' is misleading. It makes it sound as if, amongst the things we perceive, additionally to rabbits and tables, are *aspects* of rabbits and tables. Since the notion of an aspect is not that of a part, or even a property, it is not at all plausible that we ever perceive aspects. Instead, we perceive objects *as falling under* aspects, guises, or concepts. I believe Wittgenstein encouraged the mistake of thinking we see aspects when he said: "Two uses of the word 'see'. The one: 'What do you see there?' — 'I see this' (and then a description, a drawing, a copy). The other: 'I see a likeness in these two faces' — let the man to whom I tell this be seeing the faces as clearly as I do myself. What is important is the categorial difference between the two 'objects' of sight" (Wittgenstein, 2009, p. 203).

[104] Charles Siewert (1998) has argued it does, and Susanna Siegel's (2010) 'rich content view' of perception suggests it may. However, see Dokic (2010), Lyons (2005), Matthen (2010) for very plausible arguments against this.

some space in one's stream of consciousness. Then, one comes to feel that the thought that *p* is familiar, bringing it under the guise of familiarity, thereby altering one's stream of consciousness.

This means that the analogy with aspectual perception holds up and elucidates the nature of the experience even if we stop short of claiming that the feeling of familiarity is a kind of perceptual experience. My overall point is that the feeling of familiarity need not be construed on the model of a sensation, and that it is better construed as being *at least like* aspect perception in some important ways that I have tried to bring out.

It is important to see how being familiar with something is connected to feeling familiar with it. Most obviously, in good cases, the feeling of familiarity plausibly reveals that which is familiar *as being* familiar. This is brought out in the quote from Ratcliffe where says that "the face's significance is *registered*", where 'registration' is most naturally understood as being factive (Ratcliffe, 2004, p. 39). Indeed, I think there is reason to think that the feeling of familiarity is similar to perceptual experience in this respect. Where perceptual experience, in a good case, makes how things are manifest to a subject, the feeling of familiarity, in a good case, may make it manifest to a person which things they are familiar with.

However, something's being familiar and its feeling familiar can come apart. When I experience déjà vu the scene feels familiar but is not. Cases where I see a friend but fail to recognise her can be described as cases where she does not seem familiar, but actually is. So the feeling of familiarity admits of illusory cases. However, I take it that this does not falsify the claim that, in good cases, when something is familiar and seems familiar, my experience simply discloses how it is, and does so under the appropriate aspect. The thing is experienced as being the way it is: familiar.

Further, things typically seem familiar *because* they are, just as oranges typically look to be spherical because they are. Although there are perceptual illusions, cases where things are not what they seem, this does nothing to cast doubt over whether in good cases things look how they do to a perceiver because they are that way.[105] The same thing can be said of the feeling of familiarity.

Section 3 – Habit and Familiarity

I have been examining the twin notions of familiarity and the feeling of familiarity in very broad terms. I have argued that the feeling of familiarity should not be thought of as a sensation, and is importantly similar to the phenomenon of aspectual perception. I also argued that the feeling of familiarity has an affective profile, and serves to reveal the familiar in good cases. I also argued that one becomes familiar with things by engaging with them in proprietary sorts of ways.

But although I have been talking about familiarity and the experience of familiarity in very general terms, I think their connection to habit should be quite transparent. We are not only familiar with faces, languages, practices, and rooms. We are also familiar with the things that we repeatedly do, such as brushing our teeth, going for morning runs, or visiting *Gilmour's* for coffee. We are also

[105] At least, it shouldn't do. See Brewer (2008), Dancy (1995), McDowell (1983, 1986), and Martin (2004, 2006) for critical assessments of arguments from illusion.

familiar, not just with doing them, but with doing them in the contexts in which we typically do them: in the context of going to bed; of a Saturday morning, or a Monday lunchtime.[106]

To form a habit of *A*-ing in *C*, one must have repeatedly *A*-ed in *C* before. This means that one will be familiar with three things: with *A*-ing; with *C*; and with *A*-ing in *C*. Therefore, my habit of going to the *Gilmour's Café* for a lunchtime coffee means that I am familiar with getting coffee there, with the general lunchtime-coffee-getting context, and with the fact that I get coffee at the *Gilmour's Café* in that context. This means that having my habit, and being familiar with what I am in the habit of doing, are both determined by the same factor: my repeatedly doing something in a given context-type, *C*. Therefore, one is familiar with everything done habitually, and with one's doing those things in the relevant contexts. Given this, and the fact that the feeling of familiarity serves to reveal the familiar, when in an instance of *C*, a person in the habit of *A*-ing in *C* will (in a good case) feel that *A*-ing is the familiar thing to do. Having doodled much in past lectures, for example, I am familiar with doing so, and this (in good cases) explains why doodling in a lecture feels familiar to me.[107]

So there are tight connections between familiarity, the feeling of familiarity, repetition of act-types in context-types, and habit. My suggestion is that we can exploit these connections in developing a model of the Rationalising View of Habit which meets all five constraints discussed in Section 1. I will develop this in Section 3.2. But, first, there is one final piece of the picture that I need to present: I want to argue for the claim that *that A-ing in C is familiar to S* is a reason for *S* to *A* when in an instance of *C*.

3.1 – Familiarity Grounds Reasons for Action

I will argue that the fact *that A-ing in C is familiar to S* is a reason for *S* to *A* when in an instance of *C*. Now, for philosophers such as Mark Johnston (2001b) and Peter Poellner (2016), I have already said enough to show this claim to be true. I have described the feeling of familiarity as a kind of affective experience, and those philosophers think that affective experiences should be understood across the board as revelations of values which exert motivational force on our behaviour. If one thinks this, then one has an easy route to arguing that the fact *that A-ing in C is*

[106] The phrase '*S* is familiar with *A*-ing in *C*' is ambiguous between a prospective and progressive reading. The prospective reading is '*S* is familiar with being such as to *A* when in *C*'. The progressive reading is '*S* is familiar with being in the process of *A*-ing when in *C*'. These readings carry over to the feeling of familiarity. On the one hand, there is the feeling, in an instance of *C*, that *A*-ing now would be familiar, where one has not started *A*-ing. On the other, there is the feeling, when *A*-ing in an instance of *C*, that doing *this* is familiar. I am primarily interested in the explanations of why habit-bearer's act as they do, when they act out of habit, so my use of '*S* is familiar with *A*-ing in *C*' is primarily prospective. This is because, when the progressive reading is available, *S* is already *A*-ing. But we want to know *why S A*-ed, so the explanation cannot be found in the fact that *S* is and feels familiar with *doing the thing we want to explain*. That would be puzzlingly circular. To explain why *S A*-s in terms of *S*'s familiarity with *A*-ing in *C*, we need the prospective reading of '*S* is familiar with *A*-ing in *C*'. This is not to say that the progressive reading plays no role – perhaps it plays a role in keeping *S* on track. But I will not pursue that question here. Thanks to Jingbo Hu for pressing me on this.

[107] For ease of expression, I will proceed without the 'good cases' qualification.

familiar to S is a reason to act: the feeling of familiarity is affective, and therefore what it reveals – the familiarity of things – is a *value* which exerts motivational force on *S*. The feeling of familiarity presents us with reasons for action, and we arrive at this via a conception of affective experiences generally as revelations of value.

I am quite sympathetic to this view, and if the view about affective experiences can be successfully defended, then I would be happy rest my case with this argument. However, as it stands it is deeply controversial and I have little original to say in its defence.[108] Despite this, it is worth noting if only because a plausible view of affective experiences would provide an argument for my view. Instead, I intend to give a much simpler argument for my conclusion, one which does not rely on, or entail, the wholly general thesis about affective experience that Johnston and Poellner defend. My argument is based on the specific character of things we are familiar with doing.

Clare Carlisle (2006, 2014) has best described the connection between being familiar with doing something and having reasons for doing it.[109] As she says, there is a "sense of comfort, safety and ease that is engendered by familiarity" which contributes to "[insulating] us from the threat of the unknown" (Carlisle, 2006, p. 23). Carlisle argues that this is why "even during a week away one finds a regular haunt: the café one returns to each morning [...]. In combining the novel with the familiar they [...] make one feel at home in a new place" (Carlisle, 2014, p. 78). The point is that there is a kind of safety and ease in doing what we always do – the familiar paths through places we often find ourselves in are well-trodden, and, in contrast with courses of action that we are unfamiliar with, are vouched for by one's own history. Take one of Carlisle's examples. When I am away on a trip, because I go to the same café each morning I know what it's like there and what to expect. There are other cafes, and I can see that they seem nice. But they are places that I could be surprised by, and so represent something of a risky alternative to my consistent haunt. Therefore, my familiarity with my usual place grounds a reason to go there – the reason is *that it is familiar to me*.

One may be sceptical of this given I have cashed-out the normative significance of familiarity in terms of other notions like safety, risk, comfort, and ease. It may seem like these provide the real reasons for action. I have two responses. The first is that facts about risk and safety may explain *why* the fact about familiarity is a reason, without usurping its status as a reason. Compare: all sorts of things will explain why the fact *that my friend is in need* is a reason to help her. But that is consistent with the latter fact's being a reason. My second response is that one may, of course, do something for the reason that it's less risky than an alternative. But that is a different case from one in which someone does something for the reason that it's familiar. Even if the two reasons are connected in important ways, I am trying to capture how one of them – the fact about familiarity – sometimes plays a motivating role for agents. Whilst I want to explain how this fact can get its normative force in terms of safety and risk, the cases I am highlighting are cases where people take the fact that

[108] See Dokic and Lemaire (2013), Döring (2007), Johnston (2001a), Mitchell (2017), Vanello (2018), and Wedgwood (2001) for this debate.

[109] Carlisle seems to run being familiar and feeling familiar together, but I intend to use her claims to make a point about the former. Amelie Rorty also nicely describes familiarity's reason-grounding force at (Rorty, 1980, p. 210).

something is familiar to them as a reason to do it because, from their perspective, it seems to favour doing so.

Of course, this says nothing about the independent relative values of these places. Perhaps my usual café is actually much less cosy and has much worse coffee than the others. So, my reason for going there which is grounded in my familiarity with it is normatively outweighed by reasons to go somewhere else, even if it is the motivationally potent reason. But we should not be surprised that habits can sometimes obscure stronger normative reasons for doing something different. Habits are basically conservative forces, in that they keep us in the realm of the known, thereby keeping us from situations which, though they may be fun and new, may equally be dangerous, difficult or unpleasant. My go-to café represents the conservative, safe, option. And that is part of what explains how the fact that it is familiar gets to be one of my reasons for going there.

I want to resist the temptation of thinking that one's reason to do the familiar must always be that doing the familiar *feels* good because the feeling of familiarity has a positive affective profile. The fact that something feels good often can be a person's reason for doing something, but typically only if the experience is positively valenced – that eating ice cream will be tasty is a reason to eat it; that eating dirt will taste awful will almost never be a reason to eat it.[110] But I have allowed that the affective value of the feeling of familiarity may be often negative – the familiar may feel stale and boring. So Carlisle's talk of the familiar as 'making one feel at home' (where this is understood as a positively valenced, homely sort of feeling), by cashing out the reason in terms of how doing the familiar makes one *feel*, does not get to the heart of things. One has reason to do the familiar even if it will not make one feel any better. For example, someone in an unfulfilling relationship may feel that their life has become fusty and lacklustre, and that this is most prominent from their perspective in the unremitting familiarity of it all. Still, I think they have *some* reason to stay in the relationship: it's familiar because it is predictable, a safe bet, despite being emotionally worthless.

3.2 – The Familiarity Model of the Rationalising Conception

I am now in a position to argue that there is a model of the Rationalising Model of Habit which meets the constraints from Section 1, and which has a habit-bearer's familiarity with what they do out of habit at its core. The model is the Familiarity View:

> *Familiarity View*: Habits are tendencies to *A* in *C* for the reason that *A-ing in C is familiar to the habit-bearer*, where that reason is typically revealed to the habit-bearer in feeling familiar with *A*-ing in *C*.

[110] There are occasions on which something's tasting bad is a reason to eat it. For example, if you're fascinated by someone's bad cooking then one reason to taste it is to see *just how bad it really is*. But these examples are somewhat removed from my topic.

To make the case for this model, I will examine how it meets each of the five constraints plausibly and elegantly. Then, in the next section, I will respond to a serious objection to the view. If the Familiarity View can meet both the constraints and the objection, as well as being independently motivated, then I will conclude that it serves as the best contender for the correct model of the Rationalising View of Habit. I will run through each constraint in order.

1. The model must be able to make sense of things done habitually in terms of facts which, from the habit-bearer's perspective, actually *do* seem to speak in favour of acting;

The Familiarity View meets this constraint because I argued that the fact *that A-ing in C is familiar to S* genuinely does speak in favour of *A*-ing for *S*, and so is a reason to *A*. This claim depended on plausible claims about the nature of familiarity, rather than on a controversial meta-normative theory of reasons. As well, I argued that the feeling familiarity which one has when (and because) something is familiar reveals that it is familiar. Therefore, it reveals a reason for acting, and the habit-bearer acts for that reason. The model does not say that the habit-bearer's reason for acting must be something which they could not plausibly see as favouring acting, as *Vellemen's View* did. So my view meets the first constraint.

2. The model must be psychologically and phenomenologically plausible;

At first glance, the Familiarity View seems not to commit to anything very implausible in this regard. I have argued that the feeling of familiarity is an experiential feature of one's psychology which seems perfectly at home in a discussion of habitual action. I have not said that habit-bearers must take the fact *that A-ing in C is familiar to do* as the first premise in a chain of reasoning every time they act habitually – that *would* be psychologically and phenomenologically implausible. I have simply said that, in finding themselves in an instance of *C*, they recognise it as one in which it is familiar that they *A*, and they *A* for the reason that it is the familiar thing to do. Further, the Familiarity View only says that habit-bearers' reasons are typically revealed by experience; not that they always are, or that they must be. One may simply think to oneself that something is familiar, and do it for that reason. However, I think that such reasons are typically revealed in experience because one's knowledge that something is familiar is (also typically) motivated by the feeling that it is so. Therefore, the feeling of familiarity plays something of an empirical role in the Familiarity View, the empirical role of being the medium which I think in fact provides us knowledge of our reasons. I will consider an objection to its playing this role in the final section, but it certainly seems plausible that it does.

Importantly, the Familiarity View does not entail, as Velleman's View did, that habit-bearers know which habits they have. It is not a pre-condition of taking an environment to be familiar, and taking doing something in it to be familiar that one knows that one does that thing out of *habit* in that environment. And taking it to be familiar that one *A*-s in *C* does not have the consequence that one knows one habitually does so either. The most my view requires is that habit-bearers know

which things are familiar to do in which contexts. But this seems harmless: surely anybody with a habit of brushing their teeth in the evening knows *that brushing my teeth in the evening is a familiar thing to do*, even if they do not have further knowledge that they have a habit doing so. Therefore, I think we can grant, notwithstanding the objection to come in Section 4, that the model says nothing too untoward about our psychology or phenomenology.

3. The model should: (a) minimally, be consistent with the metaphysics of habit; (b) ideally, go some way towards explaining how the metaphysics of habits connects with their role in rationalising explanations;

Part (a) requires, as a minimal adequacy condition, that a model be consistent with the metaphysics of habit I developed in Chapter One where I argued that habits are tendencies rather than capacities or dispositions. The basic idea is that whilst capacities figure in explanations of the possibility of someone's doing something, tendencies figure in explanations of *why* they do. The capacity to play tennis helps explain how it's possible that Alice plays tennis, but her tendency to do so helps explain why she does so, as well as why she plays so regularly, and so on. The Familiarity View is consistent with this because it has the fact that habits are tendencies built in: habits are tendencies to do things one has repeatedly done for the reason that they are familiar to do. So (4)-(a) is met.

Part (b) says that, ideally, a model should also provide an illuminating connection between the metaphysics of habits and their role in rationalising explanations. I think the Familiarity View does this. This is because habits, like other tendencies, manifest in something of a certain type's occurring. In this case, habits manifest in *S*'s doing something they are in the habit of doing: *A*-ing. That *S*'s *A*-ing is a matter of their habit's manifesting is therefore relevant to explaining *why S A*-s. But the reason for which *S A*-s is *that A-ing is familiar to S*. The Familiarity View connects these two facts about the explanation of *S*'s *A*-ing by saying that *S*'s habit is a tendency to *A for that very reason*. So the manifestation of *S*'s habit and *S*'s *A*-ing for this reason are identified. Therefore, the model provides an illuminating connection between the metaphysics of habit and its role in rationalising explanations.

4. When a habit-bearer *A*-s out of habit, their *A*-ing is (at least partly) explained by an idiosyncratic reason, a fact whose status as a reason for the habit-bearer to *A* depends on some feature of the habit-bearer;
5. The fact which is one's idiosyncratic reason for *A*-ing, when one *A*-s out of habit, is a fact whose truth depends on features of the habit-bearer.

The Familiarity View easily meets both constraints in basically the same way. The reason which the Familiarity View invokes is clearly idiosyncratic: it is a fact which only counts as a reason for someone with the history of acting that a habit-bearer like *S* has. Alice, who has never gone to *Gilmour's Café* before cannot have the reason I have as a habit-bearer, since my reason depends on

my history of acting, which Alice does not share. This fact both depends for its truth on my history of acting (5), and has its status as a reason in virtue of features of me (4).

I think it is clear that the Familiarity View meets these constraints smoothly. In my discussion, I have also tried to show that it is independently plausible, at least so long as we are trying to account for *how* the Rationalising View of Habit is true. I want, though, to make three further comments about the virtues of the Familiarity View. I do not see these virtues as constraints on a model of the Rationalising View, but I do see them as significant advantages worth spelling out.

Firstly, it is an advantage of my account that it does not tie habits too intimately to properties like character traits and virtues. Habits are clearly closely connected to those sorts of things, and it is understandable why we might reach to identify them with better understood things. However, habits are *odd*, and they come with their own distinctive baggage. I think that modelling them too closely on character traits or virtues just obscures their distinctive nature. Part of the worry here is one about reduction: we lose sight of habits if we identify them too closely with virtues or character traits. Part of the point of the arguments against the models based on those items was to show just how different habits are.

My second point is related. The Familiarity View provides us with a way of understanding the nature of habit in terms of some connected concepts – concepts of things being and seeming familiar, of repetition, of affect, and of reasons – which promises to actually teach us something about the phenomena. It promises elucidation without reduction, and provides us with a way to think about habit's 'internal structure'. It would be useless to give an account of habit that recognised its distinctiveness without saying anything about how having a habit figures in explanations of what we do. An account like that would see habit as a sort of 'dormative virtue' which would make it hard to see its explanatory relevance and its relations to other aspects of our mental lives. The Familiarity View has enough internal structure to show us how habits can figure in substantive explanations of what we do. The next, final, chapter aims to bring this out in full, but I hope this feature of my account is appreciable even from here.

Finally, in Chapter 2, I argued that mental properties are generally apt for figuring in broad rationalisations because they make up one's perspective. I also argued that habits are mental properties, broadly construed. That was sufficient to provide me with material to argue for the Rationalising View. But it may have remained puzzling how a habit could make up part of one's perspective on the world and on one's reasons. It is easy to see how beliefs, desires, and even virtues can do this. But, under the influence of Rylean ways of thinking, it can be quite hard to see how habits – brute behavioural tendencies – could also contribute to a point of view.

Now, there is a different tradition, with a different approach to the question of habit, in which this is not puzzling at all. Here is the phenomenologist Maurice Merleau-Ponty describing the acquisition of a habit:

> "[t]he subject [...] acquires the power of responding with a certain type of solution to a certain form of situation. The situations may differ widely from case to case, the responding movements may be entrusted sometimes to one effector organ and

> sometimes to another, and situations and responses resemble each other in the different cases much less through the partial identity of elements than by the *community of their sense*." (Merleau-Ponty, 2012, p. 143 my italics)

Merleau-Ponty's thought is that forming a habit involves linking certain types of responses to certain types of situation. However, the linkages between types is not to be found in fully extensional descriptions of environments and actions – partly constitutive of habit-formation is a change in the agent's perspective. There is a change in the intentional relations between habit-bearers and their environments, such that the *meaning* of the environment for the agent is altered by forming a habit. We can then ask *how* the meaning of the environment changes for a subject upon acquiring a habit.[111] What is the 'community of sense' which establishes the link between what one habitually does and the context one does it in? The Familiarity View gives us an answer. It tells us that part of what it is to become habituated is for there to be a change in one's perspective such that things become, and come to seem, familiar to do. Against the background of innumerable things one might do in a bathroom at bedtime, it is *brushing my teeth* which stands out as the thing to be done.

This allows us to see that habits are indeed like other mental properties in helping constitute our perspectives. To have a habit is not just to have a brute behavioural tendency which we must grudgingly call 'mental' just because there are no other options. Instead, they constitutively involve alterations in a subject's point of view on the world, how they take it to be, and which options are salient and attractive for them. The Familiarity View therefore allows us to integrate habit into the class of mental properties far more compellingly than we may have thought possible. That is an advantage of the view if nothing else is.

Section 4 – Wittgenstein and the Feeling of Familiarity

In this section, I want to develop and respond to a serious objection to my view. The objection attacks the Familiarity View's claim to meet constraint (2). It says that the view is not psychologically or phenomenologically plausible, because the feeling of familiarity is not as pervasive a feature of our phenomenal lives as the Familiarity View would have it. In fact, the objection claims, the feeling of familiarity is much rarer than habitual action, so cannot play the role I have asked it to.

The objection comes in a few forms and derives from Wittgenstein. However, I will need to demur somewhat from Wittgenstein's own discussion for two reasons. Firstly, Wittgenstein was clearly grappling with these issues over a long period of time and changed his mind between texts –

[111] Merleau-Ponty himself understands this in terms of habit-bearer's gaining new sensitivities to affordances, or opportunities for action, where this sensitivity is manifested as a kind of perceptual salience which motivates the agent to act (Merleau-Ponty, 2012, pp. 143–148). However, his focus on *possibilities* for action makes it seem that this is really a view of *skill acquisition*. Indeed, I think that Merleau-Ponty systematically conflates habits and skills, just as I warned we should not in Chapter 1. Despite this, the insight expressed in the passage I have quoted is absolutely correct, and a lot of what I have said in this thesis stems from trying to make good on it.

what he says in the *Philosophical Investigations* (2009) is quite different from what he says at some points in *The Brown Book* (1958). Secondly, his discussion is tied up in various classical Wittgensteinian quandaries about language, privacy, and understanding. That means that getting a grip on his own view of the matter would require diving far too deep into those murky waters. Therefore, I will treat some passages of Wittgenstein alone, out of context, as suggesting an argument against the Familiarity View. I will then develop the line of thought beyond what Wittgenstein says, but in ways which I take to be broadly Wittgenstein*ian*.

With qualifications made, here is a place where Wittgenstein questions whether there is a distinctive feeling of familiarity:

> "Pondering about the question whether there is such a feeling or not, we are likely to gaze at some object and say 'Don't I have a particular feeling when I look at my old coat and hat?' But to this we now answer: What feeling do you compare this with, or oppose it to? Should you say that your old coat gives you the same feeling as your old friend A with whose appearance too you are well-acquainted, or that whenever you happened to look at your coat you get that feeling, say of intimacy and warmth?" (Wittgenstein, 1958, pp. 180–181)

A page later, he says:

> "My room with all the objects in it is thoroughly familiar to me. When I enter it in the morning, do I greet the familiar chairs, tables, etc. with a feeling of 'Oh hello!'...?" (Wittgenstein, 1958, p. 182)

The worry in these passages is that when we look at familiar objects, although we are inclined to say that they look familiar, or that we experience them as being familiar, it is very hard to latch onto any particular feature of one's current experience which is *a feeling of familiarity*.[112] Whilst Wittgenstein recognises that sometimes there does seem to be an experiential episode which we can latch onto[113] – as in cases of déjà vu and of recognising a person, for example in the case I quoted from Ratcliffe in Section 2 – he is here doubting that we can do so in all cases where things seem familiar. If we cannot latch onto a distinctive mental episode, sensation, or 'component of experience' in every case, then not every case in which we are inclined to say that something looks or feels familiar is one in which there actually *is* such a feeling. We do not have 'Oh hello!' experiences whenever anything is familiar and we would be inclined to say that it is. Therefore, the feeling of familiarity is much rarer than I have suggested.

It is hard to know what to make of this objection. It primarily seems aimed at the view I discussed in Section 2 on which the feeling of familiarity is modelled on sensations, as perhaps suggested by his comments on the 'warm glow'. However, I rejected that view for similar sorts of

[112] Wittgenstein raises similar worries about facial recognition (Lyon, 1996, p. 88).

[113] For example, see Wittgenstein (1958, p. 182).

reasons: sensations are re-identifiable mental states, but the feeling of familiarity is diffuse and 'non-presentational'. So, about the sensation model, Wittgenstein and I agree.

However, I think that Wittgenstein's target is not only the sensation model, but rather *anyone* who thinks that there is a distinctive and pervasive phenomenal character to experiences of things on the basis of which we usually take them to be familiar. And that *is* a claim I would like to make. I have one main response to this.

Wittgenstein asks rhetorically whether two familiar things give one feelings that one can *compare*, presumably because only if one can compare two instances of a feeling one can then judge whether they are the same. The force of the question is obvious, since it is very implausible that there are comparable experiences when a coat looks familiar and when a friend does. But the same issues arise whenever we think about the non-presentational features of experience when it presents something as falling under an aspect. There are certain paradigmatic ways that rabbits look, such that one experience of a rabbit can be subjectively similar to another experience. But there is no *feeling of rabbithood* which I can compare across both cases and which I have both times. Nonetheless, we want to say that part of what it is for the two experiences to be similar is that both present perceived objects as being rabbits. So in aspectual phenomena, we just do not get features of experience which we can judge to be of the same kind on the basis of a comparison. And since the feeling of familiarity is an aspectual phenomenon, we should expect the same thing – the aspect of familiarity contributes to the character of an experience, but not in a way which allows us to *compare* the contributions of the aspect across distinct experiential episodes. In fact, it is plausible that the possibility of such a comparison is a mark of the sensation model of the feeling of familiarity, and that this represents another virtue of the aspectual model over the sensation model. Either way, the argument misses the mark.

One way some may want to develop Wittgenstein's about the feeling of familiarity scepticism (whilst departing from the text) is to ask why I am entitled to claim that there is a *feeling* of familiarity. Of course, things do *seem*, *look* and *sound* familiar, but seems, looks and sounds statements very often have an epistemic reading on which they are talk of judgement or belief. For example, 'the stocks seem to be doing poorly', 'their car isn't parked so it looks like they are out', and 'I heard Sally was chased by a bear – that sounds awful!'[114] These epistemic senses of 'seems' and 'looks' do not require that we posit experiences, just beliefs of the relevant kind. And since Wittgensteinians are very commonly sceptical of overpopulating the mind with things such as experiences, this may seem an attractive route.

However, we need to accept that there is a pervasive experiential sense in which things seem, look, and sound familiar because the epistemic sense does not capture all the right cases. For example, if Sally is an actor in a biopic about her own life which is filmed on a set built to look just like her home, things would seem familiar to her when she first walks on set. However, Sally would not believe that they were actually familiar; she knows that these are not her own sofa and coatrack. This means that the sense in which the set seems familiar is belief-independent, in the same way

[114] For discussion of these different senses, with a focus on 'looks' statements, see Chisholm (1957), Glüer (2012), Jackson (1977), and Martin (2010).

that the Müller-Lyer illusion is belief-independent – no matter how firmly one knows that the lines are equal, the visual illusion keeps its grip. This indicates that the sense in which things seem familiar to Sally must be prior to, and independent of, her beliefs. This further indicates that a phenomenal construal of 'seems' is appropriate.

So what should we say about when Sally gets home and her house seems familiar? Now, the objector may suggest that in this case the sense of 'seems' is merely epistemic. And of course, in this case, the epistemic sense of 'seems' will be appropriate, because Sally will believe this is her familiar house. But when she was on set, Sally *phenomenally* seemed familiar with things. Why should things be different for Sally when she goes home? Why would she *feel* familiar with the set but not *feel* familiar with her actual house? This seems quite implausible. We should assume that Sally must, therefore, also feel familiar at home; her home must be presented to her, in her experience, as being familiar. In which case, we seem to have an argument for the pervasiveness of the feeling of familiarity, as well as a response to the claim that the sense in which things normally seem familiar is merely epistemic. Both arguments for scepticism in the pervasiveness of the feeling of familiarity fail, and therefore this part of the Familiarity View stands. The feeling of familiarity can do the work I give it in the theory of habit.

Conclusion

In this chapter, I have sought to say *how* habits figure in broad rationalisations by considering a number of models for how they may do so. I have argued that Velleman's suggestions fail, and that we cannot model habits on virtues or character traits. I argued that we should accept five constraints of adequacy on any model of the Rationalising View of Habit, and that the Familiarity View meets these constraints easily where others fail. I also argued that the Familiarity View is both independently motivated and has significant virtues, particularly in integrating habits into a conception of mental properties as constituting one's perspective. Finally, I tried to argue for the ubiquity of the feeling of familiarity, which has a central role in the Familiarity View.

These arguments for the Familiarity View are far from conclusive. Rather, they provide accumulative support for it. Given the constraints, the attractiveness of the view, and that other models fail, the Familiarity View seems a strong hypothesis. This gives us good reason to think it at least approximates the way in which habits figure in broad rationalisations. In the final chapter, though, I will give another argument for the Familiarity View on the basis that that familiarity, and the feeling of familiarity, help explain some central and important features of habits. That will complete my case for the view.

CHAPTER 7

HABIT, DELIBERATION, AND THE FEELING OF FAMILIARITY

If there is a dogma in philosophical discussions of habit, it is that for someone to do something habitually, they must do that thing automatically.[115] Someone does something automatically in the relevant sense if and only if they do it without deliberating about what to do. I will call the view that a necessary condition on someone's doing something habitually is that they do it automatically 'The Non-Deliberative View' of habit. It is a view assumed by many, yet it has received scant argument. In this chapter, one of the things I aim to do is argue that the Non-Deliberative View is false.

This is worth doing for its own sake, since I think the Non-Deliberative View embodies the same sorts of (fairly crude) popular characterisations of habits as Gilbert Ryle's view did. It is yet another contrast between habits and a paradigmatic display of rational sensitivity, this time practical reasoning rather than skill. And I think that, insofar as these sorts of pictures prevail, there will always be a lingering suspicion about any view like the one I have been defending in this thesis.

There is a further reason, though. Considering the Non-Deliberative View is another way for me to defend the Familiarity View, which is a model of the Rationalising View of Habit. This is because it will allow me to show off some genuine explanatory advantages for the Familiarity View, since (in essence) I will argue that whatever explanatory work the Non-Deliberative Theorist thinks automaticity can do, familiarity does better. Therefore, the Non-Deliberative View acts as a foil for the defence of the Familiarity View on basically explanatory grounds.

This means that this chapter's role in the thesis can be understood in a couple of different ways. If you liked Chapter 2-5's argument for the Rationalising View, but were not convinced of my case for the Familiarity View in Chapter 6, then this chapter can be seen as a new argument for this particular model of the Rationalising View of Habit. However, if you left Chapter 6 thinking the Familiarity View seemed plausible, but wondered why we should care about it, this chapter provides an answer: the view can explain some important essential features of habit.

[115] See Brett (1981), Peters (2014), Pollard (2003, 2005b, 2006a, 2006b, 2008), Ryle (1970), Wu (2011). Julia Annas defends a similar but weaker claim in Annas (2011, 2012). See Douskos (2018a) and Owens (2017) for a rejection of this view.

Finally, if you were totally unconvinced by Chapters 2-5, Chapter 6's argument may well have been frustrating, since it assumed that the Rationalising View had been established. In which case, there is no reason to think too hard about the Familiarity View, since it's just a version of a view you weren't convinced by. In *that* case, this chapter provides an independent argument for the Familiarity View which doesn't depend on prior commitment to the Rationalising View of Habit. This means that, since this chapter argues for the Familiarity View directly, it is also an *in*direct argument for the Rationalising View. So, for the reader most sceptical of the Rationalising View of Habit, this can be seen as a last ditch attempt to show that habits figure in broad rationalising action-explanations.

In Section 1, I will present the Non-Deliberative View and explain the motivation for it as it is presented by its proponents. I argue, however, that this motivation cannot sustain a claim as strong as the Non-Deliberative View. In Section 2, I provide a new argument for the Non-Deliberative View which both builds on the original motivation and has the right form to provide the view's necessity claim. I will do this by identifying three features of habit which a theory must explain, and arguing that anything which can explain them could be granted the status of a necessary condition on someone's doing something habitually. These features are: that habit-formation requires repetition of some act-type; that having a habit involves insensitivity to non-habitual courses of action; and that habits have a distinctive role in action-explanation. I will argue that since automaticity plausibly figures in explanations of these features, it might fulfil important explanatory ambitions of a theory of habit which grant it the status of a necessary condition on acting habitually. If so, the Non-Deliberative View would be true.

However, in Section 3, I argue that there are at least five kinds of counterexample to the Non-Deliberative View which undermine its ability to explain the features I identify. In Section 4, I argue that there is a better explanation of the three phenomena which makes no mention of automaticity. The explanation I defend is that someone does something habitually only if one of the reasons for which they act is that the thing is familiar to do. This is, in essence, the Familiarity View from Chapter 6. This view of habit undercuts the explanatory ambitions of the Non-Deliberative View, and so leaves it with no way of rejecting the counterexamples. So I argue that we should reject the Non-Deliberative View in favour of one which has a person's familiarity with a course of action at its heart.

Firstly, I will explain the relevant sense of 'automaticity' the Non-Deliberative View of Habit invokes, and then show how the view is motivated.

Section 1 – The Non-Deliberative View of Habit

Philosophers who defend the Non-Deliberative View display a certain conception of automaticity with which we can formulate the view.[116] For example, Nathan Brett says that "Habitual

[116] There are multiple other conceptions of automaticity, particularly in psychology. However, I am singling out the one which figures in the Non-Deliberative View of habit. My arguments are not intended to show that there is *no* sense of

behaviour must in some respect be 'automatic', not the product of conscious decision" (Brett, 1981, pp. 363–364). He goes on to say that a dentist may wash his hands "from force of habit; but equally he might stop to think what he is doing each time and then proceed to wash up as a result of his deliberations [...] But he is not acting from force of habit if he is going through these deliberative manoeuvres" (Brett, 1981, p. 364). If the dentist washed his hands as an outcome of deliberating about what to do, then according to Brett what he did could not be habitual. This is because Brett thinks that the dentist washed habitually only if he washed automatically, and he washed automatically only if he did not wash as a result of deliberating. This is Brett's full characterisation of the sense of 'automatic' in his claim that acting automatically is necessary for acting out of habit. But since he determines whether someone does something automatically by seeing whether they deliberate, we have *prima facie* evidence that he thinks the necessary condition is also sufficient: the dentist washed automatically *if and only if* he washed, but not as a result of deliberation.

Another example is Julia Peters, who begins one paper by stating "When we act habitually, our actions are typically automatic. [...] [T]hey are automatic in the sense that they are not preceded by an explicit act of deliberation and decision." (Peters, 2014, p. 165).[117] Bill Pollard thinks that, necessarily, "a *habitual action* is [...] *automatic*, that is, it does not involve the agent in deliberation about whether to act" (Pollard, 2003, p. 415).[118] Wayne Wu argues that automaticity is a property actions have if and only if they are not directed by intentions (Wu, 2016, p. 104). Whilst Wu does not seem to think that intentions can only result from deliberation, on the common assumption that the conclusion of deliberation is an intention,[119] Wu's definition of automaticity excludes any automatic action from being the result of deliberation. And to the extent that Wu thinks of habitual actions as less-than-intentional, he thinks that they are automatic (Wu, 2011, p. 62).[120] Therefore, they cannot be results of deliberation.

Notice that Peters, Pollard, and Wu all state their theses about automaticity as theses about particular actions rather than what Jennifer Hornsby (1980) calls 'things done'. Things done are *types of action*.[121] They are things such as washing one's hands, looking for one's watch, or sitting down. Actions are the events or processes which are one's doings of those things – the particulars which instantiate the properties *being a washing of one's hands* or *being a sitting down*. Perhaps it will be thought that these distinctions cannot matter much here. However I think we cannot state

'automaticity' according to which acting habitually is necessarily automatic. However, I suspect that any such conception would have much more to do with sub-personal motor systems than deliberation. See Douskos (2017b, 2018a) for good overviews of the different conceptions of automaticity in psychology.

[117] Peters says 'typically', yet her defence of Pollard's view of virtue suggests she is sympathetic to the Non-Deliberative View's necessity claim.

[118] See also Pollard (2008, p. 216). Strictly speaking, Pollard thinks that the sort of deliberation ruled out by doing something habitually is deliberation about *what* to do or *whether* to do certain things. He allows we may deliberate about *how* to do something that we will do out of habit (Pollard, 2003, p. 416). However, this feature of Pollard's account is irrelevant, since I will be arguing that doing something habitually is compatible with its resulting from deliberating about *what* to do.

[119] See Paul (2013) for a defence of this view, and Dancy (2018) for arguments against it.

[120] Importantly, whilst Wu is a Non-Deliberative theorist of habit, his own definition of 'automaticity' makes no mention of deliberation. Because of this, the ensuing discussion of automaticity should not be taken to apply to him.

[121] So I will sometimes call them 'act-types' as a tense-neutral variant on 'things done'.

the Non-Deliberative theorist's conception of automaticity plausibly without attending to it, because stating their theses as primarily concerning actions is problematic.

For one thing, practical deliberation is reasoning about what to do.[122] With Hornsby's distinction in hand, it follows that practical deliberation is not reasoning about particular actions, but about whether one should do one type of thing or another type of thing.[123] Therefore, if the Non-Deliberative theorist's conception of automaticity is that *S*'s action, ϕ, was automatic if and only if *S* did not deliberate about whether to ϕ, and ϕ happens, then given we only deliberate about what to do and never about actions, all actions would be automatic. This is presumably not what the Non-Deliberative theorist means to say.

A more plausible statement of their thesis in terms of things done is the following: *S A*-ed (where '*A*' denotes a thing done) automatically if and only if *S* did not deliberate about whether to *A*, and *S A*-ed. This still allows us to state a thesis about actions: *S*'s action, ϕ, was automatic if and only if ϕ was an *A*-ing and *S A*-ed automatically. But since the right-hand of that biconditional makes essential reference to *A*-ing automatically, I take it that the more basic definition of automaticity is the one I gave in terms of things done. That is the definition of automaticity at work in this paper. With that in place, I can now state the Non-Deliberative View of habit more precisely:

> *Non-Deliberative View of Habit*: Necessarily, if *S A*-s habitually, then *S A*-s and *S* does not deliberate about whether to *A*.

What is the motivation for the Non-Deliberative View of habit? In essence, Non-Deliberative theorists are motivated by the fact that there is an over-abundance of cases where one does something habitually and automatically. Examples includes rubbing one's chin (Wu, 2016, p. 62), flicking on the light on entering a room (ibid.), biting one's nails (Pollard, 2006a, p. 233), saying grace before dinner (Anwander, 2007, p. 190), and taking a particular route to work (Romdenh-Romluc, 2013, p. 5). For people in the habit of doing those things, mostly they will do those things automatically, without deliberating. And in the face of such a wealth of examples, it has seemed to many philosophers that doing something automatically is a necessary condition on doing it habitually.

Insofar as this is a *motivation* for the Non-Deliberative View, it is fairly plausible. However, as an argument it will not work. Even if these examples make it attractive to think that whenever anyone acts habitually they act automatically, they are not evidence that it is necessarily so. At best we have evidence for a universal generalisation: '*everything done habitually is done automatically*'. But even if that were true, it would be insufficient to support the Non-Deliberative View since the truth of a generalisation does not entail its necessity. Considering these examples alone simply cannot tell us that there is anything other than a contingent relation between automaticity and acting

[122] Plausibly, practical deliberation can also be about how, when, or where to act. But the focus of this chapter is on the question of whether doing something habitually is compatible with deliberating about what to do, so we can ignore those other forms.

[123] Jonathan Dancy argues this persuasively, adapting a point from H. A. Prichard (1932) (Dancy, 2018, pp. 31–33).

habitually. In which case, however plausible the Non-Deliberative View might seem, we lack an argument for its necessity claim.

Although no compelling argument has been given for the Non-Deliberative View, I want to propose and consider one promising argument for it which builds on the general motivation for the view. The strategy is to appeal to the possibility that automaticity might play an essential explanatory role in the theory of habit, and that this is a plausible route to the Non-Deliberative View's necessity claim. I will develop this argument in the next section, before arguing against it in Section 3 and providing an alternative explanation in Section 4.

Section 2 – Automaticity, Repetition, Insensitivity, and Force

The strategy I want to consider for defending the Non-Deliberative View first identifies a number of features which seem to be essential to characterising habit. If one thinks, based on the examples considered in Section 1, that everything done habitually is done automatically, then one might look to this general truth to explain those features. If there is a plausible explanation stemming from automaticity, and there is no better explanation, and the features really are essential to characterising habit, then automaticity should also be thought essential to characterising habit. That is a way of moving from a universal generalisation – everything done habitually is done automatically – to the claim that the generalisation is necessarily true.

Here are three features central to habit which automaticity might plausibly figure in an explanation of:

1. *Repetition*: To form a habit of *A*-ing in context-type *C*, an agent must have repeatedly *A*-ed in instances of *C*.
2. *Insensitivity*: A habit of *A*-ing in context-type *C* makes the habit-bearer relatively insensitive to other non-habitual courses of action in instances of *C*.
3. *Force*: There is a psychological force of habit.

I will discuss these features in turn, firstly giving reasons for thinking they are essential to characterising habit, and then showing how automaticity might help explain them.

Everyone accepts that it is something like a conceptual truth that habit-formation requires repetition of act-types (Douskos, 2019a, p. 10; Owens, 2017, p. 173; Pollard, 2003, p. 411). One cannot habitually bite one's nails without having bitten them before; one cannot habitually interrupt people if one has not interrupted others before. And typically, though not always, one has to repeat these things more than once across relevantly similar contexts. But how does repetition contribute to the formation of a habit? Unless we can say what it is about repeatedly doing something which could contribute to habit-formation we will be in the dark about habituation.

One possible answer is: *S*'s repetition of *A* in *C* contributes to *S*'s forming a habit by making *A* become automatic for *S* when in *C*. What it is for *A* to 'become automatic' for *S* in *C* is to be

understood in terms of *S*'s gaining a kind of tendency to *A* in *C* such that *S* just *A*-s without deliberating about what to do.[124] Therefore, the Non-Deliberative theorist can attempt to explain *Repetition* in terms of automaticity by positing that a tendency to do something automatically is formed on the basis of repetition. This account therefore invokes automaticity, which it takes to be a general feature of habit, as an account of *Repetition*, which is an essential feature. If this is the best explanation of *Repetition*, then automaticity is also essential to habit.

The second fact, *Insensitivity*, is that habits make agents relatively insensitive to non-habitual courses of action in instances of context-types in which they have a habit of *A*-ing. We can see this by recognising that having a habit of *A*-ing can often lead one to *A* despite standing intentions to the contrary. For example, Komarine Romdenh-Romluc describes a case where I am intending to cycle to the park, become distracted and absentminded, and my habit of cycling to work 'takes over'.[125] Being in the right context-type, and not paying enough attention can mean that one just does the habitual thing, sometimes even when one is committed to another course of action. Another kind of insensitivity to alternatives that habits generate is the sense that doing the habitual thing is one's only option. For example, if I have an armchair I habitually settle into in the evening, then this habit makes me relatively insensitive to the possibility of sitting on the sofa. Whilst habits do not screen-off other options entirely, they do make alternative routes less salient, thereby making us less likely to take them. This is perhaps why mention of 'habit' is often prefaced with 'blind' or 'dumb'. An account of habit must explain *Insensitivity*.

Again, automaticity looks to do the explanatory work. If one automatically *A*-s when in *C*, then *A*-ing is set as a kind of default for the habit-bearer. When in *C*, the agent just *A*-s. But if that is the case, then this fact obstructs the agent from considering what else they could do and appreciating their alternatives. That is, automatically *A*-ing whenever in *C* competes with deliberating about what to do in *C* – the paradigmatic display of sensitivity to alternative options – and so precludes the agent's evaluation of other courses of action. So, if acting habitually meant acting automatically, this would explain why acting habitually entails a decreased sensitivity to alternatives.

Finally, an account of habits should tell us about the force of habit.[126] In saying that we sometimes act habitually, we impart habits with a role in explanations of why we do certain things. This role can be captured with talk of the *force of habit*, and we can sharpen our understanding of this role by asking: in virtue of what does a habit have any power over what one does? We can hone the question with an example. Say Alice and I both go to the *Café des habitudes*. Alice has never been before, and is going because she has heard nice things about it – that its food is good, and that it has a nice atmosphere. These are the reasons Alice goes there, and they are reasons that I also have.

---

[124] See Pollard (2003, p. 417) for an explicit formulation of this view.

[125] See Romdenh-Romluc (2013). This is linked to the fact that acting habitually often accompanies a certain lack of focal attention on what one is doing. The dispensing of such attention is often identified as closely related to automaticity, particularly by Wu (2011) and Douskos (2019a), however it is not clear how relevant attention is to the kind of automaticity at issue here.

[126] Talk of 'the force of habit' is riddled throughout the literature (Brett, 1981; Carlisle, 2006, 2014; Douskos, 2018a; Owens, 2017; Pollard, 2003, 2006b; Ryle, 1970). I myself had reason to make use of it in Chapter 4, Section 3, when I discussed variations on the story of Alice and Bert. My discussion of *Force* is intended to capture the point of such talk.

However, I have been there many times, and when I go on this occasion, I go out of habit. The question about the force of habit is this: what is it about my having the habit of going to *Café des habitudes* that plays a role in explaining why I go there, and which cannot play any role in explaining why Alice goes there? It is no good saying 'Well, I have the habit and she does not'. We want to know *why* my having the habit figures in an explanation of my going there. To give an adequate answer is to give an explanation of the force of habit which *Force* demands.

For someone who thinks that whenever someone acts habitually they act automatically, automaticity might seem to answer this explanatory demand. Perhaps the reason that my habit figures in an explanation of why I go to *Café des habitudes* is that, when I am in the appropriate context (whatever that may be) I go there automatically. The suggestion is that to say that I act out of force of habit is to say that when in *C*, I *A*, and that this connection bypasses other sorts of forces like those of considered desire and reflection. The view explains *Force* much the same as it does *Insensitivity*: habit-formation means that an agent forms the propensity to *A* automatically – without deliberation – in instances of *C*, and this propensity just manifests in *C*, precluding deliberation and playing an explanatory role in the agent's *A*-ing. This explanation has the nice feature of tying together the Non-Deliberative View's explanations of *Repetition*, *Insensitivity*, and *Force*. If this explanation is right, along with the others, then there is now a plausible explanatory argument for the Non-Deliberative View's claim that acting automatically is necessary for acting habitually.

In this section, I have presented a much more substantial case for the Non-Deliberative View than exists in the literature, albeit one which takes its lead from Non-Deliberative theorists' motivations for their view. By considering some examples, I granted the Non-Deliberative theorist the generalisation '*everything done habitually is done automatically*'. Then, I outlined three features that seem to be essential to habit and which must be accounted for, and argued that if the generalisation holds then automaticity can figure in plausible explanations of these features. If they are essential to habit, and the proposed explanations are good, and there are no better explanations, then the relation between automaticity and habit is not contingent. That is the proposed argument for the Non-Deliberative View of habit.

Section 3 – Counterexamples to the Non-Deliberative View

However attractive this line of thought might seem, I will argue that the Non-Deliberative View faces a number of serious counterexamples which show that not all things done habitually are done automatically. If this generalisation is false, then automaticity cannot explain *Repetition*, *Insensitivity*, and *Force*, because if it is not even generally true of things done habitually that they are done automatically, then it cannot be necessarily true. Here are the counterexamples.

Firstly, I might have a habit of running once a week, and usually habitually run on Mondays. However, one week I have other commitments and must deliberate about whether and when to run, and I decide to run on Tuesday. When I run on Tuesday, I run out of habit, despite my having

deliberated.[127] The Non-Deliberative View is committed to denying that when I run on Tuesday I run habitually, but there is no obvious non-*ad hoc* way to do this. One thing they could try is denying the appropriateness of the habit's specification as a *habit of running once a week*, preferring the *habit of running on Monday*.

But there are two problems. Firstly, my specification of the habit seems perfectly natural, despite being different from the canonical schema of 'the habit of *A*-ing in *C*'. But this is just because the context-type with which we specify the habit gives a *range* of days (a week) rather than any determinate days. But this does not stretch our concept of habit – I can habitually have a glass of water every hour (where there is no specific time within the hour that I have it) or have a habit of calling my mum once a week. These are perfectly natural and we should not feel pushed to purge them from our habit-ascriptions. Secondly, the suggested habit-specification also utilises the range-concept of 'Monday', so I could usually run on Monday morning but end up running on Monday evening out of habit as a matter of deliberating about when in the day to run, and whether to run at all. So the suggestion will not help.

The fact is that these sorts of habits, habits whose specifications of the relevant context-types include some indeterminacy or a range, invite certain sorts of practical questions. If one has a habit of running once a week, then the question about *when* and *whether* to run can arise because there is conflict between running on some specific day and other aspects of one's life. One solves these questions by deliberating, and, sometimes, by deciding to go running. And when one thereby runs, one does it habitually. Why should we deny this, except if we are blinkered by an antecedent commitment to the Non-Deliberative View?

Secondly, consider my habit of brushing my teeth every morning. Brushing my teeth in the morning can figure in a plan for a particularly busy morning, say one where I have to pack and leave for an early train. I could even make a list of what I intend to do after reasoning about it the night before: 'I must remember to pack, call a cab, shine my shoes, and yes, *I must brush my teeth*.' If I manage to seamlessly integrate my brushing of my teeth into my busy morning schedule, then I have acted habitually. This is because my plan makes space in my morning for my habit to manifest more easily in the context of other less familiar activities, and there is no reason why deliberately leaving space for a habit to manifest means that the ensuing action cannot be a manifestation of the habit. But the Non-Deliberative View cannot accommodate this fact.

A third example, used by Christos Douskos in a similar argument against the Non-Deliberative View, is a case where a person in the habit of *A*-ing must deliberate about whether to *A* or *B* on some occasion, and decides to *A* for reasons grounded in their habit of *A*-ing (Douskos, 2017b, p. 512).[128] For example, I habitually walk through the park on my way to work. The routine has given me a fondness for walking there – I am at ease, it feels familiar and calm, and I enjoy the routine. I have reasons for walking there which someone without the habit does not have. If I learned there is a quicker route, though, I might reconsider which way to go. I might deliberate for a moment, yet ultimately decide to go through the gardens on the strength of those reasons I have

[127] Thanks to Andrea Blomkvist for this example and lots of important discussions about how and why it works.

[128] The content of my example differs from Douskos's, but the form is the same and I owe much to his discussion of it.

which are grounded in my personal history, and the connection I have with the gardens built up by repeatedly walking in them.

This is a manifestation of my habit despite ensuing from deliberation because what I do is a response to reasons *internal* to my habit. Roughly, the reasons are internal to the habit because they are generated by the *habitual nature* of the activity, and not the activity taken in isolation.[129] That is, the reasons are *idiosyncratic*, and the reason is a fact which is dependent on a feature of me, the agent. Its truth and status as a reason depends on the fact that I have repeatedly walked through the park. So, even if there were no non-idiosyncratic reasons to walk through the park, reasons a first-time park-walker could have, for me, there is something good about doing what I always do – it is good to do what I am at ease with, that which is familiar and homely.[130] Those facts seem to be my reason, and they depend in a certain respect on my habit. I would not have *A*-ed had I lacked the habit, I *A*-ed as I habitually do, and I *A*-ed because I am in the habit of *A*-ing – that is, I acted for reasons internal to my habit which I weighed in deliberation.

The fourth counterexample is a case where frustrated deliberation is resolved by an agent's habit. Consider a common form of conscious deliberation about where to eat lunch:

> 1: Should I go to *Boring Bistro* or *Cold Café*?
> 2: *Boring Bistro*'s food is good but its atmosphere is a bit dull.
> 3: *Cold Café*'s livelier but very chilly.
> C: Actually, I'll just go to *Default Diner* like I always do.

The agent's practical reasoning is clearly invalid. Their reasoning is frustrated by indecision and yet a conclusion is reached because of the agent's knowledge of their habit of eating at *Default Diner* which, faced with indecision, they revert to. I will argue that subsequently going to *Default Diner* is both habitual and the result of deliberation.

Firstly, going to *Default Diner* ensues from deliberation simply because the intention to eat there is the conclusion of the agent's reasoning. The Non-Deliberative View does not entail that an agent does something habitually only if their doing so does not result from *valid* practical reasoning, so I do not need to show that my cases involve appropriate rational connections between premises and conclusion to show that acting habitually can ensue from deliberation.[131]

Eating at *Default Diner* is habitual because the agent settles on it rather than the other cafés because their habit, coupled with their knowledge of it, sets *Default Diner* as the default option. Without the agent's knowledge of their own habit, *Default Diner* would simply be one more café to deliberate about, and so could not figure as somewhere which could function as a default conclusion

[129] This sense of 'internal reason' is designed to map onto a standard use of 'internal relation' according to which a relation between *X* and *Y* is 'internal' whenever the relation holds just in virtue of the *natures* of *X* and *Y*. What the relation is to Bernard Williams's (1979) sense of 'internal reason' is, I do not know.

[130] David Owens misses the possibility of habit-dependent reasons when considering habitual actions as counterexamples to the guise of the good (Owens, 2017).

[131] I don't need to show that unless there are strong general reasons for thinking that actions which seem to follow from invalid practical reasoning in fact do not, but I doubt there are.

in a case of frustrated reasoning. Given *Default Diner* does function as a default option, and does so because of the agent's habit, the habit figures in an explanation of the agent's going to *Default Diner*, not in any strange or deviant way, but by doing what habits often do: setting, and motivating us to act on, default options. So the agent's going to *Default Diner* is habitual and deliberative.

The fifth and final counterexample depends on a feature of action-explanation which is almost totally obscured in the literature, so it will take some effort to uncover. Say I crossed my legs as I sat down. Someone can now ask me:

Standard Question: Why did you cross your legs?

Philosophers tend to think that the *Standard Question* requests *the* reason for one of my *actions*. An extremely influential example of this Donald Davidson, who begins 'Actions, Reasons and Causes' with: "What is the relation between *a reason* and *an action* when *the reason* explains *the action* by giving *the agent's reason* for doing what he did?" (Davidson, 1980, p. 685 emphasis added).

There are a number of reasons this is mistaken. The first is that reasons are reasons for which we do certain sorts of things. However, these things are not actions. Instead, they are kinds of which particular actions are members. Hornsby's distinction between actions and things done is of paramount importance here: what I do is cross my legs, and my reason for doing it is a reason for crossing my legs; but the event of my crossing of my legs, my action, is not explained by a reason (Hornsby 1980; 1997; Sandis 2012).[132] The second is that, as Alfred Mele points out, there is typically no single reason for which anyone does anything, but a "whole raft of reasons" (Mele 2017, 55).

That said, I want to focus on a third argument which receives almost no discussion: that the *Standard Question* obscures the fact that someone's crossing their legs has many explanations which are answers to a number of different contrastive questions whose meanings can be exposed in English by means of emphasis.[133] A more representative variety of requests for explanation is this:

1. Why *did* you cross your legs (when you did, rather than some other time)?
2. Why did *you* cross your legs (whereas this other person didn't)?
3. Why did you *cross* your legs (rather than keep them apart)?
4. Why did you cross *your* legs (rather than anybody else's legs)?
5. Why did you cross your *legs* (rather than your arms)?

These questions all ask for my reasons for doing something – crossing my legs – but answers to these questions will not necessarily cite the same reasons. My answer to (1) may be that I cross my legs whenever I sit down, and I sat down a little while ago; my answer to (3) may be that it's more comfortable to sit that way; and my answer to (4) may be that it would be a bit awkward to

[132] In effect, this is the argument I gave in Chapter 2, Section 3: reasons explain facts about what agents do, and they are reasons to do things, rather than reasons for particular actions.

[133] For two exceptions, see Dretske (1972) and Snedegar (2017). Contrastivity is, however, the topic of significant discussion in the philosophy of science and explanation.

just fold someone else's legs over. My reasons are facts which explain the contrastive facts presupposed in the contrastively different questions about why I crossed my legs.[134] But, importantly, whilst these reasons explain different contrastive facts, they are facts about the very same thing done – my crossing my legs. The reasons shed light on why I did one thing, by explaining a number of different facts about it, facts which differ contrastively.

If that is right, then we can often explain why I do one thing by appealing to many different factors which figure in explanations of different contrastive facts about my doing that thing. And this gives us the space to see yet another way that habit and deliberation can both figure in an explanation of why I do something: they may figure in different explanations of different contrastive facts about why I did something. For example, Clyde, a mean boss, has a habit of firing someone whenever he is angry. One day, he becomes furious and, his habit kicking in, he decides to fire someone. But who? After a brief think, he chooses to fire Alex because on balance he dislikes him the most.[135] We can now ask:

a) Why did Clyde *fire* Alex?
b) Why did Clyde fire *Alex*?

The fact that Clyde has a nasty habit of firing people when angry is one reason for his firing Alex, but it is not his reason for firing *Alex rather than anyone else*.[136] So that fact can be given in answer to (a) but not to (b). The answer to (b) is that Clyde dislikes Alex more than his other employees. Now, Clyde arrived at his reason for firing *Alex* through deliberation, and so his deliberating figures in an explanation of why he fired *Alex*, even though it did not figure in an explanation of why he *fired* him. And a reason he *fired* Alex is that he has this bad habit, and so his habit figures in an explanation of that fact, but not in the fact that he fired *Alex*. But Clyde only did one thing: he fired Alex. Clyde's habit and Clyde's deliberation both figure in explanations of different contrastive facts about why he did what he did, but this just means that they figure in mutually illuminating explanations of why Clyde fired Alex. But this means that the Non-Deliberative View's claim that doing something habitually and doing that thing as a result of deliberation are incompatible is false. Clyde fired Alex out of habit, and he fired him as a result of deliberation.

I have given five counterexamples to the view that all cases of habitual action are automatic, and together they show that the Non-Deliberative View is false because they show that it is not true that everything done habitually is done automatically. But if that is not true then automaticity cannot figure in explanations of essential features of habit. Let me consider how the Non-Deliberative theorist might respond.

---

[134] The standard semantics for the meaning of questions involves appealing to facts presupposed by questions (Achinstein, 1977; Cross, 1991; Ginzburg, 2011; Sæbø, 2016).
[135] This example is adapted from Fred Dretske's masterful discussion of contrastive statements at (Dretske, 1972, p. 419).
[136] I have *not* said that the fact that Clyde has a nasty habit of firing people when angry is *Clyde's reason* for firing anyone – in the case I am describing, Clyde does not treat that fact as a reason for him to fire anyone. However, it is one of the reasons why he does what he does and that is all my argument turns on. See Alvarez (2010, Chapter 1) and Chapter 2 for more on this.

3.1 – A Response

I suspect that some will eye my counterexamples with suspicion. Perhaps they will worry that my examples are just too fringe, too far from the paradigmatic examples canvassed in Section 1, to be taken seriously.

This is not a tenable response. Regardless of whether the Non-Deliberative View's favoured cases are *paradigms*, we cannot discard putatively less paradigmatic examples as not falling under the concept *habit*. Giving metaphysical accounts requires one to engage with tricky cases which challenge one's preconceptions, and to account for the cases in a principled and unified way. In the philosophy of perception, for example, accounts are typically defended by reflecting on visual experience of ordinary material objects. However, some accounts struggle to deal with perceptions of shadows, holes, and the spaces between things precisely because of a one-sided diet of examples.[137] It is no good for those accounts to simply disregard these examples as non-paradigmatic cases of vision and thereby rule them out from falling under the concept *seeing*. Perhaps we do not strictly speaking see shadows, holes and immaterial paths, but that conclusion requires an *argument*. We cannot reject putatively non-paradigmatic cases by reflex.

The Non-Deliberative theorist needs a principled, non-question-begging reason to *either* deny that my examples are of habit, *or* deny that they are results of deliberation. The chances of an argument which saves the intuition that these are cases of acting habitually by denying that they ensue from deliberation seem dim. One might question whether planning space for an activity I am in the habit of engaging in on a busy morning counts as practical deliberation of the right sort, and one could wonder whether frustrated reasoning is actually the abandonment of deliberation. But it would be a thankless task to attempt to argue, for example, that Clyde did not fire Alex as a result of deliberating. Any available argument would likely force the conclusion that deliberation rarely produces action at all. But this would, as they say, 'prove too much'.

Therefore, the Non-Deliberative theorist must reject my examples. The closest they can come to a plausible argument against them is that we should reject these as counterexamples to the Non-Deliberative View because automaticity provides a compelling explanation of *Repetition*, *Insensitivity*, and *Force*. These are features that a theory of habit must explain, and automaticity fills that explanatory role. The cost is ruling out the cases in Section 3. Whatever one's intuitions about them, the explanatory advantage of accounting for *Repetition*, *Insensitivity* and *Force* far outweighs any un-intuitiveness of ruling out those cases.

But to most this will seem hopelessly *ad hoc*. Surely we need some theory-independent reasons to deny that my cases are really counterexamples to the Non-Deliberative View, rather than a mere re-assertion of that view. I have argued that my counterexamples show that not everything done habitually is done automatically, and so automaticity cannot play a central explanatory role in theorising about essential features of habit. It is simply not sufficient to respond by saying 'but it *does* play that explanatory role, so they *can't* be good counterexamples'.

[137] For discussion of these sorts of problems, see Mac Cumhaill (2015) and Sorenson (2008).

Now, if one thought that explaining *Repetition*, *Insensitivity*, and *Force* in terms of automaticity was the *only* way to do so, then perhaps we would all be forced to bite the bullet and reject my cases. But this is highly implausible. Indeed, in the next section, I will provide a new, different, explanation of those three features in terms of very different notions: those of familiarity and the feeling of familiarity. This alternative explanation therefore undercuts the last hope of the Non-Deliberative View, the hope that it is the only available option. Therefore, I will conclude that it's false.

Section 4 – Habit and Familiarity's Reasons

I have argued that the Non-Deliberative View of Habit is false because there are a wide range of serious counterexamples which it can neither reject, accommodate, nor explain away. Not everything done habitually is done automatically, which means it cannot be necessarily so. But we are now left with a problem: how should we explain *Repetition*, *Insensitivity*, and *Force* if not by invoking automaticity?

My proposal is that we can advert to features articulated in Chapter 5 for an explanation: when one has a habit of doing something, that thing is familiar to do, and one's familiarity with doing something is a strong candidate for helping explain *Repetition*, *Insensitivity*, and *Force*. First, I will briefly recapitulate the relevant points about familiarity and its connection with habit. Then, I will argue that it can figure in explanations of our three features, thereby showing that, necessarily, one does something habitually only if one does it for the reason that it is familiar.

Firstly, there is the distinction *being familiar* and *seeming familiar*. Something is familiar to someone to the extent that they have engaged with it in various appropriate ways before. Something's seeming familiar is primarily a matter of someone's having an experience of it *as* being familiar. In good cases, I argued that when something seems familiar, the experience reveals the familiar thing as being familiar *because it is*. However, there are also bad cases of déjà vu and illusion.

Secondly, I argued that the feeling of familiarity is affective, but that its affective profile cannot be captured by thin notions of positive or negative valence. Instead, we capture its affective nature by speaking of the sorts of values it reveals and implies. For example, the feeling of familiarity connected to a sense of ease, comfort and safety; yet the familiar can also feel stale and alienating.

Thirdly, the fact that doing something is familiar to one is a reason for one to do it. This is because if doing something is familiar, one knows how it goes; there are no surprises, and one can act relatively risk-free. Putting this together with the views above, we get the view that in good cases, the feeling of familiarity reveals a reason for action, for that is what the fact that something is familiar to do is.

Enter the connection with habits. What determines whether one is familiar with *a course of action* is whether one has engaged in it before, and since one must have done something repeatedly in order to form a habit of doing it, this entails we are familiar with the things we do habitually. There is therefore a very tight connection, mediated by repetition, between habit and familiarity. We

are familiar with everything done habitually because whether one has a habit of *A*-ing and whether one is familiar with *A*-ing are determined by the very same thing: the extent to which one has repeatedly *A*-ed. That is, it is sufficient for being familiar with doing something that one has a habit of doing it. Given this, and the fact that the feeling of familiarity serves to reveal the familiar when appropriate, when in an instance of *C*, a person in the habit of *A*-ing in *C* will feel that *A*-ing is the familiar thing to do. Having doodled much in past lectures, for example, I am familiar with doing so, and this (in good cases) explains why doodling in a lecture feels familiar to me.[138]

Therefore, these tight connections between repetition, habit, familiarity, and phenomenal familiarity make it more than *prima facie* plausible that everything done habitually is and feels familiar to do: it seems to be necessarily true in virtue of internal relations between the natures of habit, repetition, and familiarity which I have been outlining.[139] But whilst I think that I have said enough to make that claim both plausible and attractive, I will now argue that considerations regarding familiarity can also help explain *Repetition*, *Insensitivity*, and *Force*. I will start with *Force*.

The question of *Force* was: what is it about my having the habit of *A*-ing that plays a role in explaining why I *A* when I do so out of habit? With our new generalisation in hand – everything done habitually is familiar to do – plausibly familiarity fits this explanatory role. Now, for philosophers such as Mark Johnston (2001b) and Peter Poellner (2016), I have already said enough to explain *Force*. I have described the feeling of familiarity as a kind of affective experience, and those philosophers think that affective experiences should be understood across the board as revelations of values which exert motivational force on our behaviour. If one thinks that, then one has an easy explanation of *Force*: the feeling of familiarity is affective, and therefore what it reveals – familiar things – exert a motivational force on one which can figure in explaining why one acts. Given the connection between habit and the familiar, this would be a neat account of habit's distinctive explanatory role requested in *Force*.

I am sympathetic to this view, however it is deeply controversial and requires a defence which I cannot give here. Instead, I think there is simpler argument which provides a similar conclusion, but which does not rely on, or entail, the wholly general thesis about affective experience that Johnston and Poellner defend. It is sufficient to remember that the fact that *A*-ing is familiar to *S* is a reason for *S* to *A*.

As Clare Carlisle argues, there is a "sense of comfort, safety and ease that is engendered by familiarity" which contributes to "[insulating] us from the threat of the unknown" (Carlisle, 2006, p. 23). Carlisle argues that this is why "even during a week away one finds a regular haunt: the café one returns to each morning [...]. In combining the novel with the familiar they [...] make one feel at home in a new place" (Carlisle, 2014, p. 78). The point is that there is a kind of safety and ease in doing what we always do – the familiar paths through contexts we often find ourselves in are well-trodden, and, in contrast with courses of action that we are unfamiliar with, are vouched for by one's

[138] For ease of expression, I will proceed without the 'good cases' qualification.

[139] This argument is therefore different from the motivation for the Non-Deliberative View, even if they perhaps look superficially similar. That argument attempted to invalidly move from a universal generalisation to a necessity claim. However, the motivation for my own necessity claim is derived from necessary connections between habit, familiarity, and repetition.

own history. Take an example from Section 3. The fact that I always go to *Default Diner* means that I know what it's like there and what to expect. Whilst I know some things about *Boring Bistro* and *Cold Café*, they are places that I could be surprised by, and so represent something of a risky alternative to *Default Diner*. Therefore, my familiarity with *Default Diner* grounds a reason to go there – the reason is that it is familiar.

Of course, I may be wrong about the relative values of these three places. Perhaps both *Boring Bistro* and *Cold Café* are (despite their names) lively, warm, and a great improvement on *Default Diner*. So, whatever reason grounded in familiarity I have to go there is outweighed by the reasons to go to those other places. But that is beside the point. The point is that I do have *some* reason to go to *Default Diner* just because I am familiar with it, even if acting for that reason involves obscuring stronger reasons to the contrary. But this should not be surprising. Habits are essentially conservative forces, in that they keep us in the realm of the known, and thereby keep us from situations which, though they may be fun and new, may be dangerous, difficult or unpleasant. By being familiar, *Default Diner* represents the conservative, safe, option. That is a reason to go there over-and-above the reasons someone unfamiliar with *Default Diner* could have.

Repeatedly *A*-ing in *C* – be that going to *Default Diner* at lunchtime, or firing employees when angry – therefore bootstraps reasons for *A*-ing in instances of *C* by making one familiar with *A*-ing in *C* where this grounds a reason to *A* when in an instance of *C*. What is phenomenal familiarity's role in this?

The feeling of familiarity discloses the fact that *A*-ing in *C* is familiar and performs the role of making it experientially manifest to the agent that they have a reason to *A* when in *C*. In presenting the familiar course of action *as* familiar, it presents the act-type as having a property which grounds a reason to do it. And it does this just by revealing what we are familiar with doing under the aspect of familiarity.[140] It therefore forms a bridge between what the agent has reason to do and the agent's own practical and epistemic position.

This account of familiarity makes it available for figuring in rationalising action-explanations, and it is *this* which accounts for *Force*. Take again the contrast between my going to *Café de Habitudes* out of habit and Alice's going there for reasons available to anyone. We want to know what is it about my having the habit of going there that plays a role in explaining why I do (when I go out of habit) which must be missing in an explanation of why Alice goes there. My answer is that, because I am familiar with going there, and I feel it to be familiar, I have and appreciate a reason to go there which is grounded in the fact that I have repeatedly eaten there before. *That eating there is familiar to me* is a reason which depends on my past and which is appreciated by virtue of

[140] This might sound odd, but some comparisons may help. If I have made a promise to someone, then only some ways of thinking of them can represent them as a person to whom I owe something. Thinking of them as 'the last person I spoke to', or 'Fred' will not do. What will represent them as someone to whom I have a duty is 'my promisee'. So one way for me to grasp the fact that I have reasons to do things for this person is to know that *they are a promisee of mine*. Similarly, to see a pile of clothes as untidy is to see it as grounding a reason to tidy them up. To feel a sensation as painful is to grasp a reason to get rid of it. So, there is a class of concepts related to affect whose application seem to entail the grasping of reasons. Therefore familiarity is not strange in this respect. I do not claim to have fully elucidated this point, but I do hope to have made it plausible.

my feeling familiar with eating there, and as such that reason figures in an explanation of why I go there. But it clearly cannot explain Alice's going there. She is not familiar with *Café des habitudes* and in a good case will not feel it to be familiar. She therefore lacks a reason that I have just because of the difference in our histories. And this is surely what we want as an explanation of *Force*. It accounts for why I eat at *Café des habitudes* by providing an account of the force of habit as the force of a reason which is dependent on my history, and which Alice cannot have without repeatedly eating at that café.

This helps us explain *Insensitivity* because phenomenal familiarity with one course of action involves insensitivity to others. This is derived from the fact that we take in familiar seeming things more easily and to the exclusion of unfamiliar things. Feeling familiar with a language typically means that one does not have to attend closely to hear the different words, whereas new speakers struggle; feeling familiar with a visual scene means not having to attend much to grasp its content. Similarly, phenomenal familiarity with a course of action means not having to attend as closely in order to see what is involved in doing it. One already knows how it goes. But in that case, we can see why habit-bearers are somewhat insensitive to alternative possibilities. If there is a path which seems familiar, all the circumstances are normal, there is no special reason to engage in an act of attentive search for an alternative, and the familiarity with the path presents reasons to take it, then those features militate against looking for alternatives. The nature of the feeling of familiarity gives us a very plausible explanation of the insensitivity of habit-bearers.[141]

Finally, how does repetition contribute to habit-formation? If what I have said is right, then we have a simple answer: *S*'s repetition of *A* in *C* contributes to *S*'s forming a habit by making *S* familiar with, and feel familiar with, *A*-ing when in instances of *C*. This account is initially plausible because it draws on connections between repetition, habit and familiarity which I have already outlined. But my explanation of *Force* provides us with even stronger reason to accept it. Repeatedly *A*-ing makes one familiar with it, thereby providing us with the reason to act which is drawn upon in explaining *Force*. So my explanation of *Force* implies this explanation of *Repetition*.

Since I argued that anything which explains *Repetition*, *Insensitivity*, and *Force* counts as a necessary feature of acting habitually, I have argued that, necessarily, if one does something habitually, one does it for the reason that doing that thing is familiar to do. Before concluding, I will outline how my view is compatible with the cases I have discussed, and address a potential worry some may have.

Firstly, it is worth saying that my view is compatible with every case of acting habitually that I discussed in Section 1 and Section 3. My view is compatible with all the cases I discussed in Section 3 because there is no incompatibility between doing something because it is familiar and deliberating about whether to do it. In fact, my view even nicely explains counterexample three, where an agent deliberates and decides to do what they habitually do on the basis of reasons grounded in their having the habit. This is because, on my view, we always do this in habitual action: we act for reasons

[141] This whole section owes an awful lot to Tyler Haddow, but these points in particular are highly indebted to him. The point is also made at by Amelie Rorty (1980, p. 210).

'internal' to our habit. It's just that, in that example, the agent makes those reasons *explicit* in their thinking.

One might worry that, since my view places so much emphasis on acting for reasons grounded in familiarity, these reasons push out or obscure other more general reasons we have for doing what we habitually do. For example, on my view, to brush one's teeth habitually, one must brush for the reason that doing so is familiar. But don't we usually brush our teeth because it's good for us to have clean teeth? If so, either few of us are really in the habit of brushing our teeth, or, if we are, we do not do so for the right reasons. Put like this, the worry generalises into a concern about the rationality of acting habitually, for, if acting habitually means acting for reasons other than good reasons, then acting habitually cannot be rational.[142]

However, these worries are misplaced. I have argued only that doing something because it feels familiar is necessary for doing something habitually. This is compatible with it being one amongst a "whole raft of reasons" (A. Mele, 2017, p. 55) for which one does that thing. There is no reason at all to think that once one becomes familiar with brushing one's teeth, say, that the reasons the familiarity grounds push out or occlude the reasons that everyone always has to brush their teeth. Adding a reason to one's stock of reasons to do something does not compete with one's other reasons to do it. So on my view one can brush one's teeth habitually *and* because it's good to have healthy teeth.

That said, my view has something positive to say about instances where a person does something seemingly utterly irrational because of their habit. For example, a person may have a nasty habit of picking their nose which they have every reason to suppress during a job interview. Unfortunately, despite themselves, they pick their nose out of habit in the interview. Now, on the Non-Deliberative View, this is explained by the sheer operation of automaticity, and what they did is rendered utterly irrational since there are no practical reasons to do it, and yet they did it anyway. But if there is genuinely nothing to be said for their picking their nose, then it is barely intelligible why they do it at all. Certainly, we can give no story about why from the agent's own perspective. But in that case, it is hard to see how we can trace the nose-picking to the agent at all, and this makes it hard to see the picking of the nose as an action. If any view is in danger of ruling acting habitually out from being rational, it is the Non-Deliberative View. On my view, however, no-one acting habitually is rendered unintelligible because they are always acting for at least one reason grounded in their familiarity with a course of action. Granted, this nose-picker is not maximally rational, but we can think of them more along the lines of an akratic person rather than someone whose behaviour in that instance is utterly disconnected from their reasons. So, far from worrying that my view obscures the rationality of doing things habitually, it in fact preserves it.

Conclusion

[142] Thanks to an anonymous reviewer for this objection.

I have argued against the Non-Deliberative View by showing that there are serious counterexamples to it, and that the only principled response to these counterexamples is undercut by a better explanation of *Repetition*, *Insensitivity*, and *Force* than the Non-Deliberative View has. My view also better accommodates the counterexamples discussed in Section 3 by allowing that acting habitual can result from deliberation. Therefore, we should abandon the widespread dogma that necessarily, everything done habitually is done automatically, in favour of a view which has agents' familiarity with courses of action at its heart.

CONCLUSION

In this thesis, I have argued for the Rationalising View of Habit: the view that habits, when they manifest, figure in broad rationalising explanations of why the habit-bearer does what they do, when they do it habitually. Further, I have argued for an account of how habits figure in those rationalisations: they do so by being tendencies to act for the reason that *doing this is familiar*, where this reason is typically revealed by the feeling of familiarity.

I began, in Chapter 1, arguing against pervasive prejudices about habit and habitual action by focussing on Gilbert Ryle's influential work on habit and skill. Having laid some of the groundwork for the rest of my thesis, I continued in Chapter 2 and Chapter 3 to provide an extensive discussion of rationalisations, reasons, and intention that formed the framework within which I argued for the Rationalising View. My argument, ultimately, came down to the following:

1. Habits are mental properties (broadly construed);
2. If mental properties figure in explanations of why someone does something intentionally, then they figure in broad rationalisations of why they do that thing;
3. Whenever habits figure in explanations by manifesting, they figure in explanations of why the habit bearer does something intentionally.

C) Therefore, whenever habits manifests, they figure in broad rationalisations

I defended Premise (1) in Chapter 2 (and, I suppose, in Chapter 1). The materials for defending Premise (2) were drawn together in Chapters 2 and 3, and I provided the explicit argument for it at the end of Chapter 3. I argued for Premise (3) in Chapter 4, and I defended it against a number of powerful objections in Chapter 5. That concluded my argument for the Rationalising View of Habit.

Then, in Chapters 6 and 7, I turned my attention to *how* habits figure in broad rationalisations. First, I considered and rejected some potentially promising models of the Rationalising View, and derived a number of demanding constraints on an adequate model. I argued that the Familiarity View – the view that habits are tendencies to do things for the reason that doing them is familiar, where this reason is typically revealed by the feeling of familiarity – is both

independently plausible and could meet these constraints. Finally, I argued that it promises to actually explain a number of central features of habit, most importantly, the force of habit.

I have argued for these views by paying close and careful attention to the nature of habits, and by showing how they connect in various ways to other more familiar topics in the philosophy of mind and action. Of course, I have been on habit's side this whole time, trying to exhibit and display their intrigues and intricacies, and trying to show that they are worthy of deep thought.

However, I have also tried to avoid what I think is an all too common trap: attempting to *vindicate* habits by *accepting* the prejudice that they are mindless and produce only arational and non-intentional action, and then arguing that because this is the case, all our traditional views about the importance of reason and intention to agency should be committed to the flames. Indeed, that is the attitude I took into this project. The problem was that, as I have tried to show, the prejudices *themselves* do not hold up. Habitual actions display agents' intelligence; everything done habitually is done intentionally, and so is done for a reason; habits have a rational role in action-explanations; they alter the landscape of one's reasons, and one's perspective on that landscape. So we cannot vindicate habit's interest and importance by accepting tired and worn preconceptions about them. Even if the details of the views I have defended in this thesis do not work out, I hope to at least shown this much.

BIBLIOGRAPHY

Achinstein, P. (1977). What Is an Explanation? *American Philosophical Quarterly*, *14*(1).

Althusser, L. (1969). *For Marx*. Verso.

Althusser, L., & Balibar, É. (2016). *Reading Capital: The Complete Edition*. Verso.

Alvarez, M. (2009). Acting Intentionally and Acting for a Reason. *Inquiry: An Interdisciplinary Journal of Philosophy*, *52*(3).

Alvarez, M. (2010). *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford University Press.

Alvarez, M., & Hyman, J. (1998). Agents and Their Actions. *Philosophy*, *73*(2).

Annas, J. (2011). *Intelligent Virtue*. Oxford University Press.

Annas, J. (2012). Practical Expertise. In J. Bengson & M. A. Moffett (Eds.), *Knowing How: Essays on Knowledge, Mind, and Action* (1st ed., pp. 101–112). Oxford University Press. https://doi.org/10.1093/acprof

Anscombe, G. E. M. (1979). Under a Description. *Noûs*, *13*(2), 219–233.

Anscombe, G. E. M. (2000). *Intention*. Oxford University Press.

Anwander, N. (2007). The Reach of Habit. In A. Leist (Ed.), *Action in Context*. Berlin: Walter de Gruyter.

Aristotle. (2002). *Nichomachean Ethics* (S. Broadie & C. Rowe (eds.)). Oxford University Press.

Aristotle. (2006). *Aristotle: Metaphysics Theta* (S. Makin (ed.)). Oxford University Press UK.

Arvidson, P. S. (2003). A Lexicon of Attention: From Cognitive Science to Phenomenology. *Phenomenology and the Cognitive Sciences*, *2*(2).

Arvidson, P. S. (2013). Restructuring Attentionality and Intentionality. *Human Studies*, *36*(2).

Audi, R. (1986). Acting for Reasons. *Philosophical Review*, *95*(4).

Bäckström, S., & Gustafsson, M. (2017). Skill, Drill, and Intelligent Performance: Ryle and Intellectualism. *Journal for the History of Analytical Philosophy*, *5*(5). https://doi.org/10.15173/jhap.v5i5.3205

Baier, A. (1971). The Search for Basic Actions. *American Philosophical Quarterly*, *8*(2).

Baier, A. (1972). Ways and Means. *Canadian Journal of Philosophy*, *1*(3).

Barlassina, L., & Hayward, M. K. (2019). More of me! Less of me!: Reflexive Imperativism about Affective Phenomenal Character. *Mind*, *128*(512), 1013–1044. https://doi.org/10.1093/mind/fzz035

Bird, A. (1998). Dispositions and Antidotes. *Philosophical Quarterly*, *48*(191).

Block, N. J., & Fodor, J. A. (1972). What Psychological States are Not. *The Philosophical Review*, *81*(2), 159–181. https://doi.org/10.2307/2183991

Boyle, M., & Lavin, D. (2010). Goodness and Desire. In S. Tenenbaum (Ed.), *Desire, Practical Reason, and the Good*. Oxford University Press.

Brandom, R. B. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Bratman, M. E. (1991). Cognitivism About Practical Reason. *Ethics*, *102*(1), 117–128. http://www.jstor.org.sheffield.idm.oclc.org/stable/2381726
Bratman, M. E. (2000). Reflection, Planning, and Temporally Extended Agency. *Philosophical Review*, *109*(1), 35–61.
Bratman, M. E. (2018). *Planning, Time, and Self-Governance: Essays in Practical Rationality*. OUP USA.
Brett, N. (1981). Human Habits. *Canadian Journal of Philosophy*, *11*(3), 357–376.
Brewer, B. (2008). How To Account For Illusion. In A. Haddock & F. Macpherson (Eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press.
Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, *4*(1).
Byrne, A. (2009). Experience and Content. *Philosophical Quarterly*, *59*(236).
Carlisle, C. (2006). Creatures of Habit: The Problem and Practice of Liberation. *Continental Philosophy Review*, *38*, 19–39.
Carlisle, C. (2014). *On Habit* (1st ed.). Routledge.
Carlson, G. N. (2005). Generics, Habituals and Iteratives. In A. Barber (Ed.), *Encyclopedia of Language and Linguistics*. Elsevier.
Castañeda, H.-N. (1966). "He": A study in the logic of self-consciousness. *Ratio*, *13*(2), 130–157.
Chang, R. (2004). Can Desires Provide Reasons for Action? In R. J. Wallace, P. Pettit, S. Scheffler, & M. Smith (Eds.), *Reason and Value: Themes From the Moral Philosophy of Joseph Raz* (1st ed.). Oxford University Press.
Chisholm, R. M. (1957). *Perceiving: A Philosophical Study*. Cornell University Press.
Cohen, M., & Hornsby, J. (1982). The Agent's Independence of the World. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *56*, 21–50.
Corneille, O., Monin, B., & Pleyers, G. (2005). Is positivity a cue or a response option? Warm glow vs evaluative matching in the familiarity for attractive and not-so-attractive faces. *Journal of Experimental Social Psychology*, *41*(4), 431–437.
Cross, C. B. (1991). Explanation and the Theory of Quetions. *Erkenntnis*, *34*(2), 237–260. https://doi.org/10.1007/BF00385722
Dancy, J. (1995). Arguments From Illusion. *The Philosophical Quarterly*, *45*(181), 421–438. https://doi.org/10.2307/2220307
Dancy, J. (2000). *Practical Reality* (1st ed.). Oxford University Press.
Dancy, J. (2008). On How to Act: Disjunctively. In A. Haddock & F. Macpherson (Eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press.
Dancy, J. (2018). *Practical Shape: A Theory of Practical Reasoning* (1st ed.). Oxford University Press.
Darwall, S. L. (1983). *Impartial Reason*. Cornell University Press.
Davidson, D. (1980). Actions, Reasons and Causes. In *Essays on Actions and Events* (1st ed.). Oxford University Press.
Davidson, Donald. (1970). Action and Reaction. *Inquiry: An Interdisciplinary Journal of Philosophy*, *13*(14).
Davidson, Donald. (1980a). Actions, Reasons, and Causes. In *Essays on Actions and Events* (1st ed., pp. 3–21). Oxford University Press.
Davidson, Donald. (1980b). Agency. In *Essays on Actions and Events*. Oxford University Press. https://doi.org/10.1111/j.1468-0149.1981.tb01007.x
Davidson, Donald. (1980c). Freedom to Act. In *Essays on Actions and Events* (1st ed., pp. 63–83). Oxford University Press.

Davidson, Donald. (1980d). Mental Events. In *Essays on Actions and Events* (1st ed.). Oxford University Press.
De Vries, M., Holland, R. W., Chenier, T., Starr, M. J., & Winkielman, P. (2010). Happiness cools the warm glow of familiarity: Psychophysiological evidence that mood modulates the familiarity-affect link. *Psychological Science*, *21*(3), 321–328.
Delacroix, S. (2017). Law and Habits. *Oxford Journal of Legal Studies*, *37*(3).
Department, H. P. (2018). *2016 Whitehead Lecture 2 -- Stephen Yablo (MIT) "Ways and Relevance" [Video]*. Youtube Https://Www.Youtube.Com/Watch?V=Yd_yPhqWppY.
Dewey, J. (2007). *Human Nature and Conduct: And Introduction to Social Psychology*. Cosimo Classics.
Dokic, J. (2010). Perceptual Recognition and the Feeling of Presence. In B. Nanay (Ed.), *Perceiving the World*. Oxford University Press.
Dokic, J., & Lemaire, S. (2013). Are Emotions Perceptions of Value? *Canadian Journal of Philosophy*, *43*(2).
Döring, S. A. (2003). Explaining action by emotion. *Philosophical Quarterly*, *53*(211).
Döring, S. A. (2007). Seeing What to Do: Affective Perception and Rational Motivation. *Dialectica*, *61*(3).
Douskos, C. (2013). Settling and Bodily Control. *Inquiry: An Interdisciplinary Journal of Philosophy*, *56*(6).
Douskos, C. (2017a). Habit and Intention. *Philosophia (United States)*, *45*(3), 1129–1148. https://doi.org/10.1007/s11406-016-9810-z
Douskos, C. (2017b). Pollard on Habits of Action. *International Journal of Philosophical Studies*, *25*(4), 504–524. https://doi.org/10.1080/09672559.2017.1355406
Douskos, C. (2018a). Deliberation and Automaticity in Habitual Acts. *Ethics in Progress*, *9*(1), 25–43. https://doi.org/10.14746/eip.2018.1.2
Douskos, C. (2018b). Learning, Acquired Dispositions and the Humean Theory of Motivation. *Philosophical Papers*, *47*(2), 199–233. https://doi.org/10.1080/05568641.2017.1422989
Douskos, C. (2019a). The Spontaneousness of Skill and the Impulsivity of Habit. *Synthese*, *March*, 1–24. https://doi.org/10.1007/s11229-017-1658-7
Douskos, C. (2019b). The Varieties of Agential Powers. *European Journal of Philosophy*, *27*(4).
Downey, G. (2010). Throwing like a Brazilian: On Ineptness and a Skill-Shaped Body. *Anthropology of Sport and Human Movement*, 297–326.
Dretske, F. (1972). Contrastive Statements. *The Philosophical Review*, *81*(4), 411–437. https://doi.org/10.2307/2183886
Ehl, T., Roberton, M. A., & Langendorfer, S. J. (2005). Does the Throwing "Gender Gap" Occur in Germany? *Research Quarterly for Exercise and Sport*, *76*(4), 488–493. https://doi.org/10.1080/02701367.2005.10599322
Evaldsson, A.-C. (2003). Throwing Like a Girl?: Situating Gender Differences in Physicality Across Game Contexts. *Childhood*, *10*(4), 475–497. https://doi.org/10.1177/0907568203104006
Fara, M. (2005). Dispositions and Habituals. *Noûs*, *39*(1).
Ferrante, E. (2015). *The Neapolitan Novels Boxed Set*. Europa Editions.
Ferreira, M. (2016). The Semantic Ingredients of Imperfectivity in Progressives, Habituals, and Counterfactuals. *Natural Language Semantics*, *24*(4).
Fish, W. (2009). *Perception, Hallucination, and Illusion*. Oxford University Press.
Ford, A. (2014). Action and Passion. *Philosophical Topics*, *42*(1).
Frankfurt, H. (1977). Identification and Externality. In A. Rorty (Ed.), *The Identities of Persons*.

University of California Press.
Frankfurt, H. (1982). The Importance of What We Care About. *Synthese*, *53*(2).
Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, *66*(23), 829–839.
Fredrickson, B. L., & Harrison, K. (2005). Throwing Like a Girl: Self-Objectification Predicts Adolescent Girls' Motor Performance. *Journal of Sport and Social Issues*, *29*(1), 79–101.
Fricker, E. (2009). Is Knowing a State of Mind? The Case Against. In D. Pritchard & P. Greenough (Eds.), *Williamson on Knowledge*. Oxford: Oxford University Press.
Ginzburg, J. (2011). Questions: Logic and Interactions. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of Logic and Language* (2nd ed., pp. 1133–1146). Elsevier.
Glüer, K. (2012). Martin on the Semantics of "Looks." *Thought: A Journal of Philosophy*, *1*(4).
Goldman, A. I. (1971). The Individuation of Action. *The Journal of Philosophy*, *68*(21), 761–774. https://doi.org/10.2307/2024949
Haddock, A. (2005). At One With Our Actions, But At Two With Our Bodies: Hornsby's Account of Action. *Philosophical Explorations*, *8*(2).
Haddock, A. (2011). The Knowledge That a Man Has of His Intentional Actions. In A. Ford, J. Hornsby, & F. Stoutland (Eds.), *Essays on Anscombe's Intention*. Harvard University Press.
Hampshire, S. (1999). The Reason Why Not. *New York Review of Books*. https://www.nybooks.com/articles/1999/04/22/the-reason-why-not/
Hampson, M. (2020). The Learner's Motivation and the Structure of Habituation in Aristotle. *Archiv Für Geschichte Der Philosophie*.
Heuer, U. (2014). Intentions and the Reasons for Which We Act. *Proceedings of the Aristotelean Society*, *114*(3), 291–315. https://doi.org/10.1111/j.1467-9264.2014.00374.x
Horgan, T. (1984). Functionalism and Token Physicalism. *Synthese*, *59*(3), 321–338. http://www.jstor.org.sheffield.idm.oclc.org/stable/20116004
Hornsby, J. (1980). *Actions*. Routledge and Kegan Paul.
Hornsby, J. (1981). Acts and Other Events by Judith Jarvis Thomson. *Journal of Philosophy*, *78*(4).
Hornsby, J. (1997a). Agency and Causal Explanation. In *Simple Mindedness: In Defence of Naive Naturalism i the Philosophy of Mind* (1st ed., pp. 129–155). Harvard University Press.
Hornsby, J. (1997b). Physicalist Thinking and Conceptions of Behaviour. In *Simple Mindedness: In Defence of Naive Naturalism i the Philosophy of Mind1* (1st ed., pp. 111–128). Harvard University Press.
Hornsby, J. (2008a). A Disjunctive Conception of Acting For Reasons. In A. Haddock & F. Macpherson (Eds.), *Disjunctivism: Perception, Action, Knowledge* (pp. 244–262). Oxford University Press.
Hornsby, J. (2008b). Agency and Alienation. In M. de Caro & D. MacArthur (Eds.), *Naturalism In Question*. Harvard University Press.
Hornsby, J. (2012). Ryle's Knowing-How, and Knowing How to Act. In J. Bengson & M. Moffett (Eds.), *Knowing How: Essays on Knowledge, Mind, and Action* (pp. 80–98). Oxford University Press. https://doi.org/10.1093/acprof
Hume, D. (1777). *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*. Greenwood Press.
Hursthouse, R. (1988). Moral Habituation. *Oxford Studies in Ancient Philosophy*, *6*.
Hursthouse, R. (1991). Arational Actions. *Journal of Philosophy*, *88*(2).
Hyman, J. (2006). Knowledge and Evidence. *Mind*, *115*(460).
Ingerslev, L. R. (2020). On the Role of Habit for Self-Understanding. *Phenomenology and the*

Cognitive Sciences, *19*(3), 481–497. https://doi.org/10.1007/s11097-018-9605-8
Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge University Press.
James, W. (2000). *The Prinicples of Psychology*. Dover Publications.
Jaster, R. (2020). *Agents' Abilities*. De Gruyter.
Johnston, M. (2001a). Is Affect Always Mere Effect? *Philosophy and Phenomenological Research*, *63*(1).
Johnston, M. (2001b). The Authority of Affect. *Philosophy and Phenomenological Research*, *63*(1), 181–214.
Koikkalainen, P., Syrjämäki, S., & Pettit, P. (2004). Eye to Eye: Interview With Philip Pettit. *Niin & Näin*, 1–14.
Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.
Langendorfer, S. J., & Roberton, M. A. (2002). Individual Pathways in the Development of Forceful Throwing. *Research Quarterly for Exercise and Sport*, *73*(3), 245–256. https://doi.org/10.1080/02701367.2002.10609018
Langton, R. (2004). Intention as Faith. In H Steward & J. Hyman (Eds.), *Agency and Action*. Cambridge University Press Press.
Lavin, D. (2011). Problems of Intellectualism. *Jurisprudence*, *2*(2).
Lavin, D. (2013). Must There Be Basic Action? *Noûs*, *47*(2).
Lavin, D. (2015). Action as a form of temporal unity: On Anscombes Intention. *Canadian Journal of Philosophy*, *45*(5–6), 609–629. https://doi.org/10.1080/00455091.2015.1125589
Lekan, T. (2007). Actions, Habits, and Practices. In A. Leist (Ed.), *Action in Context*. Walter de Gruyter.
Lenman, J. (2009). Reasons for Action: Justifiation vs. Explanation. In *Stanford Encyclopedia of Philosophy2*. https://stanford.library.sydney.edu.au/archives/sum2012/entries/reasons-just-vs-expl/
Levine, S. (2015a). McDowell, Hegel, and Habits. *Hegel Bulletin*, *36*(2), 184–201. https://doi.org/10.1017/hgl.2015.17
Levine, S. (2015b). Norms and Habits: Brandom on the Sociality of Action. *European Journal of Philosophy*, *23*(2), 248–272. https://doi.org/10.1111/j.1468-0378.2012.00530.x
Lewis, D. K. (1981). Are We Free to Break the Laws? *Theoria*, *47*(3).
Lewis, D. K. (1994). Humean Supervenience Debugged. *Mind*, *103*(412).
Lewis, D. K. (1997). Finkish Dispositions. *Philosophical Quarterly*, *47*(187).
Lyon, G. (1996). The Experience of Perceptual Familiarity. *Philosophy*, *71*(275).
Lyons, J. (2005). Clades, Capgras, and Perceptual Kinds. *Philosophical Topics*, *33*(1).
Mac Cumhaill, C. (2015). Perceiving Immaterial Paths. *Philosophy and Phenomenological Research*, *90*(3), 687–715. https://doi.org/10.1111/phpr.12037
Magrì, E. (2018). Emotions, Motivation, and Character: A Phenomenological Perspective. *Husserl Studies*, *34*(3).
Mangan, B. (2001). Sensation's Ghost: The Nonsensory Fringe of Consciousness. *PSYCHE: An Interdisciplinary Journal of Research On Consciousness*, *7*.
Marcus, E. (2009). Why There Are No Token States. *Journal of Philosophical Research*, *34*.
Martin, J.-R. (2013). Experiences of Activity and Causality in Schizophrenia: When predictive Deficits lead to a Retrospective Over-binding. *Consciousness and Cognition*, *22*(4), 1361–1374. https://doi.org/https://doi.org/10.1016/j.concog.2013.09.003
Martin, M. G. F. (2004). The Limits of Self-Awareness. *Philosophical Studies: An International*

Journal for Philosophy in the Analytic Tradition, *120*(1/3), 37–89.
Martin, M. G. F. (2006). On Being Alienated. *Perceptual Experience*, 354–410.
Martin, M. G. F. (2010). What's in a look? In B. Nanay (Ed.), *Perceiving the World*. Oxford University Press.
Matthen, M. (2010). Two Visual Systems and the Feeling of Presence. In N. Gangopadhyay, M. Madary, & F. Spicer (Eds.), *Perception, Action, and Consciousness: Sensorimotor Dynamics and Two Visual Systems*. Oxford University Press.
Maurin, A.-S. (2016). Tropes: For and Against. In F. F. Calemi (Ed.), *Metaphysics and Scientific Realism: Essays in Honour of David Malet Armstrong*. De Gruyter.
McDowell, J. (1979). Virtue and Reason. *The Monist*, *62*(3).
McDowell, J. (1982). Reason and Action. *Philosophical Investigations*, *5*(4).
McDowell, J. (1983). Criteria, Defeasibility, and Knowledge. In *Proceedings of the British Academy, Volume 68: 1982* (Vol. 68). Oxford University Press.
McDowell, J. (1986). Singular Thought and the Extent of "Inner Space." In J. McDowell & P. Pettit (Eds.), *Subject, Thought, and Context*. Clarendon Press.
McDowell, J. (2008). The Disjunctive Conception of Experience as Material For a Transcendental Argument. In F. Macpherson & A. Haddock (Eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press.
McDowell, J. (2011a). *Perception as a Capacity for Knowledge*. Marquette University Press.
McDowell, J. (2011b). Some Remarks on Intention in Action. *The Amherst Lecture in Philosophy*, *6*, 1–18.
McDowell, J. (2011c). What is the Content of an Intention in Action? In *Agents and their Actions* (Vol. 110, pp. 61–78). https://doi.org/10.1002/9781444346763.ch4
McDowell, J. (2013). Acting in the Light of a Fact. In D. Bakhurst, M. O. Little, & B. Hooker (Eds.), *Thinking About Reasons: Themes From the Philosophy of Jonathan Dancy*. Oxford University Press.
McGeer, V. (2018). Intelligent Capacities. *Proceedings of the Aristotelian Society*, *118*(3), 347–376. https://doi.org/10.1093/arisoc/aoy017
Meiksins Wood, E. (2016). Rethinking Base and Superstructure. In *Democracy Against Capitalism* (5th ed.). London, New York: Verso.
Mele, A. (2017). *Aspects of Agency: Decision, Abilities, Explanations, and Free Will*. Oxford University Press.
Mele, A. R. (1992). Acting for Reasons and Acting Intentionally. *Pacific Philosophical Quarterly*, *73*.
Merleau-Ponty, M. (2012). *Phenomenology of Perception*. Routledge. https://books.google.co.uk/books?id=r1OhbwAACAAJ
Miliband, R. (2015a). Political Action, Determinism and Contingency. In *Class War Conservatism and Other Essays* (2nd ed., pp. 131–153). Verso Books.
Miliband, R. (2015b). The Capitalist State: Two Exchanges with Nicos Poulantzas. In *Class War Conservatism and Other Essays* (2nd ed., pp. 26–35). Verso Books.
Mitchell, J. (2017). The Epistemology of Emotional Experience. *Dialectica*, *71*(1).
Molnar, G. (2003). *Powers: A Study in Metaphysics*. Oxford University Press.
Monin, B. (2003). The Warm Glow Heuristic: When Liking Leads to Familiarity. *Journal of Personality and Social Psychology*, *85*(6), 1035.
Moran, D. (2011). Edmund Husserl's Phenomenology of Habituality and Habitus. *Journal of the British Society for Phenomenology*, *42*(1), 53–77.
Moran, R. (2004). Anscombe on 'Practical Knowledge.' *Royal Institute of Philosophy Supplement*,

55.
Morris, K. (2012). *Starting With Merleau-Ponty*. Bloomsbury Continuum.
Mumford, S. (1998). *Dispositions*. Clarendon Press.
O'Shaughnessy, B. (1980). *The Will: A Dual Aspect Theory (Vol. 2)*. Cambridge University Press.
O'Shaughnessy, B. (2008). *The Will: A Dual Aspect Theory (Vol. 2)*. Cambridge University Press.
Owens, D. (2017). Habitual Agency. *Philosophical Explorations*, *20*(115), 93–108. https://doi.org/10.1080/13869795.2017.1356358
Parfit, D. (1997). Reasons and Motivation. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *71*, 99–146. http://www.jstor.org/stable/4106956
Paul, S. K. (2013). The Conclusion of Practical Reasoning: The Shadow Between Idea and Act. In *Canadian Journal of Philosophy* (Vol. 43, Issue 3, pp. 287–302). https://doi.org/10.1080/00455091.2013.847345
Pavese, C. (n.d.-a). Know-How, Action, and Luck. *Synthese*.
Pavese, C. (n.d.-b). Knowledge, Action, Defeasibility. In J. Brown & M. Simion (Eds.), *Reasons, Justification, and Defeaters*. Oxford University Press.
Pavese, C. (2016a). Skill in epistemology I: Skill and knowledge. *Philosophy Compass*, *11*(11).
Pavese, C. (2016b). Skill in epistemology II: Skill and know how. *Philosophy Compass*, *11*(11).
Peters, J. (2014). On Automaticity as a Constituent of Virtue. *Ethical Theory and Moral Practice*, *18*(1), 165–175. https://doi.org/10.1007/s10677-014-9516-x
Phillips, E. (2018). *The Darker Steed: Reason, Passion, and Self-Awareness*. UCL.
Poellner, P. (2016). Phenomenology and the Perceptual Model of Emotion. *Proceedings of the Aristotelian Society*, *116*(3).
Pollard, B. (2003). Can Virtuous Actions Be Both Habitual And Rational? *Ethical Theory and Moral Practice*, *6*(4), 411–425.
Pollard, B. (2005a). Naturalizing the Space of Reasons. *International Journal of Philosophical Studies*, *13*(1), 69–82. https://doi.org/10.1080/0967255042000324344
Pollard, B. (2005b). The Rationality of Habitual Actions. *Proceedings of the Durham-Bergen Philosophy Conference*, 39–50.
Pollard, B. (2006a). Actions, Habits and Constitution. *Ratio*, *19*(2), 229–248. https://doi.org/10.1111/j.1467-9329.2006.00322.x
Pollard, B. (2006b). Explaining Actions with Habits. *American Philosophical Quarterly*, *43*(1), 57–69.
Pollard, B. (2008). *Habits in Action: A Corrective to the Neglect of Habits in Contemporary Philosophy of Action*. VDM Verlag Dr. Müller.
Pollard, B. (2011). Identification, psychology, and habits. In *New waves in philosophy of action* (pp. 81–97). Springer.
Preston, B. (1996). Merleau-Ponty and Feminine Embodied Existence. *Man and World*, *29*(2).
Prichard, H. A. (1932). Duty and Ignorance of Fact. *Philosophy*, *8*(30).
Proust, M. (1984). *In Search of Lost Time: Swann's Way* ((trans.) C. K. Scott Moncrieff & T. (Kilmartin (eds.); 3rd ed.). Penguin.
Quinn, W. (1993). Putting Rationality In Its Place. *Morality and Action*, *228*.
Ratcliffe, M. (2004). Interpreting Delusions. *Phenomenology and the Cognitive Sciences*, *3*(1), 25–48. https://doi.org/10.1023/b:phen.0000041899.08813.38
Raz, J. (2009). Reasons: Explanatory and Normative. In C. Sandis (Ed.), *New Essays on the Explanation of Action* (pp. 184–203). Palgrave Macmillan.
Romdenh-Romluc, K. (2011). Agency and Embodied Cognition. *Proceedings of the Aristotelean*

Society, *111*(1), 79–95. https://doi.org/10.1111/j.1467-9264.2011.00299.x
Romdenh-Romluc, K. (2013). Habit and Attention. In R. T. Jensen & D. Moran (Eds.), *The Phenomenology of Embodied Subjectivity* (pp. 3–21). Springer.
Rorty, A. O. (1980). Akrasia and Conflict. *Inquiry: An Interdisciplinary Journal of Philosophy*, *23*(2), 193–212.
Ruben, D.-H. (1990). *Explaining Explanation* (Issue 2). Routledge.
Ruben, D.-H. (2018a). One-Particularism In the Theory of Action. *Philosophical Studies*, *175*(11).
Ruben, D.-H. (2018b). *The Metaphysics of Action: Trying, Doing, Causing*. Palgrave Macmillan.
Ruggeri, C. (2019). A Criticism of Young's 'Throwing Like a Girl' Through Scheler's Understanding of Motor Action. *Continental Philosophy Review*, *52*(4), 335–359. https://doi.org/10.1007/s11007-019-09475-8
Runion, B. P., Roberton, M. A., & Langendorfer, S. J. (2003). Forceful Overarm Throwing: A Comparison of Two Cohorts Measured 20 Years Apart. *Research Quarterly for Exercise and Sport*, *74*(3), 324–330. https://doi.org/10.1080/02701367.2003.10609098
Russell, B. (1911). Knowledge by Acquaintance and Knowledge by Description. *Proceedings of the Aristotelian Society*, *11*.
Russell, D. C. (2009). *Practical Intelligence and the Virtues*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199565795.001.0001
Ryle, G. (1970). *The Concept of Mind* (4th ed.). Penguin.
Sachs, C. B. (2014). *Intentionality and the Myths of the Given*. Pickering & Chatto.
Sæbø, K. (2016). "How" Questions and the Manner–Method Distinction. *Synthese*, *193*(10).
Saito, Y. (2017). *Aesthetics of the Familiar: Everyday Life and World-Making*. Oxford University Press.
Sandis, C. (2010). Basic Action and Individuation. In C. Sandis & T. O'Connor (Eds.), *A Companion to the Philosophy of Action* (1st ed., pp. 10–17). Blackwell.
Sandis, C. (2012). The Objects of Action Explanation. *Ratio*, *25*(3), 326–344. https://doi.org/10.1111/j.1467-9329.2012.00545.x
Sandis, C. (2013). Can Action Explanations Ever Be Non-Factive? In D. Bakhurst, M. O. Little, & B. Hooker (Eds.), *Thinking About Reasons: Themes From the Philosophy of Jonathan Dancy*. Oxford University Press.
Sandis, C. (2015). He Buttered the Toast While Baking a Fresh Loaf. *Philosophy and Public Issues*, *Supplementary Issue*, 27–42.
Scanlon, T. (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.
Scarantino, A., & Nielsen, M. (2015). Voodoo Dolls and Angry Lions: How Emotions Explain Arational Actions. *Philosophical Studies*, *172*(11).
Schellenberg, S. (2018). *The Unity of Perception: Content, Consciousness, Evidence* (1st ed.). Oxford University Press.
Searle, J. R. (1979). The Intentionality of Intention and Action. *Inquiry: An Interdisciplinary Journal of Philosophy*, *22*(1–4).
Setiya, K. (2003). Explaining Action. *Philosophical Review*, *112*(3).
Setiya, K. (2007). *Reasons Without Rationalism* (Issue 3). Princeton University Press.
Setiya, K. (2018). *Intention*. Stanford Encyclopedia of Philosophy.
Shemmer, Y. (2007). Desires As Reasons. *Philosophy and Phenomenological Research*, *75*(2).
Siegel, S. (2010). *The Contents of Visual Experience*. Oxford University Press.
Siewert, C. (1998). *The Significance of Consciousness*. Princeton University Press.
Siewert, C. (2005). Attention and Sensorimotor Intentionality. In D. Woodruff Smith & A. L.

Thomasson (Eds.), *Phenomenology and Philosophy of Mind*. Oxford: Clarendon Press.
Small, W. (ms.). The Intelligence of Virtue and Skill.
Small, W. (2012). Practical Knowledge and the Structure of Action. In G. Abel & J. Conant (Eds.), *Rethinking Epistemology Volume 2*. De Gruyter.
Small, W. (2016). Bodily Movement and Its Significance. *Philosophical Topics*, *44*(1).
Small, W. (2017a). Agency and Practical Abilities. *Royal Institute of Philosophy Supplement*, *80*, 235–264. https://doi.org/10.1017/s1358246117000133
Small, W. (2017b). Ryle on the Explanatory Role of Knowledge How. *Journal for the History of Analytical Philosophy*, *5*(5).
Small, W. (2020). Practical Knowledge and Habits of Mind. *Journal of Philosophy of Education*, *54*(2).
Smith, M. (1994). *The Moral Problem*. Blackwell.
Smith, M. (2012). Four Objections to the Standard Story of Action (and Four Replies). *Philosophical Issues*, *22*(1).
Smith, Z. (2018). Dance Lessons for Writers. In *Feel Free* (pp. 136–149). Penguin.
Snedegar, J. (2017). *Contrastive Reasons* (1st ed.). Oxford University Press.
Sorenson, R. (2008). *Seeing Dark Things*. Oxford University Press.
Stanley, J., & Willlamson, T. (2001). Knowing How. *Journal of Philosophy*, *98*(8).
Steward, Helen. (1997). *The Ontology of Mind: Events, Processes, and States*. Oxford University Press.
Steward, Helen. (2009). Sub-Intentional Actions and the Over-Mentalization of Agency. In C. Sandis (Ed.), *New Essays on the Explanation of Action*. Palgrave-Macmillan.
Steward, Helen. (2012a). *A Metaphysics for Freedom*. Oxford University Press.
Steward, Helen. (2012b). Action as Processes. *Philosophical Perspectives*, *26*(i), 373–388.
Steward, Helen. (2013). Responses. *Inquiry*, *56*(6), 681–706. https://doi.org/10.1080/0020174X.2013.841055
Steward, Helen. (2017). Action as Downward Causation. *Royal Institute of Philosophy Supplement*, *80*.
Steward, Helen. (2018). Review of "The Metaphysics of Action: Trying, Doing, Causing" by D-H. Ruben. *Notre Dame Philosophical Review*.
Stout, R. (2009). Was Sally's Reason for Running From the Bear That She Thought It Was Chasing Her? In C. Sandis (Ed.), *New Essays on the Explanation of Action*. Palgrave-Macmillan.
Strawson, P. F. (2002). *Individuals*. Routledge.
Stroud, B. (2013). Doing Something Intentionally and Knowing That You Are Doing It. *Canadian Journal of Philosophy*, *43*(1).
Sylvan, K. (forthcoming). The Eclipse of Instrumental Rationality. In *The Routledge Handbook of Practical Reason*.
Thalberg, I. (1971). Singling Out Actions, Their Properties and Components. *Journal of Philosophy*, *68*(21).
Thompson, E. P. (1978). *Poverty of Theory*. nyu Press.
Thompson, M. (2008). *Life and Action: Elementary Structures of Practice and Practical Thought*. Harvard University Press.
Thompson, M. (2011). Anscombe's Intention and practical knowledge. In A. Ford, J. Hornsby, & F. Stoutland (Eds.), *Essays on Anscombe's Intention*. Harvard University Press.
Thomson, J. J. (1979). Acts and Other Events. *Philosophy of Science*, *46*(1).
Titchener, E. B. (1918). *A Beginner's Psychology*. New York: The Macmillan Company.

Travis, C. (2004). The Silence of the Senses. *Mind*, *113*(449), 57–94.
Van Fraassen, B. (1980). *The Scientific Image*. Oxford University Press.
Vanello, D. (2018). Affect, Perceptual Experience, and Disclosure. *Philosophical Studies*, *175*(9).
Velleman, D. (1989). *Practical Reflection*. Princeton University Press. https://doi.org/10.7208/chicago/9780226468549.003.0005
Velleman, D. (1999). Love as a Moral Emotion. *Ethics*, *109*(2).
Velleman, D. (2000). *The Possibility of Practical Reason*. Oxford University Press.
Vetter, B. (2015). *Potentiality: From Dispositions to Modality*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198714316.001.0001
Vetter, B. (2019). Are Abilities Dispositions? *Synthese*, *196*(1), 201–220. https://doi.org/10.1007/s11229-016-1152-7
Watzl, S. (2017). *Structuring Mind. The Nature of Attention and How it Shapes Consciousness*. Oxford University Press.
Wedgwood, R. (2001). Sensing Values? *Philosophy and Phenomenological Research*, *63*(1).
West, D. E., & Anderson, M. (2016). *Consensus on Peirce's Concept of Habit: Before and Beyond Consciousness* (D. E. West & M. Anderson (eds.)). Springer International Publishing.
Williams, B. (1979). Internal and External Reasons. In R. Harrison (Ed.), *Rational Action*. Cambridge University Press.
Williamson, T. (1995). Is Knowing a State of Mind? *Mind*, *104*(415).
Wittgenstein, L. (1958). *The Blue and Brown Books*. Harper & Row.
Wittgenstein, L. (2009). *Philosophical Investigations* ((trans.) Peter Hacker & J. Schulte (eds.); 4th ed.). Wiley-Blackwell.
Wong, H. Y. (2015). On the Significance of Bodily Awareness for Bodily Action. *Philosophical Quarterly*, *65*(261).
Wong, H. Y. (2018). Embodied Agency. *Philosophy and Phenomenological Research*, *97*(3).
Wood, E. M. (2016). Class as Process and Relationship. In *Democracy Against Capitalism: Renewing Historical Materialism* (5th ed., pp. 76–107). London, NewYork: Verso.
Wright, C. (2002). (Anti-)Sceptics Simple and Subtle: G. E. Moore and John McDowell. *Philosophy and Phenomenological Research*, *65*(2).
Wright, C. (2008). Comment on John McDowell's "The Disjunctive Conception of Experience as Material for a Transcendental Argument." In A. Haddock & F. Macpherson (Eds.), *Disjunctivism: Perception, Action and Knowledge*. Oxford University Press.
Wu, W. (2011). Confronting Many-Many Problems: Attention and Agentive Control. *Noûs*, *45*(1), 50–76.
Wu, W. (2016). Experts and Deviants: The Story of Agentive Control. *Philosophy and Phenomenological Research*, *93*(1), 101–126. https://doi.org/10.1111/phpr.12170
Yablo, S. (1996). How in the World? *Philosophical Topics*, *24*(1).
Young, I. M. (1980). Throwing Like a Girl: A Phenomenology of Feminine Body Comportment Motility and Spatiality. *Human Studies*, *3*(1).