
Convention, Reflection, and Agency

Peter Rupert Faulconbridge

UCL

Doctoral thesis submitted for the degree of PhD in Philosophy.

December 2018

Declaration

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

Abstract

I develop a novel account of the ontology of conventions, rules, and norms, before applying this account in an investigation of their explanatory and practical roles. Conventions, rules, and norms belong to the metaphysical category I call ‘abstract individuals’. Conventions are multiply instantiable but are also individuated by reference to historical factors, namely precedents. These features are characteristic of abstract individuals in general, which therefore contrast with members of the categories of both pure universals and concrete particulars. In the first part of the thesis I elaborate on this metaphysical category, and show that conventions belong to it. I then defend the empirical adequacy of this view from certain arguments in the literature on cultural diffusion.

The second part demonstrates how this view can encompass ‘implicit’ norms, and norms governing non-intentional behaviour. I also show how conventions, considered as abstract individuals, can have practical significance for us.

This prepares the way for the final two chapters, which apply the ontology developed earlier to questions about the place of rules in practical reasoning. I argue that rules of various kinds are treated by deliberators, as well as those who would understand and evaluate their actions, as sources of ‘requirement’. Standard approaches to rule-based practical reasons fail to account for the phenomenon of requirement, in part due to an inadequate ontology of rules. My positive account develops on Michael Thompson’s discussion of ‘practical generality’. The metaphysical ‘generality’ of rules, understood as abstract individuals, is key to explaining the sense of requirement which attaches to them. However, I argue against certain assumptions in Thompson’s work, which limit his account to rarefied practices associated with justice, such as promising. Modifying these assumptions, I will show, allows us to employ my ontology of ordinary practices to understand the practical role of rules of all kinds.

Impact Statement

This thesis makes novel contributions to a number of established topics in contemporary analytic philosophy. These include the ontology of social practices, and the nature of practical reasoning. In addition, it shows how these two topics – which are often discussed separately – can shed light on one another.

It also brings social-scientific work from various disciplines to bear on these philosophical questions. It draws in particular on contemporary discussions of cultural diffusion and cultural evolution, and on various treatments of implicit norms and norms governing non-intentional action. The latter literature in particular does not have a well-established presence in contemporary philosophical discussions. It is my hope that this will open up new avenues for philosophy to engage with the relevant social-scientific literature on these topics.

Finally, it also contributes to ongoing debates outside philosophy. This is the case particularly in Chapter 2, where I address some of anthropologist Dan Sperber's arguments for nominalism about cultural entities. In addition, Chapter 4 argues that understanding the ontology of conventions is important for understanding their explanatory role. These ontological questions are therefore not only a matter of interest for philosophers.

To sum up: in addition to contributing to a range of philosophical topics, I hope that this thesis will also help to promote interdisciplinary engagement in both directions between philosophy and various social-scientific disciplines on the topics of convention and social practice.

Acknowledgements

I would like to acknowledge the financial support of the London Arts and Humanities Partnership (LAHP), and the Royal Institute of Philosophy (RIP). My doctoral research was supported by an AHRC Research Studentship through the LAHP from 2014-2017, and by a Jacobsen Bursary from the RIP in the 2017-2018 academic year. I am very grateful for this support. Thank you to each of these institutions.

I presented earlier versions of parts of the material in Chapters 6 and 7 at the Leipzig–UCL Graduate Philosophy Workshop (2017), the London Intercollegiate Graduate Philosophy Conference (Spring 2018), and the Future of Normativity Conference at the University of Kent (2018). Various parts of the thesis were also presented at the UCL Philosophy graduate work in progress group. I thank attendees at each event for their questions and comments.

I would like to thank the faculty and graduate students of the UCL Philosophy Department, from my MPhil years to the present day. The graduate community has been, for me, an incredibly supportive, generous, and rigorous environment in which to learn to do philosophy better. I thank everyone who contributed to this over the years. In addition, I recall conversations with many fellow students which directly shaped my thinking on the topics discussed in this thesis. They include: Showkat Ali, Julian Bacharach, Senthuran Bhuvanendra, Vanessa Carr, Catherine Dale, Ben Fardell, Alex Geddes, Alec Hinshelwood, Charles Jansen, Andrew Knox, Rowan Mellor, Polly Mitchell, Dave Olbrich, Edgar Phillips, Ashley Shaw, and Winnie Sung. I apologise for any accidental omissions from that list.

Special thanks are due to my supervisors, Doug Lavin, Mike Martin, and Lucy O'Brien. Thank you to Lucy for, during the first year of the PhD, helping me to find my way in a new area of philosophy, and pushing me to articulate a research topic

which started life very inchoate. The project has changed significantly since then, but I have constantly circled back to conversations we had during that year. Thanks, too, for a very helpful conversation right near the end of the project. Doug and Mike were my supervisors for the majority of the PhD, and it is difficult know how to capture their impact on the final product. Conversations with them have shaped almost everything I write here, always for the better. I thank them for this, and for their patience and generosity.

Finally, I would like to thank my parents for their emotional, material, and intellectual support. Counterfactuals are difficult to evaluate with an influence so fundamental, but in so many ways I would not have got to this point without them.

Table of Contents

Declaration.....	3
Abstract.....	4
Impact Statement.....	5
Acknowledgements.....	6
Table of Contents.....	8
List of Tables and Figures.....	12
Introduction.....	13
Chapter 1. Conventions as Abstract Individuals.....	26
The ‘folk sociology’ of convention.....	27
Problems accounting for the individuation of conventions.....	29
The outline of a solution to these problems.....	35
Repeatable artworks as created types.....	37
Some issues with the created types proposal.....	41
Abstract individuals: the basic picture.....	46
A paradigm for understanding abstract individuals.....	49
Reproduction in Millikan’s account of ‘natural conventions’.....	57
Convention and arbitrariness.....	62
Conclusion.....	67
Chapter 2. Reproduction and Imitation in Human Life and Culture.....	69
Outline of Sperber’s challenge.....	70

An overview of memetics.....	74
‘Triggering’ versus imitation in the spread of behaviours.....	76
Against imitation.....	80
Reproduction without imitation.....	84
Variation within types and nominalism.....	89
‘The song’ in oral traditions.....	93
Accommodating variation within the account of abstract individuals.....	95
Cultural context and reproduction from multiple models.....	101
Conclusion.....	104
Chapter 3. Reduction and Other Issues.....	107
Millikan’s ‘official’ account of reproduction.....	108
Biological purposes in the reproduction of conventions.....	110
The connection of biological purposes to the fitness of individuals.....	114
Problems for Millikan: cross-species conventions and memes.....	116
Naturalistic projects.....	120
Further questions: the lives of cultural forms, and their deaths.....	124
Conclusion.....	129
Chapter 4. Conventions, Agency, and Social Explanation.....	132
Memes, conventions, and agents.....	133
One task for social enquiry: explaining behavioural regularities.....	136
Explaining regularities by appeal to conventions.....	139
Non-intentional action, and norms governing it.....	144
Norms of interpersonal distance.....	148

How interpersonal distance norms get reproduced.....	155
Connections to broader projects.....	159
Conclusion.....	162
Chapter 5. The Practical Role of Conventions.....	164
The practical significance of non-intentional action.....	165
The ‘internal aspect’, revisited.....	172
Reflection and reform.....	175
Social phenomena and formal individualism.....	181
Conclusion.....	185
Chapter 6. Rules and the Sense of Requirement.....	187
Rules in practical reasoning and practical philosophy.....	188
The basic picture.....	190
The Priority Thesis.....	197
The instrumental construal of the Priority Thesis.....	198
Problems for the instrumental construal.....	204
Conclusion.....	206
Chapter 7. ‘Generality’ and the Rules of Ordinary Practices.....	208
An overview of the argument.....	209
Thompson’s argument from tight corners.....	211
Generality and requirement.....	217
Ordinary practices are not transparent.....	220
The apparent dichotomy between instrumental and ‘general’.....	221
The gap between the tight corner and transparency.....	223

Rigourism and requirement.....	226
Complete explanation and ‘natural goodness’	228
Conclusion.....	233
Conclusion.....	235
Bibliography.....	239

List of Tables and Figures

Table 1. Characterisation of abstract individuals in terms of multiple instantiability and the identity of indiscernibles. (p.48)

Table 2. Differences between implicit and explicit norms. (p.140)

Table 3. The difference between intentional and non-intentional action. (p.145)

Figures 1 & 2. A five-pointed star and a less easily recognisable design. (p.81)

Introduction

Rules, norms, and conventions shape a significant amount of human behaviour. Rules play a role in many activities and practices that philosophers talk about, such as games and legal systems. The role of rules in such activities is clear not just to philosophers but also to participants in these practices. But local norms also explain, often without our realising it, how close we stand to talk to each other, where we allocate our attention in a social interaction, and why certain elevator rides are so uncomfortable. Such norms are largely neglected in analytic philosophy.

It will turn out to be important that norms and conventions can shape our behaviour without us recognising their existence. However, to begin to describe the philosophical puzzles raised by rules, norms, and conventions, I will start by outlining certain features of the practical thought of agents who follow rules reflectively. Various forms of alienation from social practices or institutions and their rules are possible, but I will focus on the perspective of the non-alienated rule-follower. The non-alienated rule-follower, finding herself in a situation to which a rule applies, might ask the following questions: What is the rule here? Is it a valid one? Do I fall under its jurisdiction? If she finds that the rule is valid and applies to her, then she takes the rule as providing a reason to do what the rule requires.

It is always possible that she will decide on balance that she ought to break the rule. This is an important possibility, although spelling out what it involves requires some nuance. (This is discussed at greater length in Chapter 7.) Rather than treating the rule as always a decisive reason to do whatever it requires, the non-alienated rule-follower treats this as something more like a default case. Special considerations might always make a difference. The central points for present purposes are that the rule-follower asks which rules are in place, and as a default treats such rules as, in each case to which they apply, direct sources of reasons to do what the rules require.

Rules therefore present at least two faces to us, each of which raises philosophical questions. The first is that of real entities in our social worlds. It is an integral part of the rule-follower's perspective that the rules that bind her are the *actual* rules applying to her situation. A merely hypothetical rule, or an actual rule which applies in a merely similar situation (as different universities may have different rules concerning attendance) makes no claim on her. The second is the distinctive role they occupy in the practical thought of rule-followers, where they figure as direct sources of default reasons in each case to which they apply. I suggest that these two points are common between rules, norms, and conventions. (I will elaborate on the relationships between these terms below.)

It seems, then, that an account of rules should try to illuminate both of these faces. But much of the literature focuses on one to the exclusion of the other. So, for instance, there is detailed work in a broadly rational-choice framework, following David Lewis' work on conventions, which has a lot to say about the conditions under which one rule, norm, or convention rather than another is in force in a given social context (e.g. Lewis 1969; Pettit 1990; Bicchieri 2005; 2017). Such approaches typically centre around the interdependence of distinct agents' actions in a given kind of situation. An example is Lewis' discussion of which side of the road to drive on, when one has to share it with other motorists. This serves the valuable function of illuminating some of the reasons why it can be important to have shared rules, norms, or conventions in that kind of situation.

However, the Lewisian approach fails to explain the distinctive role that such entities can play in the practical thought of the non-alienated follower of conventions or rules. The central, recurring, problem for these approaches is that they make our reasons for conforming to rules, norms, and conventions dependent on the pursuit of independently valuable goods (Lewis) or on the allocation of approval and disapproval which has grown up around the norm (Pettit and Bicchieri). But then in cases where such goods are not forthcoming (say, the agent can break the rule without detection, and so without risk of disapproval) the agent has no reason to follow the rule. The problem here is not just that this allows that it will sometimes be right to break the rules (we have seen that this is an important possibility), but rather that even in those cases where the rule binds, it does so in virtue of the prospect of

particular goods attaching to each case individually. There is no real normative work for the rule itself to do – no sense of how it could generate a ‘default’ reason to conform to it.

A more diffuse set of works places the emphasis the other way around. There are a number of philosophers who pay greater attention to the distinctive practical role of rules, but who end up separating this role from an ontology of actual rules and social practices. This is perhaps clearest with something like T. M. Scanlon’s ‘Principle of Established Practices’. Scanlon says that this Principle can give us reasons to follow the rules of certain practices which are broadly established around us. (The Principle of Established Practices is not itself a principle or rule of any actual practice – I will use the capitalised ‘Principle’ to refer to these Scanlonian entities which do not belong to an actual practice.) Given the character of this Principle, it can explain part of the normative profile of such rules – why the rules of established practices might generate the right kind of default reason to go along with them. However, Scanlon says very little about what ‘established practices’ *are*. And, in fact, he leaves it unclear exactly why his Principle should concern *practices* at all.

Scanlon’s argument for the legitimacy of his Principle rests on the consideration that other Principles may be, in a certain sense, indeterminate. His key example concerns privacy. “People need to be able to conduct parts of their lives protected from the scrutiny of others whom they have not chosen to admit, and people generally need to have some forms of private communication”, but these needs could be met through a wide variety of different practices (e.g. different areas of another’s life which one is forbidden to intrude upon) (Scanlon 339). Yet at the same time, these needs *cannot* be met in a given context without *some* set of relatively determinate rules being in place in that context. Privacy can only be achieved if there is broad agreement on *which* areas of life are private. The need for privacy thus generates a need for some more determinate set of rules or practices. The thought behind Scanlon’s Principle of Established Practices is, then, that if some privacy regime is already widely observed in a given context, the need for some such determinate regime generates reasons to go along with that local practice.

One problem with this account, from the present point of view, is that it gets the practical role of actually-existing ‘established practices’ wrong. This is obscured by Scanlon’s way of talking about actual practices (as opposed to Principles). He writes as though the existence of a practice consists in a group of people endorsing some rule or principle. When such practices are legitimated by the Principle of Established Practices, it will therefore appear that people are getting things *right* in endorsing the principles they do, with the qualification that their endorsement is playing a role in the fact that the Principle binds them in the way it does.

However, it is quite unclear why a practice would need to involve anything like the acceptance of a set of principles by a population, if the justification for the Principle of Established Practices is simply the need for determinacy in the Principles that bind a population. Other kinds of practices – or perhaps even habits – could, for all that’s been said, provide the requisite determinacy. This point is important, although not central for the present argument. It does, however, highlight a crucial further point. If practices did not consist in the endorsement of principles, then the normative situation for a group to whom Scanlon’s Principle applied, as he understands it, would not resemble the situation as they perceived it themselves. But in fact this will be the case even for people whose practices do consist in the endorsement of principles.

This is because, as we have already seen, people take their actual, earthly practices to be direct sources of practical reasons for them. The principles they endorse (when these are the principles of a social practice, rather than the principles of a personal morality) belong to a specific social practice, and exist only as part of that practice. They are not Scanlonian Principles. So, even if Scanlon’s Principle of Established Practices works by generating further Principles which correspond to the principles of the established practices, these Scanlonian Principles will be distinct entities from the principles of the practice. His account therefore seems to imply that agents are mistaken when they take the rules or principles of their practices as direct sources of reasons for them. The best case scenario is that the Scanlonian Principles are isomorphic to the principles of actual practices – non-accidentally sharing the same content.

A similar problem faces Raz's rather different account of rules in *Practical Reason and Norms* (1999 Ch.2). It seems that for Raz, to talk of the existence of 'a rule' mandating that agents do A in situations of kind S , is to imply that for each situation of kind S , there are reasons which support doing A in that situation. The reasons here may be complex. They will include certain second-order 'exclusionary reasons' *not* to act for some range of first-order reasons, along with some first-order reason(s) to act in the way required by the rule. This structure of reasons allows Raz to capture much of the practical role of rules from the perspective of the agent, and from the perspective of those who would understand and evaluate her action (as I will argue in Chapter 6). However, on this picture the rule itself is just, at most, a collection of such reasons, applying to a variety of situations which we would normally describe as being covered by the rule. An upshot of this is that there is, once again, an important kind of gap between any rule or practice which might have existence in a given social context on the one hand, and the Razian 'rule' on the other.

I will explain. On Raz's picture, the existence of a practice in the social world may be a circumstance which generates a corresponding rule in his sense. To take a slightly different case, I may have reasons to follow a given rule of thumb that I don't have to follow some equally effective alternative rule of thumb, simply because I am only aware of the former. I can use a rule of thumb to save the time and effort required to treat each case on its merits only if I am already aware of that rule. However, if the rule of thumb is a Razian 'rule' for me, it will only be so because *in each case* which the formula I have memorised purports to cover, I in fact have a reason to do what it says rather than to consider the case on its own merits. It is consistent with this that there might be 'rules' in Raz's sense which do not correspond to *any* actually existing practice. This also means that a practice itself is never the direct source of the kinds of reasons distinctive of rules, for an agent. Rather, its existence is at most a circumstance which contributes to the existence of a Razian rule. (Just as, in the case of the rule of thumb, the Razian rule is not the same thing as the formula that I have memorised, even if in some sense they have the same content.)

Despite getting certain things right about the practical profile of rules, both Scanlon's and Raz's accounts suffer from the absence of a proper account of what real

practices and rules are.¹ It might be thought that the lack could be made up for by looking to the social ontology provided by the rational choice accounts mentioned above. However, I don't think that this route is a promising one. Lewis, Pettit, and Bicchieri spend a lot of time on the conditions for there to be a rule or convention in place, but as we will see in Chapter 1 they say almost nothing about what rules or conventions themselves *are*. This ontological question will turn out to be of some importance.

My approach to the topic in this thesis is guided by the thought that in order to understand the practical role of rules, norms, and conventions, we need to have a clearer picture of what kinds of things they *are*. Before beginning the summary of my argument, it will be helpful to clarify some of the key terms.

So far I have largely been discussing the philosophical literature on rules, but I began this introduction by talking about 'rules, norms, and conventions', and these three terms often occur together in this thesis. How do they relate to one another? In short, my view is that each refers to basically the same kind of thing. To anticipate the results of Chapter 1, each term has a prominent usage in which it refers to what I will call an 'abstract individual', generated through the reproduction of patterns of behaviour. So, much of what I say about either rules, norms, or conventions could be said about all of them. I was tempted to refer to them all together as 'social practices', for stylistic simplicity. I have occasionally used 'practices' in this way here, but I refrained from doing so in general because it seems to me that 'practice' also has a different use, where for instance it might refer to a more structured entity to which a number of distinct rules *belong*. A practice in that sense might not be directly reproduced in the actions of those who follow it, and as we will see there are good

1 This is not quite fair. Each has an argument elsewhere *against* various forms of practice account. Nevertheless, I understand each of them to be trying subsequently to capture what they take to be the true role for practices in generating reasons, by giving them a more indirect role. My point here is that even if their accounts are successful as it were 'extensionally' in capturing the work that practices do, they get something wrong from the perspective of the non-alienated rule-follower precisely by making the role of practices indirect.

reasons for focusing in the first place on practices which are reproduced in this way. Hence the focus on rules, norms, and conventions.

I also considered choosing one term of the three and using that throughout. Ultimately, I do not think it would be entirely inaccurate to refer to all of them as ‘norms’. However, there are certain distinctions which it is useful to mark amongst such entities which would then be more difficult to express. For instance, I generally reserve the term ‘rule’ for norms which the agent characteristically knows she is following. This in turn is responsible, as I will argue in Chapter 7, for rules being associated with a distinctive sense of ‘requirement’. In Chapters 1-3, in which I develop my ontological account, I use ‘convention’, because this is the term used most often in the literature I am engaging with in those chapters. In Chapter 4 I will use ‘norm’ more frequently, because there I will be concerned with the distinction between implicit and explicit norms, and ‘implicit norm’ sounds more natural to me than ‘implicit convention’. On the whole, though, ‘norm’ and ‘convention’ as I use them are largely interchangeable.

I acknowledge that there might be further distinctions within the family I describe, which would be naturally marked through a differential use of these two terms. Some have wanted to reserve the term ‘norm’ for entities which are like conventions, but which are perpetuated partly in virtue of a sense of disapproval which has grown up around behaviour which contravenes the convention. There might be an important distinction here, but I do not mark it linguistically in this way. I think it is important here to grasp the common phenomenon at the core of each notion, before trying to subdivide it in a perspicuous way. Finally, some might think that a term like ‘convention’ belongs *only* in social ontology, whereas ‘rule’ has an essentially ‘normative’ meaning. Again, this usage might be useful for some purposes, but it is not my usage. As I have indicated in setting up the puzzles around rules, I use each of the terms ‘rule’, ‘norm’, and ‘convention’ to name something having actual existence in the social world. This is *not*, however, to say that they therefore have no direct normative or practical role for us. The assumption that those two things are incompatible is a substantive, rather than linguistic, one, and it is one that I attempt to undermine in the arguments that follow.

I will now outline the main arguments of the following chapters. I develop my ontological account of rules, norms, and conventions in the first part of the thesis. I then apply the results of this to practical and explanatory questions in the later parts. But I start the former task by considering these entities from a perspective which is relatively removed from that of the reflective rule-follower. In Chapters 1-3 I focus primarily on conventions as they can figure in the attempts by both social scientists and people in everyday life to *explain* behaviour.

Chapter 1 argues that our ordinary way of individuating conventions supports thinking of them as belonging to a category I call ‘abstract individuals’. Conventions are multiply-instantiable, but they are also individuated by reference to historical factors – in particular, by reference to historical precedents, and chains of precedent. In combining these two features, conventions contrast with the way in which both universals and concrete particulars are usually understood. And this combination of features defines the category of abstract individuals, as I will describe it. To illuminate these two features, I draw on discussions of the category of ‘created types’ in the ontology of art, a category which shares much in common with that of abstract individuals. However, although the work in the ontology of art is useful here, I will argue that the category of created types is, as things stand, importantly unclear. Whilst the category of abstract individuals is a broader one than that of created types, my account of the former allows for a clearer focus on the features of conventions which are important for my purposes.

I suggest that a key paradigm for understanding the possibility of abstract individuals is the paradigm of *chains of reproduction* of concrete particulars. One key set of examples is found in the case of ‘repeatable artworks’ such as novels and symphonies, and other kinds of artefact. This paradigm can also be observed in Ruth Millikan’s account of ‘natural conventions’, according to which conventions are patterns of behaviour reproduced from one context to another, partly due to the weight of precedent. Millikan does not say much about the ontology of conventions, but I will argue that her account speaks in favour of considering them as abstract individuals. Understanding abstract individuals through the concept of reproduction will also help us to see how they might *explain* the occurrence of their instances, which is another important feature of conventions, as we usually understand them.

Chapter 2 addresses a challenge to the view of conventions (and artworks) as abstract individuals generated through chains of reproduction. This challenge comes from recent work on cultural diffusion in cognitive anthropology and evolutionary approaches to culture. I focus in particular on Dan Sperber's influential critique of memetics, and his broader arguments for a kind of nominalism about cultural entities. Sperber's nominalism is clearly at odds with the proposal that conventions and repeatable artworks are abstract individuals. Moreover, his arguments hinge on the claim that, contrary to what is sometimes thought, the processes which underpin the spread of behaviours through human populations are not 'replicative'. I argue that we can, however, accept many of Sperber's empirical claims about cultural diffusion, consistently with the account of conventions developed in Chapter 1.

In Chapter 3 I raise disagreements with Millikan's reductionism. I also reflect on the *kind* of project that seems to be common to Millikan, Sperber, and the memeticists, and on how this project contrasts with my own. Finally, I suggest some issues for future work on the ontology of conventions and artefact kinds, which are raised by my account.

Chapter 4 applies the ontology developed in the first three chapters to questions of social explanation. According to the view that conventions, rules, and norms are abstract individuals, they are entities created through our actions, but nevertheless exist 'externally' to us. This gives rise to a question of how exactly they relate to human agency. If we explain an event by appeal to a convention, understood as an abstract individual, do we leave room for the existence of individual human agents in addition? I argue that we do, and show in greater detail how conventions can play a role in social explanation. Along the way, I will show how my account of the ontology of conventions can accommodate the existence of 'implicit' norms, and norms governing non-intentional behaviour. I do this in part through a case study of one set of norms which appears both to be implicit and to govern non-intentional action: the norms governing interpersonal distances in social contexts.

Chapter 5 defends the suggestion that conventions as I have described them can have *practical significance* for us. This is an important assumption of the final two chapters, but it may be thrown into doubt particularly by my inclusion in Chapter 4

of implicit norms governing non-intentional action. Can we intelligibly hold non-intentional actions up to socially-determined standards of ‘correctness’? I will argue that we can. I also discuss what could be said about the ‘internal aspect’ of implicit norms, and how reflection on a formerly implicit norm might transform it. This then prepares the way to apply my ontology of rules, norms, and conventions to questions about their practical role for us in the subsequent chapters.

The final two chapters return to the case of explicit, reflectively-endorsed rules. I argue that non-alienated rule-followers typically treat the rules of the practices they fall under as sources of what I call ‘requirement’. Requirement is a distinctive kind of normative phenomenon. As sources of requirement, rules contrast with other practically significant phenomena such as an agent’s ‘ends’. In particular, if an agent is ‘required’ to perform a given action, it seems that her deliberation is more restricted than when she is acting in pursuit of an end. In Chapter 6 I argue that some popular attempts to understand the practical role of rules in terms of ends plus a minimal social ontology of ‘constitutive rules’ fail to account for the sense of requirement.

Chapter 7 argues that the distinction between the normative profiles of rules and of ends can be explained in part by appeal to a metaphysical distinction between the generality of rules and the particularity of ends. This draws upon the ontology of rules, norms, and conventions developed in the first part of the thesis. It also develops on Michael Thompson’s discussion of ‘practical generality’ in *Life and Action*. However, Thompson’s own account of social practices faces certain important limitations. It applies only to those rarefied practices associated with justice, such as the practice of promising.

The problem here is not the one faced by Scanlon and Raz above. It is a central part of Thompson’s account that the practices in question have actual existence in a social context. Furthermore, the practice which the agent responds to and the practice that explains and justifies her action are one and the same – the practice is a direct source of practical reasons for her. The problem is rather that Thompson’s argument suggests that a practice can only be the direct source of practical reasons for an agent if following that practice always yields a good action – practices must be ‘transparent’, in his terms. This is quite implausible for almost any real practice, such

as the ones described in the rest of the thesis. However, I will show that Thompson's argument for the view that only transparent practices can be direct sources of reasons does not work. It is possible for the rules of a practice to bind an agent, even whilst it is sometimes right to break those rules.

This last remark suggests an important point which I do not discuss in depth in the body of the thesis, but which is present in the background. It seems clear that there are and have been a great number of bad practices in human history. I am not referring to the merely sub-optimal, but rather to the unjust, oppressive, or fundamentally misguided. It is hard to believe that such practices could generate *any* reasons for us to go along with them, let alone the kinds of reasons characteristic of 'requirement'. This thought is, I take it, what supplies some of the motivation for moves like Scanlon's. It can also seem to take the sting out of my objection that actual human practices figure only indirectly as sources of reasons in his account. However, Scanlon's move is an over-correction in response to the existence of bad practices. We can acknowledge the point that bad practices do not bind us without saying that the only things that directly bind us are ethereal Scanlonian Principles. We can simply say that actual practices can be good or bad, and that bad ones don't bind.

There are, then, two important elements in an accurate picture of how practices can bind us. One is the goodness of the practice, and the other is its actual existence in the relevant social context. The distinctive practical profile of rules is largely explained by the combination of these two elements. The Scanlonian approach seemed to get the second part wrong. But it is equally important not to overlook the first part.

This thought raises the question of how practices, including rules, norms, and conventions, might be assessed. What is it for one of them to be good or bad, in a given context? The only answers I provide to this question in the thesis are negative. The goodness of a rule does not consist just in its accordance with the prior ends of those who follow the rule, for instance.

Taking social practices seriously in philosophy seems to me to be important, because they are ubiquitous in our ordinary practical lives. Almost everyone will have at least some practices which they take to be good, and to make legitimate claims on their

behaviour. However, as we have noted, some practices are so unpleasant that at times it can feel as though the idea that any practice could ever place legitimate requirements on us should be jettisoned. Personally, both of these thoughts have carried great weight with me as I have written this thesis. My hope is that the latter concern can be mitigated by an (eventual) positive account of the evaluation of practices. In the meantime, the arguments I develop below show that a lot can be achieved without having a specific positive account of the justification of practices in place.

There are, however, at least two significant considerations that cannot be effectively held off by a promissory note for a satisfactory account of the evaluation of practices. One is the consideration that what is good can be contested, and contestation will almost inevitably attach to any actual practices. Saying that only good practices can bind us will provide little guidance for cases where goodness is contested, or where the best that can be achieved is an awkward balance among different claims and voices. The second consideration is how difficult it can be to hit upon good arrangements in the first place. The worry here is less that we might end up being bound to actively bad practices. Perhaps we can recognise those when we see them, though overturning them may still be difficult. The suspicion is rather that if we treat *any* way of doing things as ‘binding’, or as a source of ‘requirement’, then this will prevent us from ever discovering *better* ways to follow instead.²

I say little to directly address these two points in the following chapters, although I will touch on them briefly at the end of Chapter 5. However, perhaps surprisingly, they are less relevant than it might seem to what I describe as the sense of requirement attached to rules. It is important that bad rules do not bind, if rules are to be sources of requirement for us. However, the specific suggestion that we treat

2 Karl Popper’s arguments against ‘Utopian engineering’ in *The Open Society and Its Enemies* come to mind here (Popper 1945 esp. Ch.9). However, it is worth bearing in mind his claim elsewhere that the effective critique of tradition is itself made possible by what he calls a ‘second-order tradition’ of conjecture and refutation (Popper 1963). These points raise interesting and important questions which unfortunately I cannot explore further here.

good rules as sources of *requirement* does not seem to make the reality of inevitable contestation any easier or more difficult to deal with.

At first glance, treating rules as sources of requirement has a greater relevance to the prospect of cumulative reform of existing practices or institutions. However, I do not think that treating good practices as sources of requirement is incompatible with the kind of critique and modification that such reform depends on. For all that (as I will argue in Chapter 7) requirement depends on the *generality* of rules, it is still a phenomenon which shows up in reference to particular decisions and particular actions. Treating a rule as a source of requirement in such instances is therefore compatible with stepping back and assessing the rule itself from a different perspective in other contexts. So, although these two points are very important for a broader account of the role of rules in life and in politics, once again it is my hope that they can be largely bracketed for present purposes. In fact, by allowing for a clearer understanding of the distinction between justification of a rule, and justification of an action falling under the rule, the view I develop here will provide important resources needed for a proper account of the critique of practices in the future.

Chapter 1. Conventions as Abstract Individuals

Conventions have received a lot of attention in analytic philosophy, in connection with language and with social explanation. However, the focus has often been on the rational status and the dynamics of conventions, rather than on their ontological status. Indeed, the major works on convention in philosophy say very little at all about what conventions *are*. Further, what suggestions there are concerning the ontological status of conventions in the literature are at odds with how we generally individuate them in ordinary life.

Getting the ontology of conventions right is important for its own sake, but as will become apparent in later chapters it will also help us to understand broader questions about conventions, rules, and norms. In particular, it will help us to understand their explanatory and normative roles. In this chapter I offer an account of their ontology. I start by outlining some basic claims about the individuation of conventions, which are implicit in much ordinary and social-scientific discourse. I then argue that the metaphysical suggestions in work by David Lewis, Philip Pettit, and Cristina Bicchieri cannot do justice to these claims about individuation.

I then develop my positive account. The category of ‘created types’ is often invoked in the literature on the metaphysics of ‘repeatable artworks’ (such as novels and symphonies). There are interesting parallels between this picture of repeatable artworks and our initial claims about conventions. I will draw from the literature on the metaphysics of artworks to help make sense of the ontological status of conventions. However, I ultimately argue that the category of ‘created types’ as it has been developed so far is importantly unclear.

As an alternative, I offer an account of the metaphysical category of ‘abstract individuals’. This category is plausibly a broader genus of which created types would

be a species. I will describe it in such a way as to bring out more clearly the relationship between the basic metaphysical features of abstract individuals, and our initial claims about the individuation of conventions. This relationship in itself provides some support for the idea that conventions might be a kind of abstract individual. Finally, I will provide a more positive reason to think that conventions are abstract individuals. ‘Reproduction’ is a key concept for understanding the nature of abstract individuals. There is a natural fit between this claim, and Ruth Millikan’s claim that conventions are generated through the reproduction of patterns of behaviour. This fit speaks in favour of treating conventions as abstract individuals.

The ‘folk sociology’ of convention

There are distinctive forms of practical thought associated with *following* rules or *participating in* established social practices, which are often remarked upon. But we also have what could be called a ‘folk sociological’ way of thinking about rules, norms, conventions, and practices. This way of thinking is most obvious in relation to practices the thinker does not identify with. However, *contra* R. M. Hare, for instance, I doubt that there is a neat division amongst thoughts about practices between practical, engaged, or committal thought on the one hand and purely empirical, theoretical, or anthropological thought on the other (1964 411). To say that there is a folk sociological notion of practice is at least to say that in everyday life we *can* make sense of the claim that “this is their convention (or rule, or practice)” without endorsing the convention, taking ourselves to be bound by it, or thinking that ‘they’ are making a mistake.

In this section I will outline some features of what I suggest is the folk sociological concept of a convention, focusing on the individuation of conventions, and the circumstances under which we consider particular actions to belong to a given convention. I will speak mainly of conventions here, but these basic points about individuation are shared also with rules and norms. The dominant approaches to understanding conventions in analytic philosophy fail to account for these claims about individuation. This chapter develops an alternative account of the ontology of

conventions which does better justice to our folk sociological understanding of them.

What can we say about the folk sociological concept of convention? Ruth Millikan tells us that there is a conventional way to punt in Cambridge, which differs from the conventional way to punt in Oxford.³ In Cambridge “it is conventional to punt standing on the deck at the back of the punt boat, whereas in Oxford, one turns the deck to the front and stands at the back inside” (2005a 3). Tourists often happen to hit on the Cambridge way of punting in Cambridge, simply because it is one of a number of obvious ways to try to get a punt moving, but such a tourist will not necessarily count as “*following* the tradition”, nor will her behaviour necessarily count as ‘conventional’ (3-4). Producing behaviour which is seemingly qualitatively identical to a piece of behaviour which manifests a convention is not sufficient for manifesting that convention. This seems intuitively correct. By contrast, if the tourist punted in the Cambridge way in Cambridge because she had been told to, or because she was copying the actions of those around her who were aware of the convention, then her behaviour might well count as conventional.

We can go a little further. In another town, a punting convention just like Cambridge’s might have developed independently. When somebody who grew up in this town follows the local convention, she will produce behaviour qualitatively identical (in the relevant respects) to behaviour which manifests the Cambridge convention, and she will also be acting conventionally in doing so. (In this she is unlike the first tourist, who was not following *any* convention.) Yet, if the two conventions are distinct, her behaviour will manifest her town’s convention but not the Cambridge convention.

Intuitively, it seems that ‘history’ is playing a crucial role here: for a given piece of behaviour to manifest a particular convention requires it to have the right kind of

3 A post on the King’s College Cambridge website claims that “[d]evotees from both Oxford and Cambridge believe that theirs is the only true style” (<http://www.kings.cam.ac.uk/archive-centre/archive-month/june-2011.html>). This may or may not be hyperbole in the current case, but it is worth noting that this sort of belief is an important aspect of some conventions, and not others. Such beliefs raise interesting questions which I won’t address here.

historical connection to that convention, and conventions themselves will be individuated partly by appeal to historical factors. (If our third town's convention had historical links with the Cambridge convention, it would be less clear that they are now distinct conventions.)

In the rest of this chapter I will be concerned primarily with the points about individuation just described. However, there are further important features of our uses of the concept of convention which I also aim to account for. Both in everyday life and in much social-scientific work, it is assumed that conventions have some explanatory power. In particular, pointing to a given convention can help to explain the occurrence of its instances. So, for example, when we point out that someone is punting in the conventional way in Cambridge, this is usually understood to be offering some explanation of why she is punting in the way that she is. The same is true of the concepts of *rule* and *norm*, which are discussed at greater length in subsequent chapters.

Finally, conventions, along with rules and norms, are often treated by those who fall under them as sources of practical reasons. This thesis aims to shed light on each of these three features of conventions: individuation, explanation, and practical role. This chapter focuses on the question of individuation, along with a key point concerning explanation. Chapter 4 will elaborate on the explanatory role of conventions. Chapters 5-7 discuss their practical role.

Problems accounting for the individuation of conventions

The points about individuation outlined above have some claim to be taken seriously by a philosophical account of conventions. Taking them seriously is my task in this chapter. However, looking first at how certain prominent accounts of convention struggle or fail to accommodate these claims about individuation will help to situate my positive account. The problem stems from a general neglect of the ontology of conventions. This is partly explained by a focus on the rational status of conventions, and on their dynamics, in much of the literature. However, in the rest of this thesis I

will show that the ontological status of conventions, norms, and rules is in fact important for understanding their explanatory and justificatory power.

David Lewis' account of conventions is perhaps the most influential in analytic philosophy. I begin by outlining in some detail the problems his account faces in doing justice to our claims about individuation. I will then show that a very similar problem faces two more recent accounts which develop certain themes of Lewis'.

What *is* a convention, on Lewis' account? He gives the following conditions:⁴

A regularity R in the behavior of members of a population P when they are agents in a recurrent situation S is a *convention* if and only if it is true that, and it is common knowledge in P that, in almost any instance of S among members of P ,

- (1) almost everyone conforms to R ;
- (2) almost everyone expects almost everyone to conform to R ;
- (3) almost everyone has approximately the same preferences regarding all possible combinations of actions;
- (4) almost everyone prefers that any one more conform to R , on condition that almost everyone conform to R ;
- (5) almost everyone would prefer that any one more conform to R' , on condition that almost everyone conform to R' ,

Where R' is some possible regularity in the behavior of members of P in S , such that almost no-one in any instance of S among members of P could conform both to R' and to R . (1969 78)

Lewis appears here to be identifying a convention with a regularity in the behaviour of some group of people. What sort of thing is a 'regularity', though? It is fairly clear what Lewis means by saying that *there is* a particular regularity in the behaviour of some group. It means that their behaviour regularly instantiates a given pattern.

4 Lewis provides a number of formulations of his definition, both in *Convention* (1969), and 'Languages and Language' (1983). The differences between them are important for some purposes, but they are all the same in the respect which concerns me here.

use one of his examples, if people in P regularly drive on the right-hand side of the road, then there is a regularity among them of driving on the right. It is less clear exactly what the regularity itself, if there exists such a thing, *is*. But if the regularity is to be identical with the convention, then we need an answer to this question.

One answer which is suggested by the above gloss is that a regularity in behaviour among P just is a *pattern* of behaviour, which has the property of being regularly manifested by members of P . The pattern would be something like *driving on the right*, *wearing black to funerals*, or *punting from the deck at the back*. Such a pattern could be instantiated by indefinitely many distinct actions, and could be considered as something like a universal. This proposal also makes sense of condition (5): R' is also a pattern of behaviour; it is difficult to instantiate both R and R' in the same action; but it is true that people would prefer to instantiate R' (in the relevant circumstances) if R' were regularly instantiated instead of R . Considered as a universal, R' exists, but is only possibly regularly instantiated among P , rather than actually so. Hence in Lewis' terms it is a "possible regularity".

However, there are problems in identifying a convention with a pattern of behaviour, considered as a universal. In particular, doing so appears to undermine the folk sociological claims about the individuation of conventions with which we began. Consider, we said that it was possible for there to be two qualitatively identical but numerically distinct conventions. Our example above was of the Cambridge convention of punting from the deck at the back, and the convention of a hypothetical city where they do the same. Punting conventions do not satisfy all of Lewis' conditions for a convention, though, so let's look instead at one of his own examples – that of conventions governing which side of the road to drive on.

Imagine that there are two different cities, A and B , where it is conventional to drive on the right-hand side of the road, but there are no historical links between these two conventions. According to the picture sketched in the previous section, these are two distinct but qualitatively identical conventions. However, what we mean by saying that they are qualitatively identical is that in each city it is *the same pattern of behaviour* that is conventional. If we were filling out Lewis' definition for these two conventions, we would put the same description in for " R ". The upshot would be

that the two cities' conventions are both identical with the same universal; but if that is the case then A 's convention = B 's convention. This contradicts our claim that they are two numerically distinct conventions.

There are a couple of fairly obvious moves we can make on Lewis' behalf in response to this problem. One is to lean on his talk of 'populations'. Although *driving on the right* may be regularly instantiated in A and in B , the rest of Lewis' conditions are unlikely to hold of the combined populations of these cities: many of them probably have no expectations about how those in the other city drive (condition (2)), nor do they especially care (condition (4)).

It seems likely, however, that for any given regularity R and population P that fulfils Lewis' criteria, there will be an indefinite number of subsets of P which also fulfil them. So there remains a risk of *over-generation* of conventions. It seems desirable to avoid this, because it looks as though there is just one driving convention in A , rather than a huge multiplicity. (Or, given that the case is under-specified in this respect: it could easily be that there is only one driving convention in A .)

Again there is an obvious move here: let's look at the *largest* P s for which all of Lewis' conditions hold. This suggests the following. Perhaps we have a convention when there is some regularity R which fulfils all of Lewis' criteria in a population P , where there is no proper superset of P that also fulfils the criteria for R . This promises to give us a convention for A , and a convention for B , but no convention for A -and- B together, just as we wanted. At the same time, it avoids populating either A or B with an indefinite number of driving conventions.⁵

We are not out of the woods yet, though. It is possible that a convention might die out among a population, and a qualitatively identical convention might arise among

5 There are other possible problems with this suggestion – perhaps we see a single convention in some situations where no-one in a population has expectations about how *everyone* in the population will behave. (Perhaps everyone has expectations only about their immediate neighbours, but given that everyone has slightly different neighbours, a stable pattern is sustained across the larger group.) However, such problems appear to attach first and foremost to Lewis' substantive conditions, rather than his general ontological commitments.

the same population at a later date. I would suggest that these are (or at least could be, depending on how the latter arose) numerically distinct conventions. Yet we have the same pattern of behaviour being conventional among the same population, just at different times. Again there seems to be a clear next step. To avoid i) merging the two, and ii) the kind of over-generation we saw above, we could say that there is a convention of doing R over the longest continuous stretch of time for which all of Lewis' conditions hold of R , for some maximal population P .

We have been treating Lewis' conditions as telling us when it is conventional among some set of people to do some thing R in some kind of situation S . We noted that such conditions hold in particular populations over particular stretches of time, and we are using these as the boundaries of 'the convention' itself, in the hope that this will respect the intuitive individuation we began with.

The problem with this is that it still doesn't seem to work. It seems logically possible (though empirically unlikely) that all of Lewis' conditions could hold among a population, continuously over a given stretch of time during which there was intuitively more than one numerically distinct convention among that population. Let's say that there are only two car owners on a small island. From t_1 to t_2 they drive on the right. At t_2 , a trickster erases all their memories of having driven on the right, at the same time as telling each of them: your neighbour drives on the right.⁶ (We should stipulate that the trickster doesn't say this because it has been true up until t_2 ; perhaps she flipped a coin, planning to say 'right' if landed heads, 'left' if tails, and it landed heads.) Wanting to coordinate their driving with their neighbour's, each drives on the right until t_3 . From t_1 - t_3 all of Lewis' conditions hold for R : *driving on the right*, and P : that population of two people. But I would suggest that in this scenario there are two distinct driving conventions, one which exists from t_1 - t_2 , and the other from t_2 - t_3 .

It might be objected to this that *driving on the right* is not a *regularity* among P immediately after t_2 , until they have resumed driving on the right for some significant amount of time. Therefore there will necessarily be a gap between the times at which

⁶ Perhaps we should complicate the trickster's utterance, to secure the conditions for Lewisian common knowledge, but this shouldn't affect my point here.

Lewis' conditions hold. However, if a regularity is just a pattern which is regularly instantiated among a population over a period of time, it's not clear that R fails to be a regularity at or just after t_2 , even if we discount all the driving on the right that happened up to t_2 . Consider: if one year I take up cycling on January 1st, and cycle every day of that year, then I cycle regularly during that year. My regular cycling started on 1st January, with my first ride. Assuming that facts are timeless, it was already a fact on 1st January (and indeed before that) that I cycled regularly during that year, even though at that point only one ride had taken place, and even though it might have been rather optimistic at that point to think that I would stick to my resolution. If this is correct then we have not yet found a way to respect our intuitive individuation of conventions in certain cases where Lewis' conditions hold continuously over time. His account appears to say that there is just one convention between t_1 and t_3 , but we want to say that there is more than one during this time.

Ontologically, the situation is quite similar for more recent accounts of conventions and norms that have drawn on Lewis' work. So, for instance, Philip Pettit's definition of norms begins with basically the same clause as Lewis': "A regularity, R , in the behavior of members a population, P , when they are agents in a recurrent situation S , is a *norm* if and only if [...]" (1990 731).

Cristina Bicchieri describes R as a 'behavioral rule', and gives conditions under which such a rule will be a 'social norm' (2005 11), a 'descriptive norm' (31-32), or a 'convention' (38). Interestingly, her characterisation of each includes a condition that members of the relevant population know "that a rule R exists and applies to situations of type S " (11). She does not, as far as I can see, say what a behavioural rule is, or what it is for one to exist. My best guess is that a rule is a verbal formula such as, "Drive on the right on public roads", although this does not leave much room for a substantive question about whether such a rule 'exists' or not. (That is, given that the question of existence must, to play its role in her account, be distinct from the question of whether people treat the rule as a norm, convention, etc.) As with Lewis and Pettit's talk of 'regularities', it looks as though if a norm is identified with a 'behavioural rule', we are going to have to fend off the thought that qualitatively identical norms or conventions are also numerically identical.

The outline of a solution to these problems

The last few paragraphs might have presented a rather frustrating reading experience. I suspect that readers may feel that it is obvious how we can modify our approach so as to do justice to our claims about individuation, should we desire to do so: we need to mention the role of *precedent*. The cities *A* and *B* have distinct driving conventions (we imagine) because they developed from distinct precedents. Similarly, in the imagined case of the two drivers, between t_1 and t_3 there is no single precedent that the drivers follow continuously, and that is why a new convention is created after t_2 .

It may be surprising, given Lewis' focus on precedent in the development of his account, that in fact this element goes missing in his final 'definition' of convention. Indeed, he basically gives it up along the path to this definition, allowing that a fictional precedent, if believed in, could suffice to get a convention going (1969 39).⁷ Lewis' interest in precedent is really just that his account shows how a precedent *can* generate what becomes a self-reinforcing pattern of behaviour, without relying on prior agreement.

For their own reasons, Lewis, Pettit, and Bicchieri are more interested in the conditions under which some pattern of behaviour can be considered conventional in a population than in the question of what conventions *are*. Despite their apparent identification of conventions or norms with regularities or behavioural rules, it is better to understand their accounts as seeking to provide conditions for facts of the form, "it is conventional to φ in S among P ." Each aims to show that when these conditions hold, all else being equal people in P will continue to φ in situations of kind S , or it will continue to be rational to φ in S , etc. Bicchieri in particular is interested in what we might call the 'dynamics' of norms and conventions. She describes interesting sets of conditions under which, for instance, 'pluralistic ignorance' might sustain an unpopular norm in existence, or when the actions of

⁷ As an empirical claim we need not doubt that a convention might begin with people following a fictional precedent. Nevertheless, the role of the *actual* precedents they thereby set is crucial to individuating the convention.

‘trendsetters’ are most likely to have an impact on the rest of the population (2017 esp. Ch.1, Ch.5).

So there may be good reasons for an approach which focuses on conditions under which “it is conventional among P to...” or “there is a convention among P that...”. Nevertheless, I suggest that our ordinary picture of conventions treats them as something more like ‘continuants’, individuated in part by reference to their origins in precedent. And as we have seen above, it is difficult to reconstruct this picture by starting with the conditions under which there is a convention among P , and trying to restrict them so as to guarantee that there is exactly one convention among P .⁸ Even if we succeed in doing that, however, we may still have no answer as to what the convention *is*. (We have seen that any answer identifying a convention with a universal is likely to get the claims about individuation wrong.)

I suggest taking a different tack. In the rest of this chapter, I explore how we can put ontology first by attending to the ontological category of ‘abstract individuals’, which satisfies and helps to explain the basic pattern of claims about individuation that we have been concerned with. I will then show that drawing on Ruth Millikan’s work on ‘natural conventions’ provides the basis for understanding conventions, along with norms and rules, as abstract individuals.

I take it that my proposal is basically consistent with what is of value in the approaches of Lewis, Pettit, and Bicchieri. Millikan’s account rests primarily on patterns of behaviour being *reproduced* from one situation to another in interesting ways. The accounts we have been discussing so far might then be thought of as revealing certain ways in which a pattern of behaviour can come to be reproduced,

8 Perhaps the most promising avenue for this sort of approach would be to develop on Brian Epstein’s recent work on the social ontology of groups, which includes an intricate discussion of how to keep track of things like teams over time, throughout changes in personnel, etc. (Epstein esp. Ch.12). Epstein’s account makes use of reference back to the particular event through which a given group was created, which has a function analogous to that of precedent in the individuation of conventions. I suspect that if this approach could be carried out successfully, it would basically match up with the account of the creation and persistence-conditions for conventions which I am developing here, albeit with a different emphasis.

for instance because it facilitates coordination or was publicly performed by a high-profile trendsetter. At the same time, my proposal can help to ground these other accounts, by providing a basic ontology of conventions, rules, and norms which sheds light on their unity with each other and with other similar phenomena.

It also helps to illuminate the unity within a *given* convention, rule or norm. Bicchieri's account of social norms requires that a significant proportion of a population holds a certain set of attitudes towards a 'behavioural rule', and her account of conventions requires that they hold a different set of attitudes towards a behavioural rule. She acknowledges at various points that, "[t]he neat boundaries I drew between descriptive norms, conventions, and social norms are quite blurred in real life: Often what is a convention to some is a social norm to others[...]" (2005 39). It is natural to read this as saying that there is some thing which is treated by some as a convention and by others as a social norm. And at least for some cases it seems like this is what we would want to say: this time, *not* that there are two distinct but qualitatively identical norms, but rather that there is one convention or norm which is regarded differently by different people within a given population. But Bicchieri's account of what conventions and social norms *are* sheds no light on why we don't always just see two separate phenomena here. The account I go on to develop will explain this, whilst also allowing space to acknowledge the relevance of the fact that, as Bicchieri points out, different people might adopt different attitudes towards it. I will introduce my account through a discussion of the ontology of art, which provides helpful insights into the ontology of conventions.

Repeatable artworks as created types

Our claims about when a particular action belongs to or manifests a given convention, and about how conventions themselves are individuated, run parallel to some common views about 'repeatable artworks'. Key examples of repeatable artworks are things like novels, symphonies, and films. The intuitive distinction between a repeatable and a non-repeatable artwork is this: whilst a copy of an oil painting might be thought of as a *mere* reproduction, a copy of a novel is just as much an instance of the novel as the model it reproduces, or the manuscript from

which it was originally printed. If I lose my copy of *We Have Always Lived in the Castle* half way through, and have to finish reading it from a new copy, there is nevertheless a significant sense in which I've been reading the 'same novel' the whole time.

By contrast, there is a particular concrete object such that if I have not seen that very object, then I have not seen da Vinci's *Mona Lisa*. And this is not considered to be an accident, in the way it might just happen to be true at a given time that there is only one copy of a given novel in existence.⁹ Even if we don't *identify* the painted work with the particular canvas and paint, the two are tied together in an intimate way, whereas the novel exists in each of its various physical manifestations in such a way that one could always in principle create more. Similarly with a symphony – two performances in different places at the same time count as producing distinct instances of the *very same thing* as each other, i.e. the piece of music.

We can now frame the questions of how concrete particulars (e.g. particular books) come to be instances of a given novel, and of how we individuate novels. First, take the book in front of me: a copy of *We Have Always Lived in the Castle*. This book contains many words arranged in a particular order, and by any account that is a large part of what matters about a novel. Now we can imagine that there is another book – a distinct concrete particular – containing the same words in the same order. Is this duplicate book also a copy of *We Have Always Lived in the Castle*? The answer is that we can't tell, just from knowing that it contains the same words in the same order as my copy. The duplicate book might be a spontaneously-generated 'swamp book'. It may have been printed by a computer running a random word-generating programme which, analogous to the 'thousand monkeys', happened to hit on a

9 Strawson thinks that it is just an accident in the case of paintings, due to our lacking sufficiently fine means of reproduction (231 n.1) However, this amounts to thinking that paintings have the properties which I suggest belong to repeatable artworks, not to denying the present characterisation of the repeatable/non-repeatable distinction.

combination of words precisely matching those found in my copy of the novel.¹⁰ In neither of these two cases would it be a copy of *We Have Always Lived in the Castle*.

Another possibility is that it might be a copy of a *different* novel, which happened to contain the same words in the same order. This brings us to the second point: it seems possible for there to be two distinct novels which are composed of the same words in the same order. Such a possibility was famously played with by Jorge Luis Borges in ‘Pierre Menard, Author of the *Quixote*’. The scenario Borges describes is fanciful, and does not quite map onto the cases we are interested in here, but as Richard Wollheim points out, it is easy to imagine coincidental duplication happening with a short poem:

imagine a brief lyric poem, a quatrain written sometime in the sixteenth century out of the deathless part of the English language, and then suppose that, in the early years of this century, a falsely naïve poet had set down, out of his head, those same few lines, same spelling, and done so in complete ignorance of his predecessor. (Wollheim 114)

There is, I suggest, a good case to be made that we have two distinct poems in such a case, which are composed of the same words in the same order. We now have it that qualitative identity to a copy of a novel or poem is not sufficient for a concrete particular to be a copy of that work, and also that there can be two distinct works which share the same qualities.¹¹ These two claims are parallel to those we made about conventions earlier. And, as in the case of conventions, it seems intuitively that ‘history’ is playing a role in the way we carve things up (cf. Wollheim 114-115). What

10 If Davidson’s ‘Swampman’ does not (at least initially) ‘have thoughts’ (Davidson, 2001d 19; cf. 38, n.3) – nor, we might add, is it a human – then the ‘swamp book’ may neither be a book nor contain words, despite being a molecule-for-molecule duplicate of a book which contained words. However, in the case of the algorithm I take it that such an object could plausibly be a book containing words, yet without being a copy of the novel.

11 Borges in his story makes much of the fact that, despite sharing the same words in the same order, the two texts *won’t* have the same *aesthetic* properties, but it is plausible to think that the differences he notes are downstream of the question of individuation.

could a repeatable artwork *be*, such that we can make sense of these claims about individuation?

These individuation claims have often been put forward as the basis for an account of repeatable artworks as ‘created types’. For instance, Lee Walters has argued that treating repeatable artworks as created types helps to *explain* these points about individuation (462). On Walters’ view, “the distinction between a *type* and its *tokens* is a metaphysical distinction between a general sort of thing and its particular instances” (461). Some types might be eternal. However, if we are going to identify individual repeatable artworks with types, this will lead us to say that these types are “artefacts, created by the respective artists”, because this is something we want to say about artworks in general (462). Walters, along with P. F. Strawson and Wollheim, whom it is also plausible to think of as type-creationists about repeatable artworks, mention that “models of car, languages, and national flags”, also look to be created types (Walters 461; cf. Wollheim 52; Strawson 231).

Following Strawson, Walters suggests that “the existence of a type is tied to the possibility of generating tokens of that type, by reference to an individual, ‘which serves as a rule or standard for the production of others’” (462; cf. Strawson 233). Such individuals,

include both pre-existing tokens, like a sentence or a musical performance, and recipes for producing tokens of that type; recursion clauses for languages and musical scores are examples of recipes for producing tokens of a type, but are not themselves tokens of that type. In addition, types can be stored, for example, in memory, on a computer, or on film. Further tokens can be generated, then, either by copying existing tokens, or else by producing tokens according to a recipe, or from storage. (Walters 462-463)

This sketch gives us some indication of how types themselves might be created, and also of how our two basic claims about individuation might emerge. Let’s take a fairly straightforward example – that of our two poems from earlier. A poem might be created when its author writes it down. (Or perhaps just when she composes it in her head.) The manuscript then provides the basis for creating new copies of the poem, when it is transcribed, printed, or read out loud. If a particular set of spoken or

written words is derived from the original manuscript in the usual sorts of ways, then they will count as ‘tokens’ of the poem.

Now we can describe what makes it plausible that in Wollheim’s imagined scenario, the contemporary author’s poem is a distinct poem from the one composed of the same words, but written in the sixteenth century. Namely, that the contemporary author’s manuscript is *not* derived in any meaningful way from the chains of particular poem ‘tokens’ which descended from the sixteenth century poet’s manuscript. This is what suggests that the contemporary poet has created a new work when she writes down her lines, when someone merely transcribing the original manuscript has not. And the contemporary poet’s manuscript is itself the potential source of a number of copies – in writing it down she has created a new type. Of any particular inscription of this set of words, then, we can ask whether it descends from the sixteenth century manuscript or the contemporary manuscript, and this is what allows there to be two particular qualitatively-identical inscriptions which are nevertheless copies of distinct poems. Many questions remain to be answered, and there will be difficult cases, but this is the basic idea.

Some issues with the created types proposal

The previous section developed a rough picture of how a ‘created type’ is created, and of how we determine which particulars count as tokens of a given such type. However, when we compare this initial picture with some widely-held views about the type–token relationship, a number of questions emerge.¹² There are two features commonly ascribed to types which Walters’ account fails to explain. It is also unclear exactly how the category of created types is supposed to relate to other metaphysical categories. More seriously, in some respects, the two apparent features of created types are in fact quite deeply in tension with each other. In light of all of this, I suggest that we do well largely to shift our focus away from the category of created types, for now. Instead, in the next section I will take the core of Walters’ and

12 I am assuming, over the next few paragraphs, that created types would inherit the key properties of types more broadly.

Wollheim's accounts, and apply it to a different metaphysical category which I call 'abstract individuals'. In this section, I will elaborate on the problems with the created types proposal.

As Walters notes, it is common to introduce the distinction between types and tokens by noting a kind of ambiguity that tends to arise in counting situations. Asked, "how many letters are in 'transubstantiation?'" one wants to know whether to increase the count every time one encounters a 't', say, or whether one should only do this for the first 't' in the word. If the former, one is presumably counting *token* letters; if the latter, one is counting letter *types*. This kind of ambiguity seems to apply to many of Walters' examples, such as "how many novels are there on your shelf?" (462) It would also apply to 'how many flags are flying?' However, there is nothing in his basic picture of created types to show why this should be the case.¹³

To take another point, the relationship between token and type has often been thought to be more specific than that between "a general sort of thing and its particular instances", as Walters had put it (461). So, for example, Wollheim distinguished a number of different relationships that could hold between a "generic entity" and the "elements" which "fall under it", and claimed that the relationship between token and type was a particularly "intimate" one (Wollheim 50). For Wollheim, this intimacy had to do partly with the idea that a type shares all of those properties with its tokens which its tokens necessarily have *qua* tokens of that type (51). In Chapter 2 I will discuss in greater depth what this means and why Wollheim thought it, but for now we can note that nothing of this sort seems to follow from the basic picture of created types outlined by Walters.

The other point left unexplained by Walters' account is how the category of created types relates to other metaphysical categories. For instance, both Wollheim and Walters contrast created types with universals in a way which appears to leave open

13 It is interesting to note that whilst it is difficult to generate type/token ambiguity for 'convention', it is much easier for terms like 'greeting'. "How many greetings did you see?" could be answered with reference to 'tokens' – "six: there were four people and each pair greeted each other" – or 'types' – "two: half of them shook hands, the other half bowed to each other". On the 'type' reading, it is plausible that we are counting greeting *conventions*.

the possibility that there might be types which are neither created nor universals.¹⁴ Is this a genuine possibility? It is difficult to tell from the way they set things up.

I have described two features which are often attributed to created types, or to types in general. One is their connection with count noun ambiguity, the other is the idea that types are distinguished from other ‘generic entities’ by the fact that they bear a particularly intimate relationship to the elements which fall under them – their ‘tokens’. Given their prominence in the literature, it seems reasonable to expect that an account of types will explain these points, or instead explain them away. I suggested, however, that Walters’ account of created types does not appear to explain either of these features. In fact, I will now argue that there are a number of tensions between these two apparent features of types, some of which go quite deep. Perhaps we could avoid the tension by dropping one of the features from our picture. However, I will instead suggest that we do better – at least for my purposes – to start afresh using a concept with less baggage than that of *types*. I begin to describe this alternative concept in the next section.

I have seen little, if any, engagement in the literature with the fact that the ‘ambiguity’ approach to the type/token distinction seems to make this distinction rest on the existence of a third term. We can see this in our initial example: we have 1) the ambiguous term ‘letter’, 2) ‘the letter ‘t’ (type), and 3) ‘the letter ‘t’ (token). The type–token relationship is supposed to hold between the referents of (2) and (3), so what exactly is the ambiguous third term ‘letter’ doing here? True, we can talk ‘types of letter’ and ‘letter tokens’. However, what we are calling ‘letter tokens’ here (particular datable, locatable inscriptions) are not tokens of the type ‘types of letter’.

The same applies to other key examples. To introduce the distinction between the Union Jack considered as a type and as a token via a counting ambiguity, we need a third term like ‘flag’, which in some sense encompasses both the type and the token. It looks like we *can* get a type/token ambiguity when asking “How many Union Jacks?”, but in what looks like the most natural ‘type’ reading of this question, it

¹⁴ To be precise, I should acknowledge that Walters claims that types are universals, but universals which are ‘impure’ and ‘created’ (463). Using this vocabulary the question would be whether there could be impure non-created universals, or pure created ones, in addition to the pure non-created, and impure created ones he mentions.

would refer not to what we've been calling the Union Jack (type), but rather to what we might think of as *varieties* of Union Jack. (We might intelligibly ask, for instance: "Should we count the so-called 'Protectorate Jack' as a Union Jack?" – 'Protectorate Jack' here seems to be functioning as a 'type' term.) Thus, although 'Union Jack' is a term which sometimes refers to a type, the type(s) which can be introduced by noting the ambiguity in some uses of 'Union Jack' do not seem to be identical to the Union Jack considered as a type of flag. This is what I mean by calling it a 'third term' with respect to the type/token distinction(s) which it can be used to introduce.

Helen Steward notes that it seems as though almost *any* count noun can generate the kind of ambiguity which we are discussing here. Even unpromising examples such as 'mountain' can be made to:

consider the question 'Which mountains do you like?' Here, I might reasonably reply, 'Oh, rugged, glaciated ones, mostly', rather than giving a list of individual mountains – thus indicating that I have understood the question to be asking about mountain types, not tokens. Or one might speak of 'the mountain', meaning to refer not to any individual peak, but rather to a type of geographical landform. (Steward 123)

Steward refers to both of these as 'type readings' of 'mountain', but they differ importantly in a way that she does not seem to note. In the first case, 'mountain' is the ambiguous count noun, and 'rugged, glaciated mountains' figures as a 'type' of mountain, of which the tokens will be particular rugged, glaciated mountains. In the second case, the ambiguous term seems to be 'geographical landform', under one reading of which we would count types like 'the mountain', 'the valley', etc., whose tokens in turn will be particular mountains or valleys, such as Scafell Pike or Wye.

If Steward did indeed fail to notice this difference, it goes to show how slippery this subject matter is to handle, which may explain why the relation of the type/token distinction to what I am calling a 'third term' does not seem to have been adequately addressed in the literature. But the role of this third term has significant implications, if the existence of count noun ambiguity is held to be central to understanding the type–token relation itself. (Steward and many others do hold this.) If the ambiguity of some count noun whose referent cannot be identified with either the type(s) or

tokens which it counts is necessary and sufficient for positing a type, and if Steward is right that this ambiguity is totally ubiquitous among count nouns in general, then it is very difficult to see how we can posit *any* kind of positive relationship as holding between a type and its tokens, such as the one suggested by Wollheim.

This point would probably not concern Steward, who advocates thinking of the type/token distinction,

as a relative, logical, not an absolute, metaphysical, distinction. That something counts as a type relative to some range of tokens ought not, on this understanding, to rule out its categorization as a token relative to further more general types; and neither should the classification of something as a token of some type or other mean that it cannot count as a type relative to further tokens. (123)

On such a view, we should not have much confidence that there will be anything both general and substantive to be said about the relationship between a type and its tokens. Having said this, the more fundamental problem for Wollheim's approach is not that we have shown it to be impossible for there to be anything interesting to say about the relationship between a type and its tokens, but rather that his *methodology* appears to be in tension with understanding the distinction via the ambiguity. If count noun ambiguity is central, we cannot hope to capture the category of 'type' by giving a positive characterisation of the type–token relationship, except insofar as this characterisation is subservient to explaining the ambiguity of count nouns. There is a deep tension between these two ways of understanding the category, suggesting that at least one of them should be given up.

Steward's treatment of the distinction as a 'relative, logical' one, understood through count noun ambiguity, represents one way to avoid this tension. There are two critical comments worth making about Steward's position, however. The first is that we should question her grouping-together of 'relative' with 'logical', and 'metaphysical' with 'absolute'. Accepting that the distinction is 'relative' insofar as the same entity can be both a type with tokens, and also a token of some other type, does not seem to entail that the distinction has no interesting metaphysical basis. It may just be that the metaphysical basis is a relational one, for instance. The second

point is that here again it seems as though Steward is overlooking the role of what I am calling the ‘third term’, on her account. The ‘relativity’ she highlights is of the sort that we might well expect here: given that a ‘token’ is a ‘token *of*’ some type, whether we can consider a given individual as a ‘token’ depends on its relationship to some type; and so it seems open that a ‘token’ might also be a ‘type’, as a parent can also be an offspring. This is distinct from the more puzzling relativity which is present in her account, but not addressed by her discussion. Namely, the way in which her account entails that identifying a type and its tokens always implicitly involves some ‘third term’.

In the rest of this chapter I will propose an account of conventions which shares a lot with Walters’ account of repeatable artworks as created types. However, in recognition of the complications just noted concerning created types, I will approach things slightly differently. In the next section I describe an ontological category – ‘abstract individuals’ – of which created types, if there are such things, are plausibly a sub-species. I suggest that the two basic points about individuation which are common between conventions and repeatable artworks belong to the genus: abstract individual. This allows us to clarify the nature of conventions, without committing to their belonging to a more determinate ontological category such as ‘created type’, which brings its own baggage. For instance, I will not take on the burden of trying to explain count noun ambiguity. There may be important things to be said about narrower categories such as that of created type, but focusing the broader category of abstract individual will allow me to say everything I want to say here.

Abstract individuals: the basic picture

What is an abstract individual? I will introduce the category in relation to two categories which are perhaps more familiar: concrete particulars and universals. It is possible to understand the distinction between concrete particulars and universals in terms of two criteria:¹⁵

1. Multiple locatability

¹⁵ I owe this way of formulating the distinction to Mike Martin.

2. The ‘identity of indiscernibles’

Spelling out both criteria fully is not something I try to do here. We should note that to say that something is ‘multiply-locatable’, in this sense, is *not* to say that it has spatial parts in multiple locations.¹⁶ For (2) to do its job, the relevant dimensions of ‘indiscernibility’ must be restricted in such a way as to make (2) non-trivial. This will require at least excluding facts about the identity of the entities being compared. It may also require excluding all ‘impure’ properties – roughly, properties whose possession consists in bearing a relation to a *given* other individual.¹⁷ So, perhaps, although *being married* is a ‘relational’ property, it will not count as an impure one. By contrast, *being married to Queen Elizabeth II* will count as impure (Humberstone 211).

Throughout I will generally use ‘qualitatively identical’ when referring to abstract individuals to mean ‘agreeing in all those properties not excluded by the identity of indiscernibles thesis’. When discussing whether two concrete particulars are instances of the same abstract individual, I may use ‘qualitatively identical’ to mean ‘agreeing in all those properties not excluded by the identity of indiscernibles thesis, and which might plausibly be thought to bear on their being instances of the same abstract individual’. So, roughly speaking, I would say that two books are qualitatively identical when they contain all and only the same words in the same order. We are not tempted to think that in general two copies of the same novel must be molecule-for-

16 Whilst I am talking of spatial parts here, the same thing could be said of the temporal parts of complex events. For present purposes I am treating both events and physical objects as concrete particulars. As mentioned in the body of the text, many questions need to be settled in order to establish an adequate characterisation of multiple locatability, and it may be that importantly different stories need to be told about the locatability of events versus physical objects.

17 Walters offers a discussion of impure properties (463 n.11), drawing on work of Humberstone’s. His formulation seems slightly too strong, but along the right lines. It may be too strong insofar as it requires that for any impure property, there is some single individual to which all possessors of that property are related in a certain way. But it seems possible for there to be interesting impure properties which are defined in terms of relations to, say, a number of distinct specific ‘ancestors’, such that there is no single ancestor that all bearers of the property share.

molecule duplicates, or even that they must be printed in the same font on the same kind of paper, etc. Similarly, two conventions might be ‘qualitatively identical’ if they involve the same pattern of behaviour, relevantly described, such as driving on the right hand side of the road. I hope it will be clear enough from context which sense is meant on a given occasion.

Typically, both (1) and (2) are thought to be satisfied by universals, whereas neither (1) nor (2) is thought to be satisfied by concrete particulars. Logically, this way of drawing the distinction allows room for things which are neither concrete particulars nor universals. (See Table 1.) We may or may not have use for the idea of entities for which (2) holds but (1) does not. However, the idea of entities which are multiply-locatable but for which the identity of indiscernibles does not hold ((1) is satisfied but (2) is not) looks very useful. We can say that this combination defines the category of ‘abstract individuals’.

	Multiply-instantiable	Not multiply-instantiable
Identity of indiscernibles applies	Universals	?
Identity of indiscernibles does not apply	Abstract Individuals	Concrete Particulars

Table 1. Characterisation of abstract individuals in terms of multiple instantiability and the identity of indiscernibles.

As we can see from the discussion above, ‘created types’ as understood by Walters and others appear to be a sub-species of abstract individuals, according to this characterisation. Repeatable artworks are multiply-instantiable, and it is possible to have two distinct yet qualitatively identical works. This would mean that a particular repeatable artwork, such as *We Have Always Lived in the Castle*, would be both an abstract individual and a created type, if it turns out that there are indeed created types. I suggest that what I have called our ‘folk sociological’ idea of conventions *prima facie* speaks in favour of them being abstract individuals, too. Multiple

locatability gives us access to the idea of there being distinct ‘instances’ of a work or convention, and the failure of the identity of indiscernibles secures the possibility of our points about individuation. I will elaborate on this further in the final few sections of this chapter. Before doing this, however, it will be helpful to gain a clearer understanding of what abstract individuals are, and how they can come into being. This is the topic of the next section.

A paradigm for understanding abstract individuals

An account of abstract individuals must explain how they can collect their instances. We might expect that in general such an explanation will be expressible in terms of a certain relation or set of relations holding between the instances of any given abstract individual. If this is correct, then these relations will need to possess certain minimal formal properties, though it seems unlikely that we will be able to give a purely formal account of when a set of particulars will be instances of some abstract individual.

Rather than trying to construct a general account of abstract individuals from the ground up in this way, I focus on one particular paradigm through which we can understand the existence of an abstract individual and how it collects its instances. This paradigm yields a familiar and intuitive sense of what talk of ‘instances’ means here, and it has the right formal properties to yield our basic claims about individuation. If this is successful then it looks like the most appropriate paradigm through which to understand conventions, repeatable artworks, and many other artefact kinds as abstract individuals. However, this may not exhaust the realm of abstract individuals. I briefly discuss two possible additional paradigms below, though I make no attempt at a more general theory of abstract individuals here. Comparison with these alternative paradigms will also help us to appreciate the explanatory potential of conventions, on my account.

One way to understand the possibility of there being such things as abstract individuals is via the notion of *reproduction*. Elements of this position were present in Wollheim’s and Walters’ discussions of how tokens of a repeatable artwork are

created. The concept of reproduction also figures centrally in Ruth Millikan's work on 'natural conventions', which I will discuss at length below. But I will first introduce the theme of reproduction in a way that makes clear its relationship to the metaphysics of abstract individuals.

There is a sense of 'reproduction' which allows us to point to two distinct concrete particulars – the original and the reproduction – and say, "here we have the same again as we have there." Yet 'the same again' here does not only (and may not even) pick out 'qualitative identity'. Nor is it like reidentifying a given physical object by pointing to two distinct spatial parts of it and saying 'this is the same again'.

Perhaps the simplest paradigm of reproduction in this sense is 'copying' of certain kinds. We might think here of tracing an outline, taking a mould of an object, or imitating a piece of behaviour. It seems as though such paradigms represent the simplest forms of reproduction, and a model through which we can understand other forms, but I do not want to suggest that other more complex forms of reproduction must somehow reduce to the simpler (cf. Wollheim 52). Rather, the idea is that the reproduction of a given kind of entity is that which allows us to identify the product as a new instance of that kind. If this is so then, at least interpretatively, there will be an interplay between our conception of what the entity *is*, and what counts as 'reproducing' it. Reproduction in this sense is thus something that can be realised by very different kinds of processes in different contexts, which is a point that much of the next two chapters will be given over to clarifying.

These observations begin to show how we can make sense of the possibility of multiple locatability coupled with the failure of the identity of indiscernibles, through the notion of reproduction. Treat 'reproduction' in the relevant sense as a transitive relation. Then we might say that the instances of a given abstract individual (say, the novel *We Have Always Lived in the Castle*) are the original manuscript and all reproductions of the manuscript. Taking a term from Millikan, although modifying its meaning slightly, we can call this a 'first-order reproductively established family' (1984 23-24). Because reproduction is transitive, copies of the novel which were not reproduced directly from the manuscript will still count as reproductions of the manuscript. It is clear that two things could be qualitatively identical, and yet not

belong to any overlapping reproductively established families. Thus we have a basis for (1) and the failure of (2), as was required by our picture of abstract individuals, from the previous section.

The true picture will be more complicated than this in at least three respects. First, a given particular might belong to more than one reproductively established family. Given this, we ought to use a notion of ‘reproduction’ which is in some way relativised to a given kind. Second, it seems too strong to require that each reproductively established family begins with a single particular which can be considered an ancestor of every member of the family, and which is also a member of the family itself. Whilst this might be plausible for novel manuscripts, it does not seem as though it will necessarily apply to all cases.¹⁸ Finally, and more seriously, if there is no general account of what ‘reproduction’ consists in which is independent of the nature of the thing being reproduced, it may be that we cannot rely on a *prior* understanding of ‘reproductively established family’ in order to determine what counts as an instance of an abstract individual. Nevertheless, I take it that the notion of reproduction in play here has the right formal properties to allow us to say that

18 Tokens of a symphony may all trace back to the score, which is not itself a token of the symphony. What about kinds with multiple starting points? It seems quite possible that we cannot pick out the lineage of any given sexually-reproducing species by reference to just one individual of that species. However, even assuming that for any such species there will be one individual to which all others are related in the relevant way, there is a slightly more complex case where reference to two or more ancestors may be ineliminable. Consider fertile hybrids of two distinct plant species, *A* and *B*. There may be multiple events where *A* and *B* are hybridised, and both Ruth Millikan and Richard Boyd suggest that the products of these distinct events, and their offspring, may form a new, unitary species, call it *C* (Boyd 1999 80; Millikan 1999b 100). For Millikan, this species is still individuated historically. Specifying the conditions for membership in *C* will, however, require reference either to relationships with members of *A* and *B*, or else reference to relationships with one of the initial hybrids which inaugurate various lineages of *C*. Either option requires reference to more than one ancestor. I have struggled to think of a realistic artefactual parallel, but it seems conceivable that one might be found.

qualitative identity to a member of a given reproductively established family is not sufficient for membership in that family.

We have not seen any reason to think that this is the only way to understand the possibility of abstract individuals. For one thing, copying does not seem to be the only basic paradigm of reproduction. For another, there appear to be other logical possibilities for constructing the basic features which we have said belong to this category. Neither of these points are damaging to my present aims, but it will be useful to briefly discuss these other options.

Many intuitive ways of thinking of biological species conceive of them as abstract individuals according to criteria (1) and (2) as given above (e.g. Wetzel Ch.1; Thompson 2008 Ch.1-4). And some notion of reproduction is central to these ways of thinking: roughly, for the most part, we think that members of species S must come from members of species S . The physical processes of reproduction will vary from species to species, but nevertheless we can intelligibly group them together and say that ‘biological reproduction’ plays a key role in our efforts to sort individual animals into species. At the same time, we should note that the basic idea of biological reproduction, or at least of sexual reproduction, does not seem to resemble any kind of ‘copying’ process. It may be, then, that we have two or more basic paradigms of reproduction through which we can understand the existence of abstract individuals, and the paradigm of copying may apply only to the realm of artefacts.¹⁹

There are certainly some conceptions of biological species which do not regard them as abstract individuals. Amongst the more realist views of species, one which seems clearly at odds with understanding them as abstract individuals is the view of

19 That individual organisms (except maybe things like amoeba) are not ‘copies’ of each other is stressed by both Dawkins and Millikan. Both in fact suggest that we should understand the ‘reproduction’ of organisms in terms of the more literal copying of genes, in which case biological reproduction would not really count as a distinct paradigm. However, although this move may be legitimate, I don’t think it shows that we never had a coherent idea in our naive biology of a paradigm of reproduction distinct from ‘copying’. For an argument that natural selection could operate on entities which reproduce without the involvement of any kind of ‘copying’, see Godfrey-Smith (2009).

Ghiselin and Hull that species are ‘individuals’ (e.g. Ghiselin 1974; Hull 1978). (Millikan herself has expressed cautious sympathy with this view, e.g. (2000 Ch.2.2; 2013 153).) The similarity in terminology should not lead us to think that there are no substantive differences between this view and mine.

David Hull, for instance, talks about the relationship between e.g. particular tigers and the species to which they belong as a “part/whole” relationship (1978 352; cf. Ghiselin 536). On his account, particular organisms and the species to which they belong are ‘individuals’ in just the same sense (1978 353). The part–whole relationship between particular organisms and their species is therefore the same at some level as the part–whole relationship between a particular tiger and its liver. This is at odds with my characterisation of the ‘multiple locatability’ of abstract individuals.

The debate over the Ghiselin–Hull view involves a number of fairly sophisticated claims about the role of species in scientific explanations. However, it is striking how often this debate also seems to rest on presupposing a *dichotomy* between things which are multiply-located/-instantiated on the one hand, and things which are individuated by reference to ‘historical’ factors on the other. (See e.g. Hull’s way of drawing the distinction between ‘individuals’ and ‘classes’ (1978 336–337).) The formulation of criteria (1) and (2) above is intended to show that these are in fact cross-cutting considerations, so that if the category I have labelled ‘abstract individual’ cannot in fact be occupied by anything, we require a positive argument for that claim.

An alternative way of understanding the possibility of abstract individuals without relying on *any* paradigm of reproduction seems to be provided by social ontology of a broadly Searlean form. Searle holds that many of what we might call ‘social kinds’ are generated by conferring a new status on an existing kind of entity, in accordance with collectively accepted rules. The main schema for such rules is “X counts as Y (in circumstances C).” Social kinds which are supposedly generated like this include examples of what are surely abstract individuals, according to the present criteria, such as the US one dollar bill. The one dollar bill is multiply-instantiated, but qualitative identity to a dollar bill is not sufficient for being a dollar bill. According to Searle,

These pieces of paper satisfy certain conditions that constitute satisfying the X term. The pieces must have particular material ingredients, and they must match a certain set of patterns (five dollar bill, ten dollar bill, etc.). They must also be issued by the Bureau of Engraving and Printing under the authority of the U.S. Treasury. Anything that satisfies these conditions (X term) counts as money, i.e. U.S. paper currency (Y term). (1995 45-46)

The condition as part of the ‘X term’ that the bills must be “issued by the Bureau of Engraving and Printing” adds an element of ‘impurity’ to the criteria for counting as a one dollar bill. This example is awkward for my purposes, insofar as reproduction in the sense described above is in fact surely involved in the generation of US currency (although this is not really acknowledged by Searle). However, it points to the logical possibility that we might generate abstract individuals without appealing to reproduction, by including ‘impure’ properties in the ‘X term’ of a constitutive rule generating a new kind of entity. An example might be, “All stones smaller than two inches in diameter, within three metres of where I am currently standing, count as tokens in my new game.” We might also try to imagine an example of a repeatable ‘found artwork’.

If Searle’s project more broadly is cogent, then there doesn’t seem to be anything to preclude this possibility, so unless its successful application involves a covert element of reproduction, this would provide another, distinct, paradigm through which to understand the possibility of abstract individuals. Does Searle’s idea include a covert element of reproduction? Any broadly Searlean approach to social ontology will need to give an account not just of the forms that ‘constitutive rules’ can take, but also of what can put such rules in place. Searle’s own story here appeals to a rule’s being ‘collectively accepted’ (1995 23-26), or ‘collectively recognised’ (2010 102-103). One might imagine that part of what makes recognition ‘collective’ would be that there is an interesting story to be told about how it has spread through a given population. Such a story would presumably involve some element of ‘reproduction’, though Searle seems to doubt this (2010 57-58).

Another thought is that there might be interesting causal or constitutive links between ‘collective recognition’ of a rule and behaviour which in some way

conforms with it. If this is how a rule is given reality, it might be usefully conceptualised in terms of some notion of reproduction. Searle sometimes gestures toward such connections (e.g. 2010, 104). A proper assessment of this point would require a grasp of Searle's understanding of 'functions' and 'reasons', although his remarks on this tend to be quite unclear at important points. Ultimately, however, I suspect that the best way to tell whether what I have labelled the 'broadly Searlean paradigm' rests on an element of reproduction will not involve determining what Searle himself is committed to. I will return to some criticisms of Searle in Chapter 6.

What these remarks on Searle suggest is that the combination of (1) and (2) through which we have characterised abstract individuals is not by itself enough to secure an interesting explanatory role for such entities. It seems reasonable to hold Searle to some standard of explanatory potential for his social ontology. However, unless we think that realism about some entity always depends on its having explanatory potential, there might be real abstract individuals which have none. One might think, for instance, that for any 'impure' property, there will be an abstract individual corresponding to that property, consisting of something like the fusion of its instances. This would be like the example of the stones within a three metre diameter, but this time it would not depend on any 'collective acceptance' that they constituted an abstract individual. There seems to be no reason to expect such an entity to play any role in explaining the existence of its instances. Rather, in such a case the instances will have priority over the abstract individual in question.

By contrast, we have seen that it is part of our ordinary conception of conventions that they can have a role in explaining the occurrence of their instances. And something similar holds in the case of repeatable artworks. Wollheim and Walters take it that the work itself, that is to say the type or abstract individual, plays some role in explaining the nature of its tokens, and our appreciation of the work (e.g. Walters 470-471). I suggest that when we understand abstract individuals through the paradigm of reproduction I have described in this section, this can help us to see how they could have such explanatory potential. Understanding a token of a created type through the notion of the reproduction of the type, via something like copying another token of the type, provides a basis for understanding how the type (or

abstract individual) might explain the existence of its tokens or instances. I will say more about this in the following two sections. By contrast, it is unclear that Searlean social ontology, as he has developed it, has much explanatory power. The explanatory role of biological species, and their ontological status, is a complex matter which I cannot explore further here.

I mention the alternative possibilities presented by biological reproduction and Searlean social ontology largely to put them to one side here. From now on as I use the term ‘reproduction’ I intend to exclude biological reproduction (unless stated otherwise) and to refer just to the form of reproduction we introduced through the notion of copying. Returning to that case, we can discern some broad, important modifications of the basic copying paradigm.

In the most basic form of copying, a token of the type being reproduced is integral to the reproduction process. Think of tracing an outline. In other forms of artefact reproduction, tokens of the type being reproduced play a more indirect role. A significant subset of forms of reproduction involve what Walters called non-token ‘embodiments’ of artefactual types. These are “individuals which allow for the generation of tokens” of a given type but which are not themselves tokens of the type (463). Examples are a musical score or reel of film. Another important division among processes of reproduction may be between those which employ a conventionalised notation or representation of the type being reproduced (e.g. bitmap image files or musical scores) and those which do not (e.g. woodblocks carved for printing). Different processes might generate not just new ways of reproducing a given artefact, but also new restrictions on what will count as a reproduction. It might be the case that images printed from the artist’s original woodblock will count as tokens of her work, but if someone copied her block and printed new images from that, they would not count. Different forms of reproduction might give rise to different kinds of *artwork* (cf. Wollheim).

Thus, although we began by understanding the relevant sense of ‘reproduction’ in terms of ‘copying’, we have now elaborated the picture until the two no longer coincide. ‘Reproduction’ remains the key term through which we understand the generation of new instances of a given abstract individual, but reproduction does not

always involve copying (as, for instance, when one plays a piece of music from a score) and not all copying counts as reproduction (for instance, some very accurate copies are mere forgeries). Much of the next chapter will be geared towards making an analogous point about the reproduction of patterns of behaviour in conventions.

The development of different processes of reproduction may also, as Walters emphasises, generate new existence conditions for the relevant types. What Walters calls ‘embodiments’ may allow a type to survive when it has no tokens, which possibility may not have existed for more basic forms of copying. Further, we might say that a given symphony (considered as a ‘type’) begins to exist once its score is written down, before any tokens of it (performances) have been produced. However, we should remember that an embodiment of a type is only an embodiment of that type if it can facilitate the production of tokens of that type; we remain fundamentally within the paradigm of reproduction. This point is at odds with some of Walters’ remarks, which at times suggest that the notion of an ‘embodiment’ is supposed to be a more substantive one.

Reproduction in Millikan’s account of ‘natural conventions’

We saw at the beginning of the chapter that we want to say of two people who follow the Cambridge punting convention that here we have the same thing present on two separate occasions. But we are not referring to the instantiation of a universal, because then we would say that the same thing is present also in the imagined third city where they punt in the same way as in Cambridge, or in the action of the tourist who just happens to hit on the Cambridge style. We are also not picking out two parts of a single large, scattered event consisting of all the occasions on which people punt following the Cambridge method, in the way we might when we say “this is the same football match that was on before you left to go to the shops at half-time”. Given the account of abstract individuals I have just developed, it appears instead that what we are picking out when we say that we have ‘the same thing’ here is the abstract individual which is the Cambridge punting convention itself. This abstract individual is instantiated in the actions of each person who follows the Cambridge convention. In the imagined case where there is also a

separate, qualitatively identical convention in another city, the two conventions in the two cities would be two distinct abstract individuals.

Ruth Millikan, in her account of conventions, does not invoke the category of abstract individuals. And, as I mentioned above, she seems at times to think of conventions as something more like scattered concrete particulars. However, the details of her account of conventions are in fact a good fit with my suggestion that conventions are abstract individuals, understood through a particular notion of ‘reproduction’. Whether or not she would accept this ontological suggestion herself, her work provides useful resources for fleshing out my proposal.

Millikan describes conventions in terms of two factors. I will discuss one in this section, and the other in the next section. The first is that a convention involves the ‘reproduction’ of a pattern of behaviour. This is characterised as follows: “[a] pattern has been *reproduced* if its form is derived from a previous item or items having, in certain respects, the same form, such that had the model(s) been different in these respects, the copy would have differed accordingly” (2005a 3).

Millikan’s counterfactual gloss on reproduction is tied to her aspiration to provide a reductively naturalistic account of convention and biological function, though it is questionable whether she really adheres to this gloss in practice. Whether or not this project is viable, the counterfactual gloss does not seem appropriate to capture the role that reproduction plays in our thought about repeatable artworks or, I will argue, conventions. In particular, it does not look as though satisfying Millikan’s counterfactuals will be *sufficient* for a process to count as reproduction, in the present sense. I will illustrate this through a comparison with artworks.

We are familiar with the idea that conceptions of ‘the artist’ and ‘authorship’ have changed over time. One story about how this goes is found in Howard Becker’s work. Becker suggests that of all of the myriad cooperative activities which go into the production of a work of art, some get classified as ‘core’, whilst others are ‘support work’. In many places and times, wielding the brush will be ‘core’, whereas producing the paint or stretching the canvas will merely support the artist’s work. In general, artists are understood to be those who carry out the core activity, and are held to be the creators of the artwork produced through such activity. What gets

counted as core work for a given medium can change, though, and with it who counts as an artist. So, for instance, Becker describes how over time sound mixing went from being skilled support in the production of recorded music, to being “integral to the art process and recognized as such” (18).

Although Becker does not seem to want to take an explicit stand on this, it is plausible to think that sound mixers actually changed from being important technicians to being co-producing artists, through the historical process he describes. Similar shifts occur concerning whether the authorship of a painting created by many hands in a workshop is attributed to each of the people who put paint to canvas, or only to the head of the workshop. We can make such claims without committing ourselves to any form of ‘idealism’ about what constitutes the role of the artist, however. Concerning sound mixing, for instance, Becker and others describe shifts in the actual practice of mixing, in the effect it contributes to the work, in standards of appreciation of the work, in legal frameworks for assigning credit and payment and, of course, in broader cultural conventions. Such shifts can make a real and significant difference to which roles it makes sense to regard as authorial in the creation of works in a given medium; authorship is not simply a matter of being *considered* an author.

We should also acknowledge that although many of these changes will show up as broad trends, perhaps affecting whole mediums at a time, there is always room for idiosyncrasy in such matters. One paradigm, though not the only case, is that of the individual who turns her own contribution, whatever it may be, into the defining part of the work, regardless of the conventional status of her role. Whether a film is seen primarily as the work of the writer or the director may sometimes be decided like this.

This story about how the role of the artist can change has parallels in terms of the creation of new instances of a repeatable artwork. Typically, traditions of woodblock printing are such that tokens of the work must be made using the artist’s original woodblock. Tracing a print, and then using this to create a new woodblock, or a relief in some other medium, which is then used to create prints, would not then count as reproducing the original work, even if it produces a highly accurate copy.

Varying significance may also attach to the question of *who* makes the prints. Where it is conventional for specialised printmakers, whose role may be considered as ‘support’ rather than ‘core’, to make the prints, *any* prints made from the original block may count as instances of the work. But some artists will perform both the carving and the printing themselves, in which case it may be that only prints that *they* make, using the original block, will count as instances of the work.

If these suggestions are correct, then the accurate copying of a token of a work is not sufficient to reproduce the work itself. But as in the case of the changing role of the sound mixer, this does not mean that what counts as a process of reproduction of a work is a matter of mere stipulation, or that it has come completely unmoored from the paradigm of copying. Rather, aspects of the general artistic, cultural, and legal context within which woodblock printing is an art form, as well as aspects of the artist’s idiosyncratic practice, can make an intelligible difference, constraining what processes will count as reproducing a given work. I suggest that something parallel is true of what counts as reproducing a convention.

It seems more plausible that Millikan’s counterfactual is a *necessary* condition for a process to count as a process of reproduction in our present sense. That is, if the output of a process did not depend in this way on the model it took as input, it would not count as reproduction in the sense required to generate new instances of an abstract individual. The next chapter will be partly given over to arguing that the processes through which behaviours spread in a population – including the processes through which conventional patterns of behaviour are reproduced – do satisfy something like Millikan’s counterfactual. Chapter 3 will then make the case in greater detail against the *sufficiency* of her counterfactual account of reproduction, for understanding the reproduction of conventions.

Leaving these claims for now as points to be established in the next chapters, my suggestion is that once we modify Millikan’s concept of reproduction, we can use it in the way she suggests. We can put reproduction of patterns of behaviour at the centre of an account of convention. But we can also go beyond this, connecting it with the above discussion to make the claim that particular conventions are abstract individuals established through these same processes of reproduction.

According to this account, particular conventions, such as the Cambridge punting convention, will be abstract individuals, analogous to repeatable artworks such as novels. If someone punts in the Cambridge fashion because she is imitating someone else who is following the Cambridge convention, or because she has been instructed that this is the way to punt in Cambridge, etc., her actions will count as *reproducing* the Cambridge punting convention. A similar story will be told about the model whom she is copying, or the instructions her instructor has learned, and so on, back to some original precedent. This parallels the way new copies of the novel are created: they are generated in the usual sorts of ways from other copies of the novel, or ways of storing its text, tracing back ultimately to the original manuscript.

There are also important disanalogies between conventions and artworks. For most repeatable artworks, a person who reproduces the work will typically be *aiming* to reproduce it, to create a new token. This does not apply to the reproduction of conventions as such. A conventional pattern of behaviour might be reproduced without anyone trying to reproduce the convention. Indeed, various forms of unconscious or automatic imitation might reproduce a behaviour without the agent's being aware that she is reproducing any behaviour at all. For some conventions, such processes will count as reproducing the convention. (I will describe a case in greater detail in Chapter 4.) The fact that artworks are not reproduced like this has more to do with the peculiarities of art than with the notion of reproduction itself. We can see this by comparing repeatable artworks with non-repeatable artworks: typically one cannot create a non-repeatable artwork without being aware of doing so, either. But in this case it cannot be because reproduction requires intention, because non-repeatable artworks do not depend on reproduction in the first place.

In short, understanding conventions as abstract individuals created through the reproduction of patterns of behaviour allows us to explain the basic claims about the individuation of conventions which we began this chapter with. In addition, as we noted in the discussion of paradigms of abstract individuals in the previous section, grounding our understanding of conventions in this notion of reproduction allows us to understand the explanatory role of conventions, too. A further element in the explanatory role of conventions will be discussed in the next section.

Convention and arbitrariness

The second factor in Millikan's account of natural conventions is that the pattern of behaviour is reproduced in part due to the 'weight of precedent'. This criterion is supposed to differentiate conventional behaviours from the proliferation of things like skills through social learning. Millikan gives the following example of non-conventional behaviour which proliferates through something like reproduction:

I learned from my mother, and she from hers, to open a stuck jar lid by first immersing it in hot water. Opening jars this way is not thereby "conventional." To be thought of as conventional, a reproduced pattern must be perceived as proliferated due, in important part, to weight of precedent, not to its intrinsically superior capacity to produce a desired result, or due, say, to ignorance of any alternatives. (2005a 7)

In terms of the wider philosophical literature on convention, this is Millikan's version of the requirement that conventions be 'arbitrary'. (See Lewis (1969), Burge (1975), and Gilbert (1989) for some classic discussions.) To pre-empt a possible misunderstanding: to say that conventions are importantly arbitrary does not entail that there must be something arbitrary about *following* a convention, *once it has been established*; Lewis' example of establishing a side of the road to drive on shows us this.

Millikan's remarks on this second factor are brief and sketchy. Like Burge, she places some emphasis on what we might call 'sociological factors' in characterising the arbitrariness of conventions. She suggests that a given convention will, due to its arbitrariness, have "little tendency to emerge or reemerge in the absence of precedent" (Millikan 2005a 166). Burge writes,

[i]n the first place, conventions are not determined by biological, psychological or sociological law: the conventions a given person learns are "historically accidental." In the second, conventions are not uniquely the best means of *fulfilling their social functions*: other, incompatible means would have done as well. (254; my emphasis)

I call such approaches ‘sociological’ because they locate arbitrariness in an account of what a convention is doing in the wider social context, and what alternatives would be or would have been possible in its place. It does not locate arbitrariness in the minds of those to whom the convention belongs.

There are other, quite different, approaches to the theme of arbitrariness. Lewis, and many of those who follow him, has arbitrariness show up partly in the *preferences* of those who follow the convention. Condition (5) from his definition of conventions, discussed above, secures that there is some alternative possible convention that members of *P* conditionally prefer to conform to, if others do. For Lewis, alternative conventions therefore have to provide alternative *coordination equilibria*, given current conditional preferences (1969 14-15). Moving beyond Lewis, there are a number of alternative accounts of what kind of equilibria – e.g. evolutionarily stable strategies or stochastically stable strategies – might be of interest in connection with conventions (cf. Peyton-Young 118). And these can in turn be interpreted in a variety of ways – for instance in terms of preferences or in terms of fitness. Andrei Marmor speaks of the *reasons* agents would have, whether they realise it or not, to conform to an alternative rule, were that rule widely followed in their social context (351-352). Views also differ on whether, and in what sense, an alternative must remain a ‘live option’ if the present way of doing things is to be arbitrary (see Stotts (2017) for a discussion of some possibilities here).

It is a striking fact that there is broad consensus on the importance of ‘arbitrariness’ to an account of conventions, in the face of such widespread disagreement about what arbitrariness amounts to. I suggest that the intuitive pull of the arbitrariness condition is that the arbitrariness of a convention helps to secure our sense that *precedents matter* in the realm of convention.²⁰ However, there are many different ways

20 One might have thought that the role of precedent was adequately captured by the reproduction requirement. Ultimately, I think this is correct. However, Millikan’s example of the non-conventionality of the method she learned for opening a jar suggests that what we might think of as ‘mere’ reproduction does not necessarily secure a particularly interesting role for precedents as such. In Chapter 3 below I discuss in a little more depth what might make a process of reproduction an interesting one, but as I have said I do not aim to provide a full account of ‘arbitrariness’ here.

in which precedents can matter in shaping human behaviour, many of which in turn can support a sense of conventionality. Hence, we have a whole host of different accounts of arbitrariness, many of which may be onto something important for a given class of conventions. Much of the work in the literature seems to presuppose that we should expect to find a single notion of arbitrariness that will fulfil all of our needs here. However, if I am correct about the role arbitrariness is playing in an account of convention, there is no obvious reason to expect to find such a notion. I would suggest that we do best at this stage to take things on a roughly case-by-case basis, relying where possible on historical and social-scientific enquiry to reveal where the effect of precedent is interestingly dependent on social or cultural factors.

The arbitrariness of conventions helps to illuminate a further aspect of their explanatory role. There is a question of why we should attribute any explanatory power to the convention itself, conceived of as an abstract individual, over and above the explanatory powers of its particular instances.²¹ I have suggested that understanding abstract individuals through a paradigm of reproduction helps account for their explanatory role. However, as it has been characterised so far, this by itself does not seem to be quite enough. Consider Millikan's example of non-arbitrary social transmission. There is something we could think of as reproduction going on there, but it is much less tempting to think of the method she describes for opening jars as something like a self-perpetuating 'technique', than it is to think of the Cambridge way of punting as (in some sense) a self-perpetuating convention. How might the arbitrariness of the latter help to explain this difference in its explanatory power?

Once again there may be a helpful parallel with artworks. It is possible to elicit a sense that art is a matter of creation, whereas science (or craft) is a matter of discovery. One explanation for this is that constructing a scientific theory, although in many senses a 'creative' activity, is, at least when successful, significantly constrained by something largely external to the activity – namely, the nature of the world. Whilst artistic creation is also shaped by and responsive to external factors, there is not the same sense of a possible 'fit' between the work and the world.²² (This remains true throughout the various historical mutations of *mimesis* as an artistic ideal

²¹ I say more about the explanatory role of the convention itself in Chapter 4.

of which I am aware.) This could be thought of as analogous to the ‘arbitrariness’ of conventions. Subsequent uptake of the work (say, through reproduction or appreciation) therefore depends importantly on a connection with *what the artist has done* in creating it, rather than simply on the accessibility of an exemplar of something with certain qualities.

This is why it is important to draw a distinction between Cervantes’ *Don Quixote* and Pierre Menard’s *Don Quixote*. It is also what gives sense to talk of the role of ‘the work itself’ (considered as an abstract individual or created type), over and above its instances. That a particular book is an instance of one work rather than the other becomes relevant, in addition to its intrinsic qualities (which are shared by tokens of either novel, as well as by our ‘swamp book’ examples). Something similar holds of the distinction between socially-transmitted techniques and conventions. It makes sense to individuate conventions according to their particular precedents, and to accord some explanatory role to the convention itself, because of the arbitrariness of the convention, in relation to whatever ‘problems’ it might solve. By contrast, when learning the jar-opening technique, all you need is to see a demonstration – it doesn’t matter where that demonstration has ‘come from’. We can see from this that determining what counts as arbitrariness in a given case will often be a complex and involved matter.

If all of this is correct, then whilst Millikan’s division between reproduction and the ‘weight of precedent’ (or ‘arbitrariness’ as we have been calling it) was helpful for expository purposes, there may not be a fundamental distinction between the two in our account of conventions. If the concept of reproduction is to shed light on the explanatory power of conventions, then this concept of reproduction cannot be the same as the one operative in accounts of the spread of mere techniques through social learning. Rather, our understanding of the reproduction of conventions may need to *incorporate* the relevant sense of arbitrariness. Given the range of details and disciplinary perspectives that may be relevant to the question of arbitrariness, it will therefore be unsurprising if, as I argue in Chapter 3, we cannot provide a fully

22 Dennett (1996 140) explains the intuitive distinction between creation and discovery in terms of ‘forced’ versus ‘unforced’ moves in a ‘design space’. The gist of his point is similar to mine, although I would not accept his gloss on it.

general account of what ‘reproduction’ in the sense relevant to conventions consists in.²³

I will just make one further remark about extant approaches to the arbitrariness of conventions in the literature. It is striking how often they appear to be formulated in what we might call ‘atomistic’ terms. The holists of mid-20th Century social anthropology may have gone too far, at least if we believe their critics’ interpretations of their views.²⁴ However, it is a feature of human cultures that we often find in them a number of distinct conventions, norms, or practices which seem in various ways to support each other. Consider now Millikan’s counterfactual account of arbitrariness. For many conventions, it actually seems quite plausible that if we were to just erase the precedent, a qualitatively similar convention *would* arise, due to the influence of the various other conventions which surrounded it. (Pierre Bourdieu, for instance, has a concise discussion of how many distinct conventions surrounding gendered divisions of labour might mutually reinforce each other (1990 70).) Contrary to Millikan’s suggestion, the fact that a similar convention would be likely to re-emerge does not show that we are not in fact following precedent in any interesting way when we follow the existing convention.

This suggests the need to modify existing accounts of arbitrariness so as to be less atomistic. It also suggests a further reason why an account of what the reproduction of a convention consists in in a given situation may need to be sensitive to many details of the broader cultural context. I won’t, however, dwell on the theme of the arbitrariness of conventions, norms, etc. in what follows. I hope that in the examples I discuss it will be sufficiently plausible that they have whichever form of arbitrariness may be important for conventionality in the circumstances.

23 This is not to say that we have become completely unmoored from Millikan’s account, though; as I will argue in Chapter 2, it seems plausible that something like Millikan’s counterfactuals remain necessary conditions for reproduction in my sense.

24 See e.g. Merton (1949) and Hempel (1965). The classic sources of apparent holistic excess are A. R. Radcliffe-Brown and Bronislaw Malinowski.

Conclusion

I began this chapter by describing part of what I suggested was the ‘folk sociological’ concept of convention. I highlighted three core features of this concept, which are also shared by the concept of convention employed in much existing social-scientific work. These were the following: it is possible for a piece of behaviour qualitatively identical to a conventional action not to be a conventional action; it is possible to have two qualitatively identical but numerically distinct conventions; conventions can have a role in explaining the occurrence of their instances.

Over the course of this chapter I have developed an account of the ontology of conventions which fits with, and helps to explain, each of these key features. Conventions are abstract individuals, created through the reproduction of patterns of behaviour which are arbitrary in an important sense. The concept of reproduction secures both of our points about individuation. It is possible for somebody in Cambridge to punt in a way that matches the Cambridge style, without in any sense doing this *because* this way to punt is conventional in Cambridge. In such a case, their behaviour will match that of people who follow the convention, but it will not be conventional, because it does not reproduce the convention. Similarly, two conventions might look the same, but be distinct entities, if they trace back to distinct precedents. Finally, it may be that *some* abstract individuals have no kind of priority over their instances (I mentioned possible objects consisting of a fusion of particulars bearing some impure property). However, understanding conventions through the notion of ‘reproduction’ I have described sheds light on how conventions, considered as abstract individuals, can help to *explain* the occurrence of their instances.

Although I have written mostly of ‘conventions’ in this chapter, I have also suggested that my account will apply to ‘rules’ and ‘norms’, too. To make the point explicit, I am claiming that each of these three terms has a prominent use on which it refers to a kind of abstract individual, generated through the reproduction of patterns of behaviour which are arbitrary in the relevant sense. As I mentioned in the Introduction, it may be that we want to mark some important distinctions between the three terms, but I suggest that they have at least this much in common. So, the

objections which I will address in the next two chapters would apply equally to my account of conventions, rules, and norms. I will continue to use the term ‘convention’ most often in the following two chapters, though, before opening it out to ‘norms’ and then ‘rules’ in the later chapters, for the reasons sketched in the Introduction.

Chapter 2. Reproduction and Imitation in Human Life and Culture

In the previous chapter I outlined the view that conventions are abstract individuals, created through the reproduction of patterns of behaviour. The viability of this view depends in part, then, on the facts about how patterns of behaviour actually spread through human populations. At this point my account connects with ongoing empirical and theoretical debates on cultural diffusion. I will focus in particular on certain arguments by anthropologist Dan Sperber. These arguments are developed in the context of a broader debate in which Sperber argues against approaches which treat Darwinian ‘selection’ as a central driver of cultural change. Such approaches include, but are not limited to, memetics or meme theory. I will make some remarks on the broader debate in this chapter and the next, but they will not be my primary concern. Rather, I will focus on Sperber’s arguments for *nominalism* about cultural entities, which are developed in the context of his criticism of meme theory.

The core idea of memetics is that ‘memes’ are units of culture which can be replicated, thereby spreading through human populations. Sperber’s central argument against the existence of memes turns on the claim that the processes through which skills, behaviours, and artefacts spread do not generally amount to ‘high-fidelity copying’. Even though my proposal is *not* a form of memetics, *prima facie* Sperber’s argument presents a challenge to the empirical adequacy of the account of conventions I developed in Chapter 1, which centred on the reproduction of patterns of behaviour. Moreover, Sperber takes his argument not only to undermine memetics, but also to support a generalised nominalism about cultural entities. This nominalism, as I will show, is at odds with my proposal that conventions are abstract individuals.

In this chapter I will defend my ontological claim that conventions are abstract individuals, by showing that the notion of ‘reproduction’ relevant to my account is compatible with many of the claims Sperber makes about how behaviours spread through populations. Doing this will also help to clarify what we mean by ‘reproduction’ here. In particular, it is not the same as the memeticist’s notion of ‘replication’. I will then address some further nominalist arguments in Sperber’s work, which it is useful to hold separate from the first, as well as a worry about my proposal stemming from some remarks of Wollheim’s.

Outline of Sperber’s challenge

The previous chapter suggested that we can follow Millikan in treating the reproduction of patterns of behaviour as a key element in our account of conventions. Linking up with the discussion of abstract individuals, this provided the basis for saying that particular conventions are themselves abstract individuals, which helps to account for the two basic claims about individuation which we encountered at the outset. (Repeatable artworks and many artefact kinds were also revealed as abstract individuals, understood through the concept of reproduction.) I suggested that whilst Millikan’s counterfactual gloss on reproduction seems inappropriate for this purpose, something like it may be a *necessary* condition for a process to count as one of reproduction in the relevant sense.

Over a series of papers, Sperber has proposed an influential programme for a ‘naturalistic’ approach to social science. In brief, he proposes an ‘epidemiology of representations’. This project would look at the particular ‘mental’ and ‘public’ representations existing in a given population at a time, and try to explain why they are distributed as they are. At the same time, he has developed a number of arguments against two of the other major naturalistic approaches to culture: Darwinian approaches to ‘cultural evolution’ and memetics.

There are a number of differences between such positions, but Sperber collects both Dawkins’ memetics and, for instance, Boyd and Richerson’s ‘dual inheritance’ approach to cultural evolution under ‘cultural selection models’. In place of the

cultural selection of cultural entities (stories, techniques, ideas, songs) which was supposed to be analogous to the selection of organisms, biological traits, or genes, he proposes a model of ‘cultural attraction’. This alternative paradigm has recently spawned a number of empirical studies (e.g. Scott-Phillips 2017; Miton et al. 2015; Morin 2013), as well as debates back and forth with advocates of more traditional Darwinian approaches (e.g. Richerson and Boyd 2005 Ch.3; Claidiere et al. 2014; Acerbi and Mesoudi 2015; Buskell 2016; Morin 2016a).

Sperber advocates an ontology of cultural entities which we can think of as broadly ‘nominalist’. That is, he encourages the naturalistic social scientist to posit only particular actions, objects, and (mental and public) representations, and perhaps also sets of these. Take a term that appears to name something like a repeatable artwork, such as “the Bororo myth of the bird-nester” (Sperber 1996). I suggest that this term refers to an abstract individual. By contrast, Sperber suggests it is non-referring (1996, 27-29), or else refers to a (fuzzy) set of particular mental and public representations (33). He also sometimes talks of items like myths as ‘abstractions’ (81). The suggestion here appears to be that these abstractions would be *merely* creatures of the mind, with no real existence (rather than being ‘abstracta’ in my sense). Although he occasionally makes use of the type/token distinction in his discussion of cultural entities, it seems charitable, given his explicit claims, not to read this as implying commitment to the existence of types, as I am proposing to understand them. His explicit ontological position is thus directly at odds with my proposal.

At the core of both Sperber’s critique of cultural selection models and his arguments for nominalism is the claim that the psychological mechanisms which are largely responsible for the proliferation of cultural entities in human populations are not ‘replicative’. In brief, his claim is that whilst humans are capable of high-fidelity copying of behaviours and objects under the right conditions, the processes which *predominate* in cultural transmission are not high-fidelity and are the source of non-random mutations. Stated in this way, it is clearly an empirical claim. It is also a claim which has some bearing on my suggestion that the reproduction of artefacts and patterns of behaviour can be used as the basis for understanding repeatable artworks, artefact kinds, and conventions as abstract individuals. However, we do not have to

identify ‘replication’ in Sperber’s sense with ‘reproduction’ in my sense, and in the following sections I will argue that we should not. So we can accept Sperber’s empirical claims without giving up the idea that many of what we normally think of as conventions are abstract individuals generated through the reproduction of patterns of behaviour. Spelling this out will help to clarify what we *do* mean by ‘reproduction’, here, and will also put it on a firmer empirical footing.

I will not try to weigh how effective Sperber’s arguments against cultural selection models are. Rather, my aim is to regiment his arguments in a way that relates them most directly to the role I have claimed for reproduction in our understanding of conventions. My claim is somewhat dependent on the empirical facts about cultural transmission. However, the fine details of the selection versus attraction debate are not directly relevant for my present purposes.

It is worth noting, though, that defenders of what Sperber calls cultural selection models have questioned how damaging Sperber’s empirical claims are to their basic approach. One recurrent theme is that the two approaches might be reconcilable by being treated as belonging to two different levels of idealisation or detail (e.g. Acerbi and Mesoudi, 2015; Buskell, 2016). A second point is that whilst there is broad agreement that *memetics* depends on the claim that cultural entities are replicators, it is much less clear that Darwinian selection more broadly requires that the items on which selection operates are replicated (Godfrey-Smith 2009 Ch.2; Sterelny 2011 495). This point is also found in more recent work by Sperber himself (Claidiere et al.). A third theme in recent discussions has been the extent to which each side in the selection versus attraction debate has tended to focus on largely distinct canons of examples most suitable to their own theories (Sterelny 2006; Sterelny 2017; Godfrey-Smith 2012; Heyes 2018). The suggestion here is that each might capture quite well the processes operating in one area. This move towards pluralism seems welcome to me, and I would gently suggest that it might be extended beyond these two broad approaches.

A related question, which I have not seen adequately discussed in the literature, is precisely what the relations are between a number of oppositions which have been invoked in this debate. Roughly corresponding to the distinction between selection

versus attraction at the level of overall models, it has been suggested that there are important distinctions between processes of cultural transmission which are:

imitative vs. not imitative

high-fidelity vs. low-fidelity

replicative vs. reconstructive

replicative vs. transformative

Whilst the terms on the left-hand side are sometimes used as though they are synonymous in this context, each of the terms on the right-hand side has distinct connotations, giving the impression that we have a number of cross-cutting distinctions rather than a single distinction given different expressions. The question of how to properly regiment these distinctions is important for assessing the selection versus attraction debate, but it would take us too far afield to try to do this here.

Over the next few sections I focus mainly on Sperber's argument as presented in 'An objection to the memetic approach to culture' (2001). This argument is directed towards Richard Dawkins' account of memetics, which provides 'replication' with a relatively clear sense, which I will try to adhere to myself. I will only discuss Dawkins' position to the extent required to make sense of Sperber's objection. For clarity, it is perhaps worth reiterating that I do not take the proposal I developed in the previous chapter to be a form of memetics. As mentioned above, the basic point of this chapter is that we need not identify 'reproduction' in my sense with 'replication' in Dawkins'. Having made this clear, I then address some further nominalist arguments arising from Sperber's book *Explaining Culture* (1996) and, indirectly, from Wollheim's account of created types.

An overview of memetics

Memetics, developing from work of Richard Dawkins', attributes a meaning to 'meme' which differs from the one it's taken on in popular discourse.²⁵ Dawkins suggests that we should understand natural selection in the biological world in terms of what he calls 'active, germ-line replicators'. He defines a 'replicator' as "[a]ny entity in the universe of which copies are made" (83). An 'active' replicator is one "whose nature has some influence over its probability of being copied" (83). A 'germ-line' replicator is one from which in principle an endless line of replicators might descend. The copying of a particular gene at a particular location on the chromosome in a reproductive cell is the sort of thing that might in principle be repeated indefinitely through sexual reproduction over indefinite generations. By contrast, although particular cells in a given living human hand might replicate through mitosis, insofar as the body to which they belong is not immortal, this process seems limited in extent (83).

Dawkins suggests that when we have systems of active, germ-line replicators we should expect to see the processes we recognise as natural selection. This is why he suggests that the *gene*, rather than the individual organism, is the 'unit of selection' in biological evolution.²⁶ Individual organisms are not replicators. First, because in the case of sexually-reproducing organisms, it is not even true that a given organism's genome will get replicated through sexual reproduction, because this process involves mixing genes from two different genomes. Second, because features of the individual organism will not be passed down to its offspring (even in asexual

25 Two of the most prominent works explicitly taking up Dawkins' suggestions are Blackmore (1999) and Dennett (1996).

26 This gloss is slightly inaccurate insofar as there are a number of understandings of what a gene is, and the equivalence is achieved partly via (I think harmless) stipulation. There is a slightly more problematic inconsistency in Dawkins' own terminology however – sometimes he seems to use 'active, germ-line replicator' to mean what he at other times refers to as a '*successful* active, germ-line replicator' (88).

reproduction) unless they go via features of the genome.²⁷ Dawkins gives the example of a female stick insect (which can reproduce asexually):

A stick insect looks like a replicator, in that we may lay out a sequence consisting of daughter, granddaughter, great-granddaughter, etc., in which each appears to be a replica of the preceding one in the series. But suppose a flaw or blemish appears somewhere in the chain, say a stick insect is unfortunate enough to lose a leg. The blemish may last for the whole of her lifetime, but it is not passed on to the next link in the chain. Errors that affect stick insects but not their genes are not perpetuated. (97)

Dawkins' understanding of replicators (from now I'll omit the 'active, germ-line' bit) clearly allows that there might be non-biological replicators. In brief, the basic proposal of meme theory is that cultural entities might be non-biological replicators in this sense. The upshot would be that one could use the theory of natural selection to explain the distribution of memes throughout a given population at a given time.

Sperber's gloss is that "Dawkins defines "memes" as cultural replicators propagated through imitation, undergoing a process of selection, and standing to be selected not because they benefit their human carriers, but because they benefit themselves" (2001 163). Common examples of purported memes include words, tunes, origami figures, and belief in an afterlife.²⁸ Everyone presumably agrees that there are many such cultural entities which spread through human populations, some of which spread widely and last a long time, others of which do not. But Sperber points out that Dawkins' characterisation of memes requires more than just this, making the

27 Of course, once we add the transmission of memes, or one of the forms of cultural inheritance discussed by dual inheritance theorists, this claim will need to be qualified. But the example of the injury to the stick insect shows that it applies with respect to an important class of cases.

28 Dawkins says that a meme should be regarded as "a unit of information residing in a brain", with things like words, tunes, and skills being the 'phenotypic effects' of such memes, but this seems to be an open question within memetics (109). (For instance, Sterelny (2006) argues that artefacts are generally the best candidates for being memes.) I find it clearer to speak in terms of what Dawkins calls 'phenotypic effects' than in terms of his 'units of information'.

claim that human culture is primarily composed of memes a much more controversial one.

According to Sperber's gloss, memes need to be replicated via imitation and with high fidelity. This means that we cannot infer just from the fact that some cultural entities do spread effectively through a population that they can be counted as memes in the technical sense. As Sperber puts it, "[t]he issue is whether the relative stability found in cultural transmission is proof of replication" (168). His answer is 'no'. I divide his argument into two parts, to be treated in the next two sections.

'Triggering' versus imitation in the spread of behaviours

The fact that qualitatively similar behaviours spread through a population with causal links between them does not prove that imitation or any kind of 'copying' or 'replication' of behaviour is taking place. Sperber's example is the spread of contagious laughter. Hearing laughter often causes people to laugh.²⁹ In this way laughter might spread throughout a population. However, this is not (or need not be) a case of one person imitating another. Rather, "[t]he motor program for laughing was already fully present in [the person], and what the laughter of others does is just activate it" (2001 168).

In Sperber's terms, laughter 'triggers' laughter, without any bit of laughter being a 'copy' of any other. His gloss on this case is somewhat artificial insofar as various aspects of laughter do get shaped by social factors, and perhaps are even imitated. However, we can grant the logical point that the causal links between chains of qualitatively-similar instances of laughter does not prove that laughter is a replicator. This should be accepted by Dawkins, and is implicit in his own argument that birds' nests are not replicators (1982 98-99).

This point is well taken, but not all chains of causally-linked, qualitatively-similar behaviours consist in triggering of pre-learned behaviours. (Additionally, although by definition a merely triggered behaviour will not be a copy of the behaviour that

29 For some experimental evidence that hearing laughter can be a 'sufficient stimulus' for laughter, see Provine (1992).

triggered it, it might still be a copy of some other behaviour, on some level of analysis. This becomes clearer when we think about the triggering of behaviours which have themselves been socially learned. I'll return to this below.) *Novel* behaviours do spread through human and other animal populations, but through what mechanisms?

The psychological literature on imitation in humans and other animals is a helpful place to look to clarify some of these issues. This is because a significant question in that literature has been how, conceptually and experimentally, to differentiate between different kinds of chains of qualitatively-similar, causally-linked behaviours. It is thought that imitation is in certain respects a fairly sophisticated capacity, and so there is a question of whether a kind of animal that produces behaviour that looks imitative really has this capacity or not.

Psychologists generally distinguish imitation from 'response priming' or 'contagion effects' such as those found in the case of contagious laughter (Hurley and Chater 14-15). One way to differentiate these experimentally would be to 'seed' the chain of behaviours by modelling a novel behaviour not already included in the behavioural repertoire of the subjects further down the chain. If an animal is only capable of being the subject of response priming or contagion effects, then we would not expect it to respond to the modelling of a novel behaviour with a qualitatively-similar behaviour. By contrast, the capacity to imitate is thought of as a capacity which can underpin a creature's acquisition of novel behaviours by observing them.

But if the experimental subject produces a novel behaviour matching that of the model after observing the model, this still does not show that imitation, or what is sometimes called 'true imitation', is going on. Hurley and Chater suggest that true imitation "requires that a novel action be learned by observing another perform it, and in addition to novelty, requires a means/ends structure. You copy the other's means of achieving her goal, not just her goal or just her movements" (14). Given this understanding of true imitation, we can see that processes other than true imitation might also underlie the production of chains of novel, matching behaviours.

There are a number of ways of carving up the options here, and I will just briefly mention some possibilities. One is that observing the modelled behaviour allows the subject to learn something about the way her environment works, which allows her to achieve a goal she already had. For example, one monkey, having observed another pull a plug out of a box to access food, may be more likely to do the same when it is given the chance. But this could be explained by its having learned that the plug can be moved in that direction, and adopting the same way of moving it on the basis of that information (Voelkl and Huber 196). This would not count as ‘true imitation’, but it could be expected to yield an increased likelihood that we would observe chains of matching behaviours between monkeys who had observed each other. Another possibility, sometimes called ‘goal emulation’, is that “you observe another achieving a goal in a certain way, find that goal attractive, and attempt to achieve it yourself by whatever means” (Hurley and Chater 14-15). Again, such a process might increase the chances of seeing chains of qualitatively-similar novel behaviours. To distinguish these possibilities from true imitation and from each other requires more subtle experimental procedures.

The literature on imitation shows, then, that we ought not to assume that just any chain of causally-linked, qualitatively-similar behaviours is generated by processes of imitation, even when those behaviours are novel. Where does this leave us? I actually think that given the right circumstances, processes like goal emulation or affordance learning could be involved in ‘reproduction’ in my sense. However, showing this now would involve a significant digression. So let’s grant that the relationship between these processes and ‘reproduction’ in my sense is complicated, and bracket them. Most important here is the point that the existence of these alternative mechanisms is not generally taken to be in tension with the claim that human beings *can* and *do* sometimes imitate each other in the ‘true’ sense.

Indeed, many people have thought that human beings have a particularly strong tendency to imitate others, more so than other animals. And imitation *can* facilitate very ‘high-fidelity’ copying of observed behaviours. For instance, as has been shown in recent studies of ‘over-imitation’ in humans, we have a tendency in some circumstances to imitate aspects of modelled behaviours even when they are quite irrelevant to the goal the behaviour is geared towards achieving (Whiten et al. 2009).

It has been suggested that this tendency is crucial to our ability to develop versatile tool-using skills, and to acquire “arbitrary, conventional/cultural skills” (Carpenter and Nielsen 226). It is also worth noting that in the realm of artefacts we possess a number of different techniques for producing indefinite, high-fidelity copies of artefacts in a variety of media.

All of this suggests that imitation is a mechanism through which behaviours *can* spread, which gives some *prima facie* support to the memeticist’s conception of culture. This in spite of the fact that it is not the only mechanism via which novel behaviours might reliably spread through a population. However, Sperber argues that imitation is in fact a comparatively *rare* way for such behaviours to spread in humans, and that other, more common mechanisms are importantly different.

Before considering this second argument, it is worth clarifying the relevance of dwelling on the transmission of novel behaviours from one subject to another. Because it seems clear pre-theoretically that not all instances of conventional behaviour involve a subject reproducing a novel behaviour which she has just seen being modelled by another. Once I have learned the Cambridge way of punting, I do not need to imitate anybody else around me when I go punting, and indeed my punting might count as an instance of the Cambridge convention even if there is no one else around whom I *could* imitate at that moment. I don’t need to imitate a model now because I have memorised the Cambridge method, or perhaps it has simply become habitual.

The claim of the previous chapter was that in order for my punting in this way to count as an instance of the Cambridge convention, it must be a reproduction of other actions of punting in this way, which were themselves instances of the Cambridge convention. This result might be secured in something like the following way. Imagine I am currently punting in a way resembling the Cambridge method because I have the habit of doing so. But I developed this habit, let’s say, by imitating the actions of others who were manifesting the Cambridge convention. Because of this connection, we can say that my current actions are reproductions of instances of the convention, even though I am not currently *imitating* anyone. For my broader picture to work, we must be able to extend the notion of reproduction in this and

other ways, and I will elaborate on this below. So, whilst Sperber's concept of 'triggering' was useful for raising the point that we need to enquire into the nature of the processes underpinning a given causally linked chain of qualitatively similar actions, it does not encompass the whole question of reproduction, because it covers over the fact that some of the triggered abilities are themselves ones which have been socially learned.

However, it was crucial to the point I just made that the initial instances in which I follow the Cambridge convention could be treated as reproductions of instances of the convention, in order that the habit which they built up could in turn count as reproducing instances of the convention. My learning the convention needs to count as reproduction in order for the rest of the story to get going. So Sperber's second argument – if it shows that the ordinary mechanisms of human social learning cannot count as 'reproduction' in my sense – remains important, and there is a good reason to focus in the first instance on the learning of novel behaviours. I examine this argument next.

Against imitation

Sperber illustrates his preferred picture of the typical mechanisms underpinning cultural transmission using a thought experiment, designed to show that imitation is nowhere near as frequent in human life as Dawkins and others suggest. I prefer to reserve the term 'imitation' for certain ways of copying *behaviours*, but Sperber's thought experiment and some other examples concern copying things like shapes and stories. I will use 'high-fidelity copying' as the generic term of which true imitation in the psychologist's sense is a sub-category.

Imagine we run the following experiment. Sort your participants into two groups of ten, and number them 1-10. Show Subject 1 in the first group Fig. 1 for a few seconds, then take the image away and ask her to draw it. Show Subject 1's drawing to Subject 2 and ask her to draw it, and so on down the line. Do the same for the second group, but this time starting with Fig. 2.

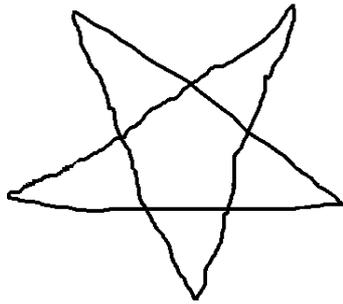


Figure 1

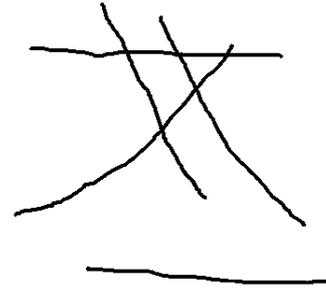


Figure 2

These figures are my own, inspired by Fig. 8.1 and Fig. 8.2 from Sperber (2001).

(Description: Figure 1 is a five-pointed star, Figure 2 is composed of roughly the same collection of lines, but not arranged into any recognisable pattern.)

Sperber anticipates that the drawings produced by subjects in the first group (who started with Fig. 1) will all tend to resemble each other. By contrast, we would expect the drawings produced in the second group to resemble their neighbours somewhat, but we would not expect the final drawing in the sequence to resemble Fig. 2 very closely at all. (If we need experimental support for these hypotheses, we can find some in Scott-Phillips (2017).)

Sperber says that the anticipated difference between the chains of drawings stemming from Fig. 1 and Fig. 2 is explained by the fact that whilst with Fig. 2, subjects will copy the image badly, with Fig. 1 they will not copy it at all, but rather employ “an ability to recognise and re-produce, using, for this, knowledge of the five-branched-star type that they already possessed before encountering the token” (2001 170). Scott-Phillips suggests that his subjects ‘recognise’ the image and use existing knowledge to ‘re-construct’ it (3).

These experiments (thought and lab) are suggestive, particularly in making us question how effective our unaided capacity to imitate or copy a wholly novel behaviour or artefact really is in general. However, it is far from clear exactly what they show. For one thing, both Sperber and Scott-Phillips have chosen for the

‘reconstruction’ image an image that participants will already have acquired specific abilities to draw (for Sperber a five-pointed star, for Scott-Phillips the capital letters ‘ABC’). And indeed the way each phrases it in the quotes above makes this sound like a case of ‘triggering’ a pre-existing ability, akin to the case of contagious laughter from earlier. (Albeit mediated by more conscious control, extrinsic motivations, etc.)

If the causal chains here involved triggering, akin to the case of contagious laughter, then it would seem that there is no direct ‘copying’ going on. Perhaps the images in the Fig. 1 group would have been about the same if, instead of being asked to copy the previous image, the participants had just been told, “Draw a five-pointed star.” However, this suggestion raises the question of where the abilities being triggered came from, given that they, unlike the ability to laugh, are not plausibly thought of as ‘biological dispositions’ (cf. Sperber 2001 168). As I suggested above, that a given behaviour has been ‘triggered’ does not exclude the possibility that it also reproduces some behaviour other than the one that triggers it. So, on the interpretation of these experiments on which it is most clear that the particular images are not being copied, they still do not show that reproduction in my sense is not occurring.

Ultimately, though, it doesn’t seem quite right to say that what’s happening in this case is *merely* triggering of pre-existing abilities to draw a given shape. Perhaps this is not the clearest case to look at here. A better example of Sperber’s general point might be found in his discussion of folktales.

When stories are not written down or otherwise recorded, they get transmitted when somebody who remembers the story tells it to another, who might then remember it and tell it again. The memetic approach allows that this process sometimes produces ‘mutations’, but generally treats accurate re-telling as default – treats the psychological processes involved as straightforwardly replicative. Sperber points out that, for such a complex thing as a story, this is unrealistic. In Scott-Phillips’ terms the relevant processes are ‘reconstructive’, in Sperber’s they are ‘transformative’. Memory incorporates a number of heuristics, and items which do not fit those heuristics are likely to be distorted in systematic ways. This is the basic point behind Sperber’s insistence that in order to explain the distribution of cultural entities in a

given population at a given time, we must pay attention to the details of the psychological processes involved in spreading them (cf. Scott-Phillips 8).³⁰

The role of transformative or reconstructive psychological processes in the transmission of *novel* cultural entities can be seen from the following example. The basic form of the story of *Little Red Riding Hood* is quite stable and widely spread throughout minds, books, etc. But there must have been ‘mutation’ events, such as tellings of the story which omitted to say that the grandmother was rescued at the end, along with Red Riding Hood. Why don’t we find more versions like that today?

The meme theorist’s answer would presumably be that such tellings have tended to lead to replicative dead-ends; they were less likely to be retold than the alternative versions. Sperber, by contrast, suggests that given “universal human psychology” and also “local cultural context”, we can say that some versions of a story have a ‘better’ form than others, and that the form in which the grandmother is rescued as well is better than that in which she is not (1996 108).³¹ The result is that people who heard the ‘defective’ version “are likely to consciously or unconsciously correct the story when they retell it, and, in their narrative, to bring the grandmother back to life too” (108). The suggestion here is not that the re-teller has some idea of there being a canonical ‘correct’ version of the story, but perhaps rather that she will have a sense (tending to be shared by others with her psychological and cultural make-up) that a version where the grandmother is saved is more fitting, or satisfying; or perhaps her memory of the story relies on a (falsified) sense of symmetry – first the grandmother goes inside the wolf, then Little Red Riding Hood, so when Little Red Riding Hood comes out again, so does the grandmother.

30 There’s no commitment to the claim that *only* universal psychological factors will matter, though some researchers exploring Sperber’s suggestions have tended to focus on such factors, e.g. Miton et al. ‘Universal Cognitive Mechanisms Explain the Cultural Success of Bloodletting’.

31 To avoid confusion: I take it that ‘good’ and ‘defective’ here are not really being used evaluatively. Rather, ‘better’ forms are just the ones that are favoured in some way by our psychological and cultural make-up. This may overlap with more evaluative categories, as we see below with the suggestion that a more *satisfying* story will be easier to transmit accurately, but it need not in general.

The existence of such non-random mutations is very important for the explanatory projects that Sperber, Dawkins and others undertake. Broadly, it suggests that a certain set of ‘defective’ entities will tend systematically to be transformed in predictable directions towards ‘better’ forms, whereas ‘good’ forms will tend to remain stable. The question of how great a challenge the existence of non-random mutations poses to more Darwinian approaches to ‘cultural evolution’ is taken up further in Acerbi and Mesoudi (2015) and Scott-Phillips (2017). How does this point affect my account of conventions and repeatable artworks as abstract individuals, though?

Reproduction without imitation

Simplifying a little, my proposal was that being a reproduction of an instance of a given convention is a necessary condition for being an instance of that convention. Copying was given as the basic notion underlying reproduction in the relevant sense, as it was also by Millikan (1984: 19). In the behavioural realm, imitation looked like it would be an obvious process through which behavioural patterns might be copied. But Sperber’s arguments against memetics have cast doubt on the prevalence of imitation and other forms of high-fidelity copying as explanations for the stability of cultural forms, in a number of different ways. Where does this leave my proposal? In brief, my answer is that the proposal is in good shape, because reproduction in the relevant sense does not have to involve imitation or replication. But it will be instructive to elaborate on this answer.

Reproduction in Millikan’s sense does involve the counterfactual dependence of features of the product on (corresponding) features of the model: “[a] pattern has been *reproduced* if its form is derived from a previous item or items having, in certain respects, the same form, such that had the model(s) been different in these respects, the copy would have differed accordingly” (2005a: 3; emphasis in original). I suggested that some such counterfactual dependence is plausibly a necessary, though not sufficient, condition for ‘reproduction’ in my sense. The relevance of the ‘triggering’ model of proliferation was that the nature of the output behaviour did not seem to depend in any interesting way on the nature of the input behaviour (the

trigger). The output of imitation, by contrast, can be regarded as completely determined by the nature of the model, and as being suitable to produce copies of any kind of model. (Both of these are idealisations, but they capture the way that Sperber talks about imitation, and I think this idealisation is harmless for present purposes.)

The processes which Sperber thinks actually drive the proliferation of cultural entities do not count as mere ‘triggering’, but they do contrast with imitation in both of these key respects. First, many of them are ‘domain specific’, with different capacities being brought to bear on different kinds of stimulus (2001 171). Second, their operation is determined to a great extent not just by features of the stimulus (the behaviour being modelled) but also by “information already available in the system” (171). What are the implications of these claims?

The first point does not seem to constitute a problem in and of itself. Of course, if the ‘domain specific’ abilities being brought to bear are as specific and innate as the ability to laugh, then there may be a significant sense in which reproduction is not going on. However, as long as the domain specific processes are sensitive to *relevant* differences *within* their domains, there is no general obstacle to reproduction. We can illustrate this with an example from Millikan:

Sometimes a behavior that occurs because someone has been told to behave so is also reproduced behavior. If a child watches me shake hands with a friend and imitates this behavior with his own friend, his handshaking is a reproduction. And if the child was trained or told by me to shake hands instead of naturally imitating this behavior, his handshaking is still a reproduction – so long, that is, as it is true that if I and others did not shake hands but rather rubbed noses or kissed on both cheeks in greeting, I would have trained or told the child to rub noses or to kiss on both cheeks and he would have done that instead. (1984 21)

What matters about the two different processes of reproduction described here – imitation and the giving and carrying out of verbal instructions – is *not* that they are ‘domain general’ processes which will accurately copy any given input. Some skills will be very difficult to pick up from verbal instruction alone. (And, we should note,

actions with no outwardly-observable component will be impossible imitate.) Rather, what matters is that they are processes through which a novel behaviour can be learned, and that they are sensitive to the relevant aspects of the ‘model’ behaviour – in this case, those features which make shaking hands different from some range of alternative viable greeting rituals. Given that they are, it makes sense to see both as processes through which handshaking can be reproduced.

The relevance of Sperber’s second point seems to be twofold. On the one hand, processes whose output is determined not only by features of the model, but also by ‘information already available to the system’ might tend to be *insensitive* to certain features of the model. With respect to such features, the output will not counterfactually depend on the input. On the other hand, though this is perhaps just a special case of the first hand, this insensitivity may tend to introduce important qualitative *variation* between the model and the output. Insofar as variation raises a distinct problem, it falls under the objection considered in the next section, so I will just discuss the point about insensitivity here.

If we understand reproduction partly in terms of counterfactual dependence, does the insensitivity of a psychological process to some features of its input disqualify it from being a reproductive process? Again, the answer is ‘not necessarily’. And again we can lean on an example of Millikan’s:

A reproduction is never determined by its model in all respects. The photocopy has a white background not because the original did, but because white paper was put in the copier’s paper tray. (2005a 3)

The copier is not sensitive to background colour in this way, and this of course remains true even when both the original and the copy have white backgrounds. It seems as though Millikan’s basic point must be true here. For one thing, two items are rarely going to be exactly alike, even at a ‘macro’ level. For another, it seems likely that most media will determine some features of the ‘copies’ which are made in them, meaning not all features of the copy will depend on those of the original. Whether and how this matters will depend on what is being copied. For some cultural entities, putting it through a photocopier will produce another instance, for others it will produce at best a replica. Whether this is so will depend on, among

many other things, whether the background colour of the entity in question is relevant. You can produce another handout via a photocopier, but not necessarily another copy of a photographic artwork.

But this might not seem to deal satisfactorily with the full scope of Sperber's point. The discussion of *Little Red Riding Hood* above encouraged us to think that *whether or not* the original story contained the saving of the grandmother, the re-teller would be likely to include it. Yet such plot points are a crucial part of what makes up the story. So if we could generalise Sperber's counterfactual to the other key plot points, it can start to look like each retelling is actually just a matter of the teller making up a wholly new story based on their sense of what makes a good tale, but they all happen to coincide. It would be hard to reconcile this picture of what is going on with the idea that the story is being reproduced in our sense.

It seems, though, that this extreme picture just can't be right, as we can see from an experiment by some collaborators of Sperber's (Miton et al. 2014). They initiate a chain of re-tellings of a novel story which is initially told by the experimenters to the first participant in the chain. They observe that over the course of successive re-tellings the story tends to morph in certain predictable ways, which they explain as a process of "attraction based on universal cognitive mechanisms" (8).³² That is, they think that precisely the kind of non-random mutation described by Sperber is occurring. But a number of features of the methodology are telling here. First, the fact that the stories appear to change fairly gradually in the relevant respects, and many other points stay fixed throughout. Second, the fact that they 'seed' into the story some elements which are thematically quite close to the elements they are expecting to find in the eventual mutations. Third, the fact that they give the participants a story to re-tell at all, rather than asking them to make up a story from

32 They are investigating the prevalence of bloodletting as a form of medical treatment, despite its general inefficacy. The story-chains experiment tests the hypothesis that 'false positive' stories of accidental bloodletting followed by recovery from illness might systematically get transformed into stories of *intentional* bloodletting which is then attributed a *causal* role in recovery. The experiment appeared to show this kind of transformative trend over successive re-tellings.

scratch. Finally, the fact that the participants were asked to “re-tell” the story they had heard (4).

All of these factors suggest that there are significant dependencies between the content of the original story and the content of the re-telling, even granting the fact that the psychological mechanisms involved in re-telling are not neutral with respect to the content of the story. It still makes sense to think of the activity as one of re-telling, rather than wholesale invention, and we can identify the tellings themselves as versions of the same story. Sperber’s observations about the nature of the psychological processes involved may encourage us to modify our understanding of the counterfactuals involved in reproduction, but consideration of this example suggests that the kind of non-random mutation found in ‘cultural attraction’ models is consistent with talk of reproduction in my sense.

Now we have a somewhat clearer view of what ‘reproduction’ needs to involve when an agent reproduces a novel behaviour based on observation of a model (or based on verbal instructions, etc.), we can say a bit more about ‘triggering’. I have granted that in Sperber’s sense, if one behaviour ‘triggers’ another, as in contagious laughter, the triggered behaviour should not be thought of as a direct copy or reproduction of the triggering behaviour, because it is mediated by a pre-existing ability or disposition to produce that very behaviour. However, to say that it is not a reproduction of the behaviour which triggered it does not imply that the triggered behaviour is not a reproduction of anything else.

We can return to the example of learning to shake hands as a greeting. The child, after having practised following the instructions a number of times, comes to develop a habit of stretching out her hand when greeting someone else who does the same. She still counts as reproducing the same behavioural models – despite the fact that the *process* involved now resembles what Sperber called ‘triggering’ – because she has the habit she does due to having learned the greeting she did. It seems right to say that if she had been taught to kiss on the cheek instead, her habit now would be to kiss on the cheek as well. And then, in turn, if her habit now was to kiss on the cheek, she would have kissed the other person on the cheek instead of shaking her hand.

Returning to what we know about repeatable artworks, we can see that a wide variety of processes of reproduction will allow us to create new tokens of a given artwork. Psychologically and physically, tracing an outline or casting a new figure have very little in common with sight reading a piece of music. Yet all can be ways of reproducing some type or other. At the same time, it is not just a matter of finding a sufficiently high-fidelity reproduction process for the medium in question. For some artworks a higher-fidelity copying process might not be able to create reproductions, whereas a lower-fidelity one might. We considered some relevant examples in Chapter 1, where we saw that in some contexts any print made from the artist's original woodblock will count as instances of her work, whereas in other cases the prints will need to be made by the artist herself. In a similar way, we can imagine that replicas made through a highly accurate process of scanning and re-printing may not count, whereas prints made from the original woodblock do, even as the block ages, introducing imperfections into the later prints.

Variation within types and nominalism

As I mentioned above, Sperber's recommended ontology for a naturalistic anthropology is broadly 'nominalist' in that it denies that cultural entities are types, or any form of abstracta more robust than a set. (I use 'nominalism' in a loose sense, I hope its meaning is clear enough.) This nominalism is at odds with my claim that conventions, repeatable artworks, and many artefact kinds are abstract individuals. In this section I will discuss an argument for nominalism which is tied into the arguments against memetics discussed above, but which it is useful to distinguish from them. My basic response to it is quite straightforward, but spelling it out will help to clarify some things about the view of cultural entities as created types or abstract individuals, as well as raising some further questions which need to be answered by a full defence of this view.

Sperber advocates a 'modest materialist' ontology for anthropology which is 'non-reductionist', but leans towards something like 'token physicalism', at least for 'mental representations' (1996 13-14). He argues that many everyday concepts – such as local concepts of marriage found throughout the world – which seem

ontologically extravagant should be paraphrased for anthropological purposes in terms of particular physical actions and artefacts and the production of particular mental and public representations (e.g. spoken or written words, pictures, etc.) (22).

As mentioned above, Sperber takes one of the primary tasks of the naturalistic study of culture to be studying and explaining the distribution of mental and public representations in a culture at a time. His ‘epidemiology of representations’ will not be concerned with every single representation found in a culture, however fleeting and idiosyncratic. Rather, it will be particularly concerned with what he calls ‘cultural representations’ – representations which are ‘distributed widely’ and ‘lastingly’ throughout a given population (1996 33).

I will digress briefly to note that there is a type/token ambiguity here which, despite his efforts, is found throughout Sperber’s book. In fact, explicit references to types and tokens are found throughout the literature on cultural evolution, sometimes coexisting with fairly explicit endorsements of nominalism (e.g. Brown and Feldman 22139). It is undeniably useful to speak as though cultural entities (including Sperber’s ‘cultural representations’) are types which have instances. Following Linda Wetzel, this may be taken as reason for endorsing their existence as such (Wetzel, Ch.2). Sperber himself appears to regard this as *just* a manner of speaking, with terms like “the Bororo myth of the bird-nester” being either non-referring, or else referring to a set of particular representations (i.e. the set of spoken, written, or memorised ‘tokens’ of the story) (1996 27-29; 33). However, his arguments for taking this position do not stem *solely* from a generalised nominalism, which would be uninteresting in the present context.

Sperber’s two strands of argument against reifying cultural representations both relate to the earlier point that the mechanisms of proliferation of cultural entities are not in general high-fidelity. Now the focus is not, as it was in the previous few sections, on the nature of the process itself, but rather on the variation among the *products* of this process. In particular, if the processes of proliferation are not replicative, the particular representations produced by these processes will often vary in important ways. Sperber suggests that the urge to reify cultural entities comes from an exaggerated view of the resemblance between the distinct particular

representations that are taken to be the tokens of a cultural type (1996 100). At the same time, he implies that reifying cultural representations will distort their content by attempting to provide a ‘synthesis’ of what are in fact different contents (27).

I will address both of these arguments over the next few sections. The account of abstract individuals I developed in Chapter 1 does not posit types simply based on qualitative similarity. The idea of a ‘reproductively established family’ of concrete particulars provides a principled basis through which we can understand how an abstract individual collects its instances, which does not depend on qualitative resemblance among its instances. So it need not be *ad hoc* to say that an abstract individual has instances which vary qualitatively.

However, there remains a question of determining the possible scope of qualitative variation amongst instances of an abstract individual. In the next few sections I explore this question by returning to the literature on repeatable artworks as created types or, as I have suggested instead, as abstract individuals. I will argue that they can admit of qualitative variation amongst their tokens in a number of different ways. Exploring variation among tokens of a repeatable artwork will suggest that there is nothing about abstract individuals *per se* which requires there to be little or no qualitative variation among their instances.

How much variation is possible amongst the tokens of a type, or the instances of an abstract individual? The brief answer is that the tokens of a type need not in general be qualitatively identical, even with respect to what might be thought of as their core features. This point is made by various people in the literature. Wetzel observes that you can have misspelled or mispronounced words (Wetzel Ch.3). Walters and Davies point out that there can be ‘imperfect’ copies of artworks, where imperfection here encompasses not just qualitative difference *per se*, but what we might think of as *significant* qualitative difference – e.g. a performance of a piece of music incorporating ‘some incorrect notes’ (Davies 265; cf. Walters 463-464).

These moves admit the possibility of significant variation amongst tokens of a cultural type by leaning on some notion of imperfection. I am inclined to go further than this. I suggest that, ordinarily, we think that different kinds of cultural productions (conceived as types) admit different degrees of qualitative difference

among their *non*-imperfect tokens. I will illustrate this in terms of a broad distinction between oral and literate traditions.

Sperber, among others, observes that an oral literary tradition will generally tend to produce “more or less faithful versions” of a story which differ significantly from each other (1996 32). In general, word-for-word accuracy will be higher when stories spread through the manual copying of manuscripts, or through the printing press, although this fact is not determined solely by the relevant technologies, independent of broader social and institutional factors (cf. Goody Ch.1). At the same time, each method of proliferation will tend to introduce its own patterns of copying-error, and the kinds of representation that spread accurately and stably in each medium will differ accordingly (53-54).³³ All of this is true and important. However, this picture overlooks significant differences, going beyond mere fidelity, which came about with the spread of literacy, printing, etc.

Changes occurring in the transitions from oral to written literary traditions concern: ideas about authorship and authenticity in relation to artistic value; emphasis on the aesthetic particularity of very specific combinations of words; the nature and significance of ‘reference’ to other, distinct works; and institutional factors such as copyright law. These changes brought about corresponding changes in the *nature* of works produced within these traditions, which have a bearing on the question of variation amongst instances of a repeatable artwork.

33 A nice illustration of this is provided by chain letters which were circulated in the 1980s and 1990s. Such letters were typically photocopied until the copies became too faint, at which point they were re-typed and then photocopied until they became too faint again. The degradation of the message introduced by the photocopier was content-neutral. By contrast, the variations (errors?) introduced through re-typing a faint photocopy are much more content-sensitive. A study of a corpus of chain letters belonging to the same broad family notes most variation at points such as names and numbers (relatively arbitrary), or near-homographs where the broader context did not determine which word belonged there (“lost his life” changed to “lost his wife”, etc.) (Bennett et al.). These variations were attributed by the authors to the re-typing stage. (The change from ‘life’ to ‘wife’ was accompanied by a corresponding change in pronouns further down – we can assume that this was the result of human intervention to ‘correct’ the text, rather than the fortuitous result of photocopying (Bennett et al. 80).)

‘The song’ in oral traditions

To substantiate these claims somewhat, we can turn to Albert Lord’s study of oral song and poetry. I will use his work to represent what appears to be a cogent position. I don’t think that this requires all of the details to be correct. Lord was partly concerned in the work I am drawing on with the Homeric poems, but he spends significant time on Balkan epic song traditions, as they still existed into the 20th Century. He is particularly concerned to use this case as a way to wean contemporary readers away from a conception of songs and other poems which has, in his view, been shaped overwhelmingly by literacy (100).³⁴

On Lord’s account, the oral poets of the Balkans compose as they sing. They make no attempt to memorise songs on anything like a word-for-word level (109). Yet some of their performances will be quite consistent on this level from one performance to the next. Their ability to compose metrically as they sing, and sometimes to produce quite consistent performances of the same song on the word level, is due in part to their experience creatively combining various sets of verbal formulas and themes which might recur from one song to another, and within a performance (Lord Ch.2-4).

However, reproducing the words sung on a past occasion in a new performance is not the aim of the true oral poet, according to Lord. She has no concept of a ‘fixed text’ for a song which she recreates in her singing of it (Lord 99; cf. Finnegan 12). That is not even an ideal that she would strive for. This is quite unlike the typical performance of a poem or song in a literate tradition. As a general rule in literate traditions, if one performs a given poem or song one tries to say the same words in the same order as they would be found in some ‘canonical’ written version or performance. (This need not always be the case.) Intentionally deviating from this version will raise questions about whether the poet is trying to modify the canonical text, or perhaps create an alternative version. Such questions appear not typically to

34 It is worth noting that to say that a singer is working in an oral tradition need not imply that she cannot read. The difference between oral and literary traditions, in Lord’s account, goes deeper than this, so oral traditions can coexist with literacy.

be raised by variation among performances in oral traditions. (Fascinatingly, Lord says that he knows of no cases where an oral poet whose song had been transcribed asked afterwards to go back and change any of the words or lines that they had sung (128).)

Yet the oral poet does not just sing whatever she feels like. She is constrained by the song she takes herself to be singing. She will sing ‘the same song’ on many occasions, and singers learn songs from one another (102). This claim can be reconciled with that of the preceding paragraph, because the identity of the song is not thought to reside in a particular form of words, and singing it does not involve trying to recreate such a form of words:

The song we are listening to is “the song”; for each performance is more than a performance; it is a re-creation. [...] This concept of the relationship between “songs” (performances of the same specific or generic song) is closer to the truth than the concept of an “original” and “variants.” (101)

Lord suggests that perhaps the oral poets (at least those he studied) think of the identity of a song as primarily given by “a flexible plan of themes, some of which are essential and some of which are not”, rather than by a fixed text (99). But this ‘plan’ itself is not identical to the song:

the song is the story of what someone did or what happened to some hero, but it is also the song itself expressed in verse. It is not just a story; it is not merely a tale divorced from its telling. (99)

So it seems that the song itself can be found in each of its performances, and whilst there may be better and worse performances, nevertheless each is in one sense on a par with all the others: none, not even the first performance, has any claim to be ‘canonical’. There just seems to be no place for the idea that instances of the song need be qualitatively identical, in the oral traditions Lord describes.

The upshot of the cultural changes brought about in part by literacy appears to be that, as a general rule, tokens of a literary work in a literate tradition *need* to be much more similar to other tokens of the work than tokens of a story in an oral literary tradition. Even when there are relatively small differences between tokens of a

modern work it might make sense to talk of there being distinct ‘versions’ of the work, in a weighty sense.³⁵ By contrast, it doesn’t seem to make sense to talk about the ‘canonical’ version of an oral work, in the way that it can for a written one.³⁶ Different media of reproduction and different cultural contexts allow or encourage an artist to treat different aspects of the work as having aesthetic significance, and the artist’s views on such matters feed into determining the metaphysics of the work (cf. Wollheim 116-117). We shouldn’t be led astray by the requirements of a particular, narrow, range of works most familiar to us.

Accommodating variation within the account of abstract individuals

We saw in the previous section that we can make sense of the reproduction of songs in oral traditions without appealing to the idea of there being a ‘canonical version’ of the song. This means that tokens of the song can vary significantly from each other, whilst still counting as non-defective reproductions of the same song. This deals with both of Sperber’s concerns about the possibility of variation among tokens of a work. It does this by i) providing a principled basis for the abstract individual to collect its instances, which does not depend on their qualitative resemblance to each other, without ii) relying on the existence of a distorting ‘synthesis’ of versions, or a ‘canonical version’ which each could be seen as approximating.

Stated as such, these claims seem coherent and defensible, and they appear to be consistent with the account of repeatable artworks and conventions as abstract

35 Viewed in this way, ‘versions’ of a novel are perhaps akin in some ways to ‘productions’ (as opposed to performances) of a play, which it is very natural to think of as something like a ‘sub-type’ of the type that is the play itself. Vanessa Carr argues in her forthcoming PhD thesis that we should understand versions of a novel in something like this way. Part of my point here is that at least within non-philosophical discourse about artistic works, it is less obvious that talk of ‘versions’ of, say, a folktale has to carry this kind of ontological or structural weight.

36 Some theories of oral culture might have a role for the idea of the ‘ur-story’ somehow glimpsed through each of its extant variations. On that kind of view of folk traditions it would make sense to ask for a canonical version, but that view seems mistaken to me.

individuals which I developed in Chapter 1. However, it remains possible that there might be independent features of that account which would push us towards saying that instances of an abstract individual cannot vary as much as I have suggested. Wollheim's account of created types appears to include such a feature. In this section I will argue that Wollheim's own account contains tensions on this point, and that it does not provide independent reasons for thinking that the feature he includes must be a general feature of created types or abstract individuals.

In many ways Wollheim's account of artworks is ill-suited to cover cases of, as it were, 'unauthored' works, like *Little Red Riding Hood* (thinking for now of its pre-Grimm form). More troublingly, his account of the properties that tokens of a repeatable artwork must share with each other, if we are to identify the work itself with the type of which they are tokens, conflicts with my claims about variation among tokens of oral works. However, part of what I have said about works in oral versus written traditions is structurally similar to what Wollheim says about the difference between repeatable artworks which 'essentially involve interpretation', such as plays or dances, and those which don't, like prints or novels. I believe that this latter point should lead Wollheim to modify his own commitments so that they are more consistent with what I have said about oral works. I will elaborate on the worry and my response to it in this section.

Wollheim suggests that a key difference between a universal and a type consists in the nature of the relationship between the properties of a universal (or type) and those of its instances (or tokens). He says that both universals (such as redness) and types (such as the Union Jack) might share properties with their instances that are in some sense 'transmitted' from the instances to the universal or type, or vice versa. Both redness and red things "may be said to be exhilarating", and both the Union Jack and its tokens are "coloured and rectangular" (51). The main difference between universals and types, according to Wollheim, is that "all and only those properties that a token of a certain type has necessarily, i.e. that it has in virtue of being a token of that type, will be transmitted to the type" (51).³⁷ All tokens of the Union Jack are

³⁷ This way of understanding the distinction clearly does not fall out of the way I set up the universal/abstract individual distinction earlier. You might try to take Wollheim's distinction as simply cross-cutting mine, but I think actually it will take more tinkering

rectangular, and so is the Union Jack itself (type), but even if all tokens are made of linen, this does not mean that the type is made of linen because it is not *necessary* that tokens of it be so. This does not hold for universals: redness is not red, though it is necessary that tokens of red are red (51).

The importance of this for Wollheim is to render ‘harmless’ the identification of the work of art (in the case of repeatable artworks) with an abstract object – a type. This move means that in principle we can consider *any* property of a token of a type as a property of the type, except for properties which can only hold of particulars (54). He takes this to mean that we can attribute ‘physical properties’ to types, though presumably not physical properties such as determinate locations. The harmlessness point might then be spelled out like this: as long as all aesthetically-relevant properties are necessary properties of the tokens, all aesthetically-relevant properties will belong to the type, too. Therefore it is aesthetically harmless to treat the work as a type.

Wollheim does not provide a principled basis in logic or metaphysics for his understanding of the distinction between universals and types in these terms. The harmlessness point seems to be motivated by a substantive claim about the role that ‘the work’ plays in our appreciation of repeatable artworks. The implications of this point are, however, potentially damaging to my claims about repeatable works in oral traditions.

Wollheim’s claim about the properties a token of a type has necessarily would seem to conflict with my suggestion from earlier that one can have tokens of a work which differ with respect to some of their aesthetically important qualities. It may be that Wollheim would want to acknowledge this point, leaning as Walters and Davies did above on the idea that such cases will be explained by one or more of the tokens being ‘imperfect’ realisations of the type. However, my remarks about stories and songs in oral traditions seem to pose a more substantive challenge to Wollheim’s

than that. I also note here that much of what he says about the properties possessed by a type does not seem to apply to things like conventions. All of this would require further work to sort out. Finally, Wollheim’s talk of necessity here raises important questions of scope and modality in general. Addressing these would take us beyond the current discussion, so I have just attempted to adhere to Wollheim’s formulations.

claim, because in denying the applicability of the notion of a ‘canonical’ version of an oral work, I seem to be denying that the type which is the story itself is *determinate* with respect to all of its aesthetically important qualities. This appears to conflict more directly with Wollheim’s harmlessness claim.

However, this is not the whole picture. Wollheim distinguishes among repeatable artworks between those which do, and those which do not contain an ‘essential’ element of ‘interpretation’.³⁸ Works of performing arts – “operas, plays, symphonies, ballet” – contain such an element (55). Novels do not. The kind of interpretation he has in mind is involved in *producing* a token of a work, not the kind involved in appreciation (the latter is probably involved in appreciation of all works). Wollheim suggests that “for these purposes interpretation may be regarded as the production of a token that has properties in excess of those of its type” (55). This cannot be quite right, however. Tokens of a type will often have properties in excess of those of its type, according to Wollheim’s account, but this need not be a matter of what he describes as ‘interpretation’. An example has already been given – the Union Jack is not made of linen or any other specific material, yet tokens of the flag must be made of some material or other, or at least exist in some medium or other.³⁹ But producing a token of the Union Jack is not supposed to involve interpretation, as Wollheim understands it.

The same point can be made regarding repeatable artworks. Producing a token of a novel need not involve interpretation in Wollheim’s sense, but it does involve

38 The ‘essential’ point is left unexplored by Wollheim, but it appears to have some counterintuitive implications. Producing a token of a symphony by playing an .mp3 of it does not involve ‘interpretation’ as Wollheim characterises it, so either interpretation is *not* ‘essential’ to symphonies, or else recording a symphony creates a new type, distinct from the type which is tokened in a live performance of the symphony. The latter point could be accommodated in a variety of ways, but it is at least superficially surprising. I thank Polly Mitchell and Julian Bacharach for prompting me to think about this kind of case.

39 It is worth mentioning that this point is another source of legitimate qualitative variation. among tokens of a given type, though not one which is of much interest, here.

generating a particular with properties ‘in excess of those of its type’. Most novels (type) do not have a particular background colour or font size, but printed tokens of them must. There is, however, an intuitive contrast between saying that all printed novels must be printed on paper of some colour or other, which is not determined by the type, and saying that a line in a play must be delivered with *some* expressive quality or other, which is not determined by the type.

In the former case, the requirement that the token be more determinate than the type seems generally to be merely an accident of the medium; if it were possible to escape the need to have a determinate background colour then it would make no significant difference.⁴⁰ By contrast, although the play itself does not determine the expressive tone with which a given line should be read, this does not mean that we can abstract away from expressive tone altogether. It seems essential to the aesthetic quality of the work that the line (token) be given some determinate tone. I take it that this is Wollheim’s real point in talking about interpretation, but the difference between the two cases is not adequately captured by his gloss. If I’m reading him correctly, it would have been better to say “interpretation may be regarded as the production of a token that has *aesthetically crucial* properties in excess of those of its type.” This is what seems to capture the difference between (most) performing arts and (most) contemporary literary ones.

Wollheim’s take on types raises a number of questions which I have not addressed here. But if we can grant the basic point that – as I’ve put it – a type might be indeterminate with regard not just to its ‘accidental’ but also to some of its ‘crucial’ features, then we can accommodate most of what I want to say about variation amongst tokens of a cultural entity relatively straightforwardly. Granting this may require a reformulation of Wollheim’s harmlessness point, but as I have just shown, this seems to be required anyway by what he says about interpretation in the performing arts.

40 This is not to say that it is impossible to give considerations like background colour crucial significance in creating a novel. I’m trying, though, to avoid weighing the text down with the endless caveats that artistic innovation could necessitate.

One might worry that if we drop or modify the harmlessness point, we lose any role for what we have been thinking of as ‘the work’ itself (i.e. the type or abstract individual), and that what we really have is just a set of distinct non-repeatable artworks (what we had been thinking of as the ‘tokens’). This would be too hasty. As Millikan points out in the case of species of animal such as dogs, that there are in fact no features shared universally by all dogs (and, we might add, no ‘ideal’ dog that they are all trying to approximate) does not mean that we would do well to dispense with our concept of the species and to approach each individual dog just on its own terms. There is still plenty that one can learn about dogs, and it is hard to imagine how we could approach particular dogs without applying such knowledge to them (2005c 107-108).⁴¹

Something similar may be true of repeatable artworks. Wollheim might be right that the role ‘the work’ plays in our appreciation of some works, or some kinds of art, requires something as strong as his harmlessness claim. But the role of ‘the work’ in other cases might be less demanding, consistently with there still being a role for it to play, over and above the role of its instances, considered as concrete particulars. (Perhaps a more accurate way to put it is that what it takes for identification of a work with a type or abstract individual to be ‘harmless’ will vary depending on the work.)

Wollheim’s harmlessness point does not, anyway, appear so compelling in relation to our central case here, which is conventions. So despite initial appearances, Wollheim’s substantive considerations about what we need a theory of types to do for us in the ontology of art do not generate new problems for the idea that there might be significant qualitative variation amongst tokens of a given created type or abstract

41 We can in fact go beyond this. Numerous people have pointed out the relevance to debates concerning essentialism about species of the fact that many biological species are polymorphous. Mohan Matthen makes the interesting claim in this context that polymorphisms in some species can be thought of not as a simple clustering of variations within the population of a species, but as an important structural element of the species itself (Matthen 144-148). In the artefactual realm, it has been noted that there are many different ways of pronouncing the same word, and that there is therefore no simple mapping of spellings onto pronunciations (e.g. Wetzel).

individual. This can be accommodated within Wollheim's broader view of the relationship between types and their tokens, as long as types can be indeterminate with respect to some of their important properties. As mentioned above, when we approach the existence of types via the notion of chains of reproduction, as I have done here, the possibility of this kind of indeterminacy in types does not seem particularly mysterious. If there are independent logical or metaphysical problems with this proposal lurking in the background, Wollheim's account of types does not seem to give us any reason to think so.

Cultural context and reproduction from multiple models

Sperber suggested that the relative stability of tellings of *Little Red Riding Hood* over time was due less to the fact that divergent tellings tend to lead to replicative dead-ends, and more to the fact that the mental and cultural make-up of the tellers will tend to introduce non-random mutations which 'correct' the story, bringing it back in line with what is (for this reason) the stable version. Lord describes something very much like this in the Balkan oral poetry traditions which he studied.

Lord had recordings of two performances of the same song by the same singer (Sulejman Fortić), made more than 15 years apart, in 1934 and 1950. The two performances are quite similar, although Lord notes two significant changes. In one case, an apparently significant detail is changed to reflect contemporary circumstances, but Lord describes Fortić's 'feeling' for the traditional themes as pushing him to include an additional alternative detail which preserved the overall structure of the theme (118). The other example, perhaps more interesting, is that the recording in 1934 has an unhappy ending, the only version of this song Lord knows of which has such an ending. In 1950 Fortić gives it the more traditional happy ending. Lord remarks:

The later version of Fortić with its happy ending would seem to indicate that the singer had heard the more orthodox versions of the song since the earlier singing and had brought his own version more into conformity with them.

The later version would then be an example of the corrective influence of

the tradition. When a singer deviates too greatly from the traditional version of a song in regard to an essential theme, he is brought back into line, not by the audience but by the songs and singers of the tradition itself. (118)

The singer's training, which produces a 'feel' for the structure and coherence of certain themes, as well as his having heard other performances of the same song, help to maintain certain important common features from one performance of the song to another (once again, without any attempt at memorising a 'canonical version'). And this is no accident. Training, practice, the teaching and sharing of songs are part of the broader cultural context in which these songs are reproduced. This context is part of the explanation for why we don't *need* reliable, domain-general, high-fidelity imitation to keep a song alive in an oral tradition. This case also helps us to see why the existence of non-random mutation need not be in tension with the notion of reproduction.

It might be worried that by invoking a distinct tradition or practice as part of the explanation of what counts as the 'reproduction' of a given song, we are inviting a regress or some vicious form of circularity.⁴² In fact, I do not think that there is a problem here.

In making a reproduction we are often reproducing more than one model, and sometimes these will be hierarchically arranged. Think of reproducing a poem. A reading or inscription of a poem will only count as such if one succeeds in reproducing the *words* which make it up. The chain of instances of *Kubla Kahn* all trace back to Coleridge's hastily-written manuscript (or perhaps to the dream during which he reportedly composed the lines). Instances of the poem contain, amongst other things, instances of the word 'Xanadu', which thereby belong to a chain of reproductions of that word which traces back much further than (albeit still passing through) Coleridge's manuscript. There does not seem to be anything particularly troubling about this sort of case, despite the fact that one can only reproduce the poem by reproducing an older tradition. A distinctive worry about the oral songs case might be that the tradition within which it is possible to reproduce the songs can itself be understood only through its role in the reproduction of these very songs,

42 Julian Bacharach raised this concern when I presented an earlier version of this material.

which could introduce some kind of circularity. However, I don't see any reason to think that that is the case here.

It is worth noting, however, that it does not seem necessary that there be any specialised scaffolding such as that provided by the tradition among the Balkan singers if an oral work is to be reliably reproduced other than through domain-general, high-fidelity psychological mechanisms. Olivier Morin has argued recently that given our poor capacities for high-fidelity replication, the most important consideration for the stable reproduction of cultural entities (bracketing forms of codification such as writing) is the *frequency* with which the entity is manifested. He points out that repeated exposure to a cultural form can allow us to reliably reproduce it, even when we would have no hope of gaining such an ability from a single exposure (2016b). This route does not rely on specialised cultural scaffolding, although there will presumably be interesting stories to tell about why the form gets produced so often.

Morin's suggestion prompts one final objection from Sperber, however. Sperber remarks,

[i]n general, if you are serious in describing bits of culture – individual texts, pots, songs or individual abilities to produce them – as replications of earlier bits, then you should be willing to ask about any given token cultural item: of which previous token is it a direct replica? (1996 104)

He then goes on to remark that “[i]n most cases [...] you will be forced to conclude that each token is a replica not of one parent token, nor (as in sexual reproduction) of two parent tokens, nor of any fixed number of parent tokens, but of an indefinite number of tokens some of which have played a much greater ‘parental’ role than others” (104). (As mentioned before, I interpret Sperber's talk of ‘tokens’ here as consistent with his nominalism about cultural entities; he appears to be using the term just to refer to what I am calling concrete particulars.)

Sperber suggests that this observation undermines the ‘meme model’ of cultural entities. I am inclined to agree that his point causes trouble for the concept of ‘replication’ which appears to be at play in memetics. However, I am not interested in

that question here. I just wish briefly to show that we can grant Sperber's claim that many 'token' cultural entities will be reproductions of multiple distinct token entities, consistently with the claim that this form of reproduction can underpin the existence of cultural entities as abstract individuals or created types. (Hence we can make use of Morin's point about the stabilising effect of frequent exposure.)

The significant point here is that a process which draws (in a particular case or in general) on a number of distinct 'models' for its output can still be a process of 'reproduction' in my sense. Take the example of learning words. There will be some words that we need to encounter more than once before we learn them. Suppose I hear the word "redundant" four times before I pick it up and am able to produce tokens of it myself. We can imagine that all four exposures were necessary, given the circumstances, for me to learn it; there is no single instance that I am using as a model when I go on to use the word myself. Nevertheless, there is no obstacle to saying that what I go on to do is to *reproduce* the word "redundant", if all of the models on which I am drawing were themselves also tokens of that word.

There is a related scenario which does appear to cause trouble. I might hear two utterances of two distinct words, and take them to be utterances of the same word (perhaps they sound the same; perhaps there are two people with 'the same name'). If I draw on these two experiences to produce a token word, which word is it? Whether or not we say, following Kaplan, that "nothing whatsoever is being said" in such a case, it seems certain that *something* has gone wrong here (Kaplan 109). However, the problem seems to be that the two 'tokens' the speaker aims to reproduce are tokens of different words, not the mere fact that she is drawing on two numerically distinct tokens. I conclude that in general there is no reason to think that reproduction in my sense cannot draw on more than one model.

Conclusion

This chapter has been concerned with Sperber's arguments against memetics, and more broadly with his arguments for nominalism about cultural entities. At the core of both arguments are his empirical claims about the diffusion of novel behaviours

and other cultural entities through human populations. In particular, he claims that imitation, and other domain general, high-fidelity processes of replication do not predominate in cultural diffusion.

I have suggested that we should accept these empirical claims, and their negative implications for Dawkins' programme for memetics. However, I argued that we can accept these claims whilst still finding room for the kinds of sensitivity to behavioural models which is necessary for a process of reproduction of the sort described in Chapter 1. Even when novel behaviours spread in part via domain-specific psychological processes which introduce non-random mutations in the resulting behaviour, such processes can be sufficiently sensitive to enough of the relevant features of the model that they can count as 'reproductions' in my sense. I concluded that my account of conventions as abstract individuals created through the reproduction of patterns of behaviour is consistent with Sperber's empirical claims.

Sperber also had a number of further nominalist arguments which took these empirical claims as their basis. In particular, he suggested that treating a cultural entity as something like a created type or abstract individual would require either for its 'tokens' to be highly qualitatively similar to one another, or else for the type to be identified with some 'canonical version' or 'synthesis' which would therefore distort the contents of its various tokens. Neither of these consequences seemed attractive.

However, I have argued that when we understand abstract individuals in the way I described in Chapter 1, we can avoid both of these consequences. When we understand abstract individuals through the notion of 'reproduction', we do not need to rely on a high degree of qualitative similarity among their tokens. In addition, we saw in the discussion of Lord's account of oral traditions that the relevant notion of reproducing an artwork does not presuppose that there must be such a thing as a canonical version of the work. If, as I argued in the discussion of Wollheim, abstract individuals or created types can be indeterminate in some of their properties, then treating oral songs as abstract individuals does not reintroduce the need for a distorting synthesis. I suggest that the same applies to conventions, where there is

even less independent pressure to think conventions are built around canonical versions of the relevant patterns of behaviour.

Chapter 3. Reduction and Other Issues

Millikan's account of conventions is self-consciously reductionist. To adapt the title of her first book, she wants to show that convention is in many respects a 'biological category'. The concept of reproduction that figures in her account of language and other conventions is in key respects the same as the one she would apply to, for instance, the asexual reproduction of amoeba. One of the great values of her work is to show that there are interesting similarities and continuities across domains that often seem to be held categorically distinct in philosophy.⁴³ We ought to appreciate these continuities where they exist, and as will become clear in the next chapter I want to keep hold of the thought that the existence of conventions and norms does not essentially depend on, amongst other things, our representation of them. Nevertheless, I do not accept what I take to be some of the implications of Millikan's reductionism. In this chapter I will show that her account can accommodate much more than might be thought, but I will ultimately argue that features of it are incompatible with certain important cases of convention.

I aim for my account of conventions, rules, norms, and artefact kinds to be broadly naturalistic. This will mean allowing our individuation of cultural entities, and our understanding of what counts as 'reproducing' them, to be sensitive to, among other things, the results of social-scientific enquiry. However, unless these forms of

43 In light of this, it seems to me that Buzl and Kreuter's criticism of Millikan's account of convention as being over-inclusive is misplaced (421-422). If Millikan is correct, then there *is* significant continuity between the conventions which make up natural languages, and many things that occur in the non-human world. That is, we can produce a good account of language conventions in these terms. But if that is true, then it doesn't look like a bad thing that many non-human animal traditions will count as conventional, according to her account. Criticism of her account on this score seems to require showing that she is *wrong* about these apparent continuities.

enquiry somehow reduce in a serious way to biology, naturalism here does not require us to stick to a concept of reproduction which is suitable for biology.

It would be open to Millikan to reply that she is not committed to preserving all appearances, and that the simplicity and generality of her account of reproduction will make up for its failure to cover certain classes of conventions which would be recognised by ordinary thought. Similarly, Sperber is explicitly committed to developing a fresh ‘natural science of the social’ which he expects to involve some ‘reconceptualisation of the social domain’. I, on the other hand, have been aiming to develop an ontology of conventions which supports much of our ordinary thought about them. So, to some extent, the differences between their accounts and mine are to be explained by our having different aims. It is not possible to fully interrogate these broader projects here, but I will argue that the mere prospect of the fulfilment of a project like Sperber’s is not enough to invalidate mine.

Finally, I raise some examples which seem more troubling for my account, given my own aims. Accounting fully for these examples will have to remain a task for future work, but I will offer some suggestions for how they might be accommodated.

Millikan’s ‘official’ account of reproduction

As I mentioned in the first chapter, in Millikan’s more official characterisations of the notion of ‘reproduction’ she encourages us to think of it largely in terms of the counterfactual dependence of some features of the product (P) on corresponding features of the model (M), where M also helps to *cause* the features P actually has, and would have caused P to have different features, had its own features been different (2005a 3; 1984 19-23).

Millikan’s account of causation does not reduce to counterfactual dependence. She says that the causal link between M and P, and the fact that M would have caused P to have different properties had its own properties been different, should be explicable by appeal to “a natural law or laws operating in situ”, in the particular

case.⁴⁴ So she is not (that particular kind of) reductionist about causation. The point is rather that the laws here could be of any kind. Millikan mentions that on this account a footprint will count as a reproduction (of the shape of the foot that makes it) (1984 21). So will the shape a falling stone leaves in the sand it lands upon. So will a shadow (21).

In Chapter 1 I suggested that we should incorporate some notion of ‘arbitrariness’ into our account of the reproduction of conventions, so as to capture the explanatory and individuating role of precedent. A similar condition, applied to repeatable artworks, seemed to shed some light on the creative role of the artist, and the difference between creation and discovery (although ‘arbitrariness’ doesn’t seem like quite the right term in the case of art). I will now suggest that there is another feature of the kind of reproduction relevant to conventions, which is not captured by Millikan’s official gloss.

According to the official definition given above, then, reproduction will be absolutely ubiquitous not just in the biological but also in the inanimate physical world. However, it is not obvious that this notion of reproduction is of very much interest for understanding the existence of conventions, artefact kinds, or biological species.⁴⁵ For the shape reproduced by a shadow, or an indentation under a stone, it will generally be a complete accident if this forms part of an interesting chain of reproductions.⁴⁶

44 Millikan: “By a law operative in situ I mean a special law that can be derived from universal natural laws by adding reference to the actual surrounding conditions, in this case the conditions surrounding the production of [P]” (1984 20).

45 I want to leave open the possibility that there are important differences between biological species on the one hand, and conventions, artefact kinds, and repeatable artworks on the other. However, it seems plausible that to understand each of these phenomena we need something more than reproduction in Millikan’s official sense, even if exactly what more is needed varies between the cases. Because of this, and because of Millikan’s views about the relationship between biological species and conventions, I will generally ignore the possible differences except where especially relevant.

46 The existence of shadow-puppetry, certain methods of projection, etc. does not undercut this claim.

The role of reproduction in the central examples we have been discussing does not seem to be captured by this thin official definition. We think of conventions as having at least some power to show that it won't be an accident if people keep following the convention. With repeatable artworks, we have already seen that a high-quality replication of an instance of some work may not count as an instance of that work. The reproduction of biological species, in Millikan's account, is built on top of the 'first order reproduction' involved in the copying of genes. However, once again we think of this kind of reproduction as capable of grounding the reproduction of species not because here and now the splitting of this cell causes both copies to share the same structure of DNA, but because this process is, within the broader context, the sort of thing that can sustain itself in the right kind of way over many such splittings, so as to maintain the life of the individual organism and facilitate the growth of new organisms of the same kind.

I suggest, then, that if we are interested in a notion of 'reproduction' that can underpin the existence of interesting kinds or abstract individuals, we need to look to a wider context than that of the laws operating 'in situ' at the moment of reproduction. We at least want some reason to think that the occurrence of episodes of reproduction has some tendency to bring about the conditions for further episodes of reproduction in the future. Certain artworks and social practices may make different demands, at least in some cases. Over the next few sections I will address the first point, before offering some remarks about the second in the final section.

Biological purposes in the reproduction of conventions

To some degree, I think that Millikan agrees with what I have said about her official definition of reproduction. She remarks that certain things counting as reproductions according to her definition, such as mirror images or shadows, are "not reproductions that will turn out to be of any interest" (1984 21). The reason she gives for their lack of interest is that they do not have 'proper functions', according to her account. I am inclined to agree with the assessment of these 'reproductions' as uninteresting, but I am doubtful of her explanation of this assessment.

In later work, Millikan describes the existence of conventions which appear not to have functions; but such conventions need not be uninteresting. As I suggested above, it seems as though one key difference between the ‘interesting’ and ‘uninteresting’ cases of reproduction lies in the presence or absence of a context which makes the occurrence of the reproductive event look more than accidental. It may be that the presence of such a context is necessary for reproduction to ground interesting proper functions, but if it is possible to have such a context without it generating functions then it will be possible to have interesting conventions which do not have functions. (It may also be that Millikan’s remark about functions here was tied more directly to the dialectical moves she was making in that particular discussion of reproduction.)

I will assume that some such broader context is required for a process of reproduction that can ground an interesting ‘reproductively established family’ (in Millikan’s terms) and hence a real species, convention, or artefact kind. There is a view of what such broader contexts are like, which can be extracted from Millikan’s work when she is not adhering strictly to what I have called her ‘official’ account. I will try to sketch this view now. I do not know whether this really is Millikan’s view, but I will try to show that it is an interesting view that is more defensible than it might appear on the surface. Nevertheless, it is not ultimately satisfactory.

The view in question divides roughly into two parts. One tells a story about the grounding of species in the reproduction of sequences of DNA.⁴⁷ One of the key processes here, namely mitosis, will typically count on a case-by-case basis as an episode of reproduction in Millikan’s official counterfactual sense. However, as we can see from considering the occurrence of such a process in the lab, this need not bring about or sustain either individual organisms or biological species. Allowing some human cells to split in a petri dish will not lead to the generation of a human being. If that were the only context such ‘reproduction’ of DNA had ever occurred in, it may not be a very interesting process. When it happens in the context of a human body or embryo, though, it helps contribute to the maintenance of that body, or the reproduction of the species itself, frequently enough to keep the whole thing

47 For simplicity of exposition, I am ignoring organisms which do not have DNA; the story to be told about such organisms is not importantly different.

going.⁴⁸ Given this context, then, the replication of sequences of DNA counts as a process of reproduction in the more interesting sense that I have been describing.

On Millikan's account, this 'first-order' reproductive process can ground 'higher-order reproductively established families' with corresponding proper functions. Human hearts or kidneys do not form a first-order reproductively established family, on Millikan's account, because they are not copied from one another (1984 24-25; cf. Dawkins' discussion of the stick insect, mentioned above). However, human hearts form a higher-order reproductively established family because it is a function of certain human genes to produce human hearts, and such genes do form first-order reproductively established families, as we have seen. The function of such genes is (perhaps among other things) to produce hearts, because it is due in part to their having done so that these genes continue to get reproduced. In turn, it is the function of the human heart to pump blood, because it is due in part to hearts having done so that the genes whose function it is to produce hearts get reproduced (1984 25-31).

Millikan's characterisation of reproductively established families is designed to be recursive, so higher-order reproductively established families can in turn generate reproductively established families of a yet higher order. Each of these families might have its own functions, which will ultimately be derived from the relevant first-order reproductive processes. The first part of Millikan's thought is, then, that anything we might think of as a biological function, or a biological kind, can ultimately be grounded in this way in the replication of sequences of DNA, considered as a process of first-order reproduction. The replication of sequences of DNA should not, however, be viewed simply as constituting, on a case-by-case basis, reproduction in Millikan's official sense, but rather as a process which, given all the relevant details of cell biology, of physiology and terrestrial ecology, tends non-

48 Various processes occurring within the individual organism, and also between the organism and its environment and conspecifics, are involved in maintaining the integrity of the individual and the species over time. For an influential account of how various 'homeostatic mechanisms' might underpin the existence of a 'natural kind', see Boyd (1991). Boyd's views bear an interesting but complex relationship to Millikan's. For more on this, see their exchange on 'historical kinds' (Millikan 1999a; 1999b; Boyd 1999).

accidentally to sustain itself in existence. It is this that makes its reproduction in the interesting sense, and on that point I think Millikan and I agree.

It is important to mention one special case of Millikan's general view about biological kinds and biological functions. Millikan thinks that any human purpose is a 'biological function' in the way just sketched. This includes what we might think of as 'psychological' functions, such as the functions of various mechanisms involved in learning (1984 25), and the purposes we adopt by forming conscious intentions (1990 342-343). That is, all of these functions are ultimately derived from functions generated by the first-order reproduction of sequences of DNA.

Millikan suggests that we should understand the category of 'biological activity' in terms of biological functions. Not just anything happening in a living body will be a biological activity. Rather, biological activities "correspond to normally fulfilled biological purposes, that is, to what the animal [or other organism] does in accordance with evolutionary design" (1990 333). According to this characterisation, even fully reflective intentional action is supposed to count as a 'biological activity' (342). This is something I am happy to largely grant for the sake of argument here, though I do not want to take an actual stand on it.

The second part of the view is that biological activities, so understood, provide the context within which we can make sense of the kind of reproduction of patterns of behaviour that will ground conventions. We have seen that some such context is necessary if a process of reproduction in Millikan's official sense is to count as a process of reproduction in a more interesting sense.⁴⁹ As I understand it, Millikan sees the wider context in the case of conventions as being the biological nature of the species among which the convention exists. Although she sometimes writes of the reproduction of conventions as though it were first-order reproduction, there are

49 The need for such a context in the particular case of conventions can also be seen from the arguments of Chapter 2. Considered just as an instance of a pattern, a given behaviour, idea or whatever else has no intrinsic tendency to secure its own reproduction. It therefore seems plausible that we need to know quite a lot about the biological, psychological, material and cultural context within which it exists, if we are to see a process by which it reproduces in the thin official sense as a process of reproduction in the sense that grounds the existence of a convention.

indications that she thinks of it as really a matter of higher-order reproduction. So, for instance, when a person reproduces the conventional use of a given word, this is due in part to the activity of some mechanisms in her which have the biological function of *reproducing language conventions* (2005b 25). On this account, conventions themselves would be *higher-order* reproductively established families.

I am not certain that this is Millikan's considered view, but it is one worth exploring. It appears, however, to be subject to a striking objection. I will outline this objection, and the resources Millikan's work provides for resisting it, in the next section. It turns out that quite a robust defence is possible. Nevertheless, I will argue that the view that the reproduction of conventions is always a matter of higher-order reproduction, as Millikan understands it, is false. It fails to accommodate two kinds of case, which I will outline in a later section.

The connection of biological purposes to the fitness of individuals

We understand Millikan's proposal to be that all human actions are biological activities in the sense that they have and fulfil biological purposes (the purposes of human actions being 'higher-order' purposes, derived ultimately from the first-order purposes or functions of genes). One might think that if all biological purposes are derived from the first-order functions of genes, then the fulfilment of biological purposes should, as such, tend to increase the fitness of the individual or her genes. If that is the case then all biological activities should tend to increase the fitness of their agent or her genes. But many human actions do not increase their agents' fitness, and do not even appear to have any *tendency* to do so. So, we might conclude, such actions are not biological activities in Millikan's sense. In this section I will suggest that we can avoid this *modus tollens* argument by, consistently with Millikan's framework, resisting the initial conditional.

Biological activity that does not promote the organism's fitness in the particular case can be understood in terms of things going wrong at one or more levels. Sometimes what goes wrong seems to be largely 'external' to the organism itself. The deer, fleeing quickly from a predator, might stumble and injure its leg, leading to it being

caught. Perhaps in the circumstances it would have done better to adopt a more leisurely pace. But this just seems like a case of bad luck. Slightly more interesting are behaviours where the pursuit of a biological purpose is led astray by faulty information (cf. Millikan 2014 67-68). An example might be eating food that makes an animal ill. The animal purposely eats what is in front of it, but it does not purposely eat something rotten that will make it sick. Its doing the latter is explained, perhaps, by its having failed to pick up on the cues that the food in front of it is rotten.

A case that is perhaps *prima facie* more troubling for Millikan's proposal is that it seems animals can systematically, or even (in the case of humans) consciously pursue things which are at odds with their 'fitness', where this is not explicable by reference to a straightforward mistake in the particular case. A good example of how such a thing is possible is provided by Millikan's discussion of food aversion in rats (1990 337-341).

Apparently if a rat gets ill shortly after eating a given substance, it will avoid eating things that taste like that substance in the future. There appears to be some mechanism that will generate an aversion to such substances. The purpose of that mechanism is, ultimately, to prevent the rat from continuing to eat things that will make it ill; this has clear adaptive value. However, according to Millikan it fulfils this purpose *by* generating an aversion to anything a rat gets ill after eating, regardless of whether the food in question actually *caused* the illness. This means that one function of the aversion itself is to get the rat to avoid things that taste a certain way.⁵⁰

We can see how this would be a good strategy, given the level at which the mechanism needs to operate, and the limited information about causality available to it. The upshot, though, is that the function of the aversion has a certain amount of autonomy with respect to the purpose that ultimately explains it (namely, the purpose

50 Certain aversions (say, to things that taste like some chemical newly invented by humans) might exist even though no aversion of the same kind had existed before. So the aversion is not a member of a first-order reproductively established family. Nevertheless, we can understand them as having proper, biological functions in virtue of the function of the mechanism that produced it. For Millikan's account of such 'derived' proper functions, see (1984 40-43).

of avoiding ingesting harmful substances). With a bit of bad luck, a rat might gain a lifelong purpose of avoiding some plentiful and nutritious source of food. Fulfilling that purpose will systematically impede its own fitness, but that purpose is still derived, ultimately, from a purpose which is tied fairly directly to promoting its fitness.

This line of thought also helps us to see how a human being might acquire a biological purpose of using a given word according to local language conventions, without that purpose needing to relate very directly to her ongoing fitness in the particular case. If “the primary function of the human language faculty is to support linguistic conventions” (due in part to the general fitness advantages of being able to communicate with others), then more specific functions, potentially with a tenuous link to fitness, can be derived from this more basic function (Millikan 2005b 25). So, Millikan’s account has considerable resources for defusing to the *modus tollens* argument sketched at the beginning of this section.

Problems for Millikan: cross-species conventions and memes

I don’t want to take a stand on whether Millikan’s account of language use and consciously intentional action as biological activities ultimately succeeds. For the sake of the current argument I am happy to grant the basic gist of that account. What I want to do here is raise two issues with the suggestion that we can lean on such an account to treat conventions as higher-order reproductively established kinds.

The thought was something like this. If basically all human activity is biological activity then perhaps any action through which a human copies a conventional pattern of behaviour can be seen as having *as one of its functions* to reproduce that pattern of behaviour. Perhaps, then, embedding reproduction in the ‘official’ sense within such a structure of natural purposes will allow us to see it as reproduction in an interesting sense, capable of sustaining a convention by making it at least to some degree no accident if a similar process of reproduction happens again.

This is the picture suggested for the case of the conventions underpinning language by Millikan’s talk of the function of the ‘language faculty’, quoted above, and by her

frequent remarks that linguistic forms would die out if their use did not fulfil speakers' and hearers' purposes sufficiently often to make it worth reproducing them (e.g. 2014 71; 2005a 16; 1984 31-32). For present purposes I am happy to grant that this might be an accurate picture, when it comes to typical human language conventions. However, taking this model as applying generally will lead to an overly restrictive account of conventions. There are two kinds of case I want to raise here.

One is the existence of cross-species conventions. Millikan appears to think that there cannot be such things. She claims that the convention of calling a cat by saying "pusspuss" or "kittykittykitty" is not a coordination convention (2005a 23 n4). She seems to be suggesting that the use of these phrases is conventional only because they are passed from one human to another. This impression is reinforced by her remark that using 'gee' and 'haw' to direct horses is a coordination convention, if at all, only *among riders of horses*, rather than between horses and their riders (2005a 23 n4).

These claims would follow if conventions had to form higher-order reproductively established families, as the view described in the previous sections suggested. This is because such families depend ultimately on lower-order reproductively established families which, for the purpose of reconstructing Millikan's view, we can treat as 'bottoming out' at particular biological species. (This was the first 'part' of her view, as I described it earlier.) The result is that there are apparently no cross-species higher-order reproductively established families.⁵¹ (See Millikan's argument that there are no common natural kinds to which the psychological states of individuals of different species belong (1999a).) And so, by hypothesis, there would be no cross-species conventions if conventions are higher-order reproductively established families.

51 There are interesting questions to ask about the relevance of coevolution, here. Additionally, if my suggestion below is correct, then first-order reproduction is found in a wider variety of contexts than Millikan envisages. These considerations would both bear on the possibility of cross-species higher-order reproductively established families. These points deserve further exploration, but I do not need to rely on them for the present argument against Millikan.

This implication of the view that conventions are grounded in higher-order reproductively established families strikes me as false, though. I whistle in a particular way to call my cats to dinner. I did not copy this whistle from other humans, or model it after any call used by cats among themselves. Together, my cats and I reproduce our respective parts of the pattern *one whistles when there's food and the others come to eat when they hear the whistle*. I would not keep reproducing the whistle if my cats didn't respond to it, and they would eventually stop responding if they never found food waiting for them. I could have settled on any of a wide range of alternative calls (none of which in fact work to call my cats, because they are not the conventions we settled on). Given the facts about humans, cats, and our ways of living together, there are good reasons why this pattern keeps being manifested, which rely in part on its fulfilling certain cooperative purposes that I and my cats have. It is, in short, a coordination convention between us.

If that claim strikes the reader as implausible, or as depending too much on uncertain details, consider the following example. Paths represent, in a certain sense, a kind of minimal case of an artefact-cum-convention. Walking a given route tends, given the right conditions, to attract others to walk that route, either by encoding information about how to get from A to B, or by making the way easier or more attractive by clearing away obstacles. This is compatible with a significant degree of 'arbitrariness' about the precise route it takes. In spite of such arbitrariness, paths can reproduce themselves by encouraging the reproduction of behaviour that reinforces the path itself.

If we accept that we can apply the interesting notion of 'reproduction' to paths used by a single species, why not apply it also to paths created and maintained in use by creatures from a variety of species? On this planet, there are plenty of creatures who get about by walking, and who share enough of their lives and livelihoods in common that they can maintain a common path in existence over a significant period of time. Such paths are the product of the purposeful actions of individuals of a variety of species. It is a robust fact about this planet that such paths tend to arise and persist, sometimes for a considerable length of time. Millikan's position would appear to suggest that we have a number of distinct paths that happen to coincide,

one for each species that uses it. This implication seems unattractive and unnecessary.

The second case I want to mention is conventions or artefacts which have *non*-biological functions. Imagine that there is a kind of artefact we make semi-regularly for one-off use. Perhaps it is used in a coming-of-age ritual. Each person needs their own one for the ceremony, but it is not used, or even cherished, afterwards. A number of people know how to make these things, but if there is one around to hand then it can be a little easier to copy it, rather than making one from memory. Given that these artefacts are not used more than once, and that people have no particular interest in preserving them after they've been used, it is no real part of their function for human purposes that they be especially durable. However, if old ones are sometimes copied, those that are around to be copied will tend to be the most durable. This durability will have some tendency to be passed on to new ones which are made from the old models.

Over time, something like a process of 'selection' may well take place, for more durable versions of the artefact. If durability helps to explain reproduction in this way, this could be grounds for thinking of the durability as functional. But by hypothesis this function is not derived from the purposes of the humans who make and use them. It appears to be a largely autonomous function belonging to the artefact kind itself.

If we grant this as a coherent description of the case, then it seems to me that we should acknowledge at least the possibility that there could be 'memes' in something like the sense Sperber argued against: "cultural replicators propagated through imitation, undergoing a process of selection, and standing to be selected not because they benefit their human carriers, but because they benefit themselves" (2001 163). We have seen that memetics provides a dubious picture of human culture in general, but it does seem possible that certain entities might arise that proliferate stably in human minds and behaviour in spite of, as it were, working against rather than with human evolutionary design at most levels. Such entities might have functions which are completely independent of human purposes.

These two kinds of case speak against thinking of conventions and artefacts as ‘higher-order reproductively established families’ in Millikan’s sense, if such families depend in the way I have described on the biological activities of a single species. Instead, I suggest that we should think of many of the activities which sustain conventions and artefact kinds as ‘first-order’ reproduction. The biological endowment of human beings is one part of the *background* or *context* against which strictly human conventions are reproduced. The biological endowments of humans, sheep, dogs, etc. are part of the background against which any one of them walking along a given route can count as the (first-order) reproduction of a path. To see what counts as an interesting process of reproduction in the realm of human culture might require us to bring to bear insights not just from biology and ecology, but also from psychology, sociology, economics, anthropology and the rest.

Naturalistic projects

As I have mentioned above, I take the account I have been developing in these first three chapters to be naturalistic in a broad sense. My account of conventions hinges on the concept of reproduction, and my account of reproduction incorporates some causal elements. It is not by any means a brute relationship between the model and the reproduction. Further, I have suggested that our understanding of when a given process counts as reproductive should be sensitive to the results of historical and social-scientific enquiry.

At the same time, however, the aim has been to provide an ontology of conventions (and repeatable artworks) which fits more-or-less with the understanding of them that is embodied in much ordinary folk sociological discourse, and with existing uses in various social sciences. In practice this has meant that our understanding of what constitutes the reproduction of an abstract individual has been sensitive to the details of the particular cases. As we saw in the description of the changing roles of the sound mixer, or the printmaker, there need not be any mystery to this. However, it does suggest that a fully general account of reproduction in terms which can be applied to any kind of case whatsoever will not be forthcoming.

Millikan's suggestions were of that latter form – aiming to give a fully general account of reproduction. This clearly applies to the thin 'official' account of reproduction in terms of counterfactual dependence of features of the replica on features of the model. But it also applies to the account of higher-order reproduction recursively derived from the first-order reproduction of sequences of DNA. The latter account gives more structure to the account of reproduction, but as we saw in the last section it is too narrow to accommodate the existence of cross-species conventions, for instance.

It may well be that Millikan would be willing to accept this result. She may view the relative simplicity and generality of her account as a virtue sufficient to make up for its excluding certain cases which we would ordinarily take to be conventions. Her project involves a kind of naturalism which is different from the one I have adopted, which is more comfortable revising our ordinary categories. Something similar is true of Sperber's project, and the one suggested by Dawkins.

We saw in Chapter 2 that Sperber is, in part, explicitly concerned with ontological questions (esp. 1996 Ch.1). But he is also aiming to create a new, "natural science of the social" (1996 vi). From this perspective, his concept of 'cultural attraction' is not just one process among many, but is instead the central concept in a projected 'epidemiology of representations' which might make the study of culture continuous with cognitive science (5). In a similar way, Dawkins' suggestion that we understand culture through the replication of memes holds out the prospect (with Dawkins' ambivalent support) of a unified subject matter for "*the* theory of evolution by natural selection", which would be neutral between the evolution of culture and of biological life (Dennett 345; my emphasis).

We should not be surprised if any of these projects were to entail at least some 'reconceptualisation of the social domain', as Sperber puts it (6). It is not part of my task to evaluate these broader projects. What I have done is shown that rejecting the attempt to provide a fully general account of reproduction does not render it a mysterious concept. I have also suggested that rejecting this attempt may be necessary in order to do justice to our ordinary understanding of conventions and repeatable artworks.

It is worth saying something further about Sperber's broader project, though. To soften the blow of his 'reconceptualisation of the social domain', he argues that this reconceptualisation does not undermine the existing expertise of, say, anthropologists. He does expect that a genuine natural science of the social will use a significantly different vocabulary from that which current anthropologists use. However, he argues that this is not a problem because anthropologists' existing technical terms do not carry ontological import. In support of this he offers three considerations.

One is the difficulty of offering a clear definition of a term like 'marriage' which will adequately cover the various practices in different societies which get labelled 'marriage' by anthropologists. A second is the fact that local conceptions of what the anthropologist calls 'marriage' frequently carry specific ontological commitments which anthropologists themselves are not committing themselves to when they talk of marriage occurring in that local context.⁵² The third claim is that in light of these previous two points, the anthropologist's application of the term 'marriage' to some activity in the culture she is studying should be understood as 'interpreting local *ideas*' about what is happening, rather than committing herself to the existence of marriages. Everyone can agree that when the event at the church occurred last Saturday, lots of people acquired certain beliefs about what was happening, and when the anthropologist says "they got married on Saturday", she is offering an *interpretation* of those beliefs.

Sperber's conception of 'interpretation' is not entirely clear, at least to me, but two things stand out about his claim that much anthropological discourse lacks ontological import. One is that the lack of ontological import seems much less compelling with respect to our present examples: 'convention', 'norm', and 'rule'. We might add to the list such terms as: 'right', 'ritual status', and 'song'. Our use of these terms is sensitive to local perspectives, but they do not absorb whatever ontological import might be locally associated with them. Hence, we might say i) that it is a rule among P to φ in C , and ii) that members of P believe that all rules are imposed by a

52 For instance, the Ebelo take marriage to be a union "blessed by the ancestors", and many people in France take it to be a union sanctioned by the Christian God (Sperber 1996 22-23).

divine creator, without committing ourselves to the claim iii) that this rule was actually imposed by a divine creator.

Similarly, we might say that each existing marriage is in fact something like a ritual status, about which people have all sorts of ontologically more extravagant beliefs. The ways in which local marriage practices are often recognised by outsiders who share none of the local beliefs about the divine or spiritual nature of marriage bonds strongly suggest that this is how they are viewed by many laypersons. And it also seems like a good fit with how many anthropologists speak of marriage. Hence, there is ontological commitment, but it is commitment to something like a sociological entity: a ritual status. My account is well-suited to explain such uses of terms like ‘rule’ or ‘marriage’.

The second point is that some conventions exist without anyone representing them. (I describe this possibility in greater detail in the next chapter.) In such cases, there will be no local representations for the anthropologist to interpret. Yet she might still want to speak of there being conventions in such a case. Again, my account of the ontology of conventions allows her statement to be straightforwardly true, if we interpret it as committing her to the existence of a convention in the context she is describing. Sperber does, of course, recognise the existence of some ontologically-committal terms in the anthropologist’s vocabulary. He mentions “colour classification”, and “sex ratio”. But he denies that the use of such terms commits us to “a distinct ontological level of culture”, in part because such terms apply “quite independently of the ‘point of view’ of the people concerned” (i.e. the people to whom the colour classification or sex ratio belong) (1996 23). Unrepresented conventions by definition exist independently of being *represented* by those whose conventions they are, but they also seem to be ‘cultural entities’, rather than merely psychological or ecological ones, and as I will argue in Chapter 5 they need not exist entirely independently of the ‘point of view’ of those whose conventions they are.

If these brief responses to Sperber are correct, then we have reason to doubt his claims that anthropological discourse quite generally fails to be ontologically committal. But if anthropological discourse – at least concerning entities like rules, conventions, and ritual statuses – is ontologically committal, then we *do* risk losing

something if we give up on the existence of such entities and reinterpret anthropological discourse so as to be consistent with Sperber's programme for an epidemiology of representations. If Sperber's programme comes to be realised in greater depth in the future, there may be difficult and important questions to ask about how to square it with more traditional work in anthropology and the other social sciences. However, given the costs associated with it, the mere *prospect* of his programme's coming to fruition should not be enough to lead us to discard our ordinary vocabulary of rules, conventions, etc. This especially seems true if, as I have been arguing over the last three chapters, a respectable realist ontology of conventions is available to us.

I will sum up the argument of this section. It may well be that Millikan, Sperber, and Dawkins would be happy to endorse revisionary views of what exists in the social-cultural realm, in part because they are interested in developing quite general naturalistic accounts of human culture, which would have significant continuities with the study of non-cultural phenomena. I have argued that this aim appears to be in some tension with acknowledging the ontology implicit in much ordinary and social-scientific discourse. But I have also tried to show that an understanding of reproduction which is better suited to our ordinary discourse need not be non-naturalistic in the sense of being mysterious, or untethered from the existing social sciences.

Further questions: the lives of cultural forms, and their deaths

I have mainly focused in Chapters 1-3 on the apparent necessity of reproduction for grounding conventions. In Chapter 2 I looked at the idea that a given instance of reproduction of a kind must be sensitive in some way and to some degree to the features of the model being reproduced. In this chapter I have been suggesting that we need something more besides that sensitivity – namely, some broader context which makes it no accident that reproduction itself might occur again. I also argued against Millikan's attempt to make that context exclusively biological in a certain way.

I have argued that looking at a richer context is necessary in many cases to ground a concept of reproduction which adequately captures much of our ordinary thought about artworks and conventions. I have described examples where details of the artistic process or broader cultural context seem to make an intelligible difference to what counts as reproducing an artwork of a given kind. So, sometimes printmaking is an integral part of the artist's process, meaning only prints made by her hand count as tokens of the work. Similarly, the oral poet's training and the broader cultural context mean that reproducing an oral song is different from reproducing a song in a literate tradition.

There are other cases in which it is less obvious how to justify a concept of reproduction which will secure existing views about the individuation of cultural entities. Given my aim of using the concept of reproduction as the basis for an ontology of conventions and artworks which respects much of our ordinary thought and practice concerning them, this is a problem. (Although I do not take this aim to require saving *all* appearances.) I will outline two such examples here, one artefactual and one behavioural. I suggest some ways in which these examples may be accommodated in my account of abstract individuals, but a fuller exploration of such cases will have to remain a fruitful avenue for further research.

Consider first certain mass-produced, branded goods. Levi's 501s plausibly form a kind. Or, at least 501s made during a certain time period. The process by which they are made counts as a process of reproduction in, for instance, Millikan's sense. But imagine the owner of a denim factory who has a contract to make 3000 pairs jeans of some such line, from a company with loose auditing. Let's say that she is also paid by counterfeiters to make another 3000 pairs of whatever the current style is. Now imagine further that she fulfils these two orders by running her factory with the same materials and the same workers for twice as long as she would need to fulfil the original order, then flips a coin on shipping day and puts half the stock in one lorry and half in another, sending them to her two clients. It seems quite plausible that the jeans that end up going to the 'legitimate' company will belong to a kind analogous to the 501s, but those in the other lorry will be mere counterfeits. Indeed it seems possible, though much less clear, that there is no single real kind to which pairs of jeans from the two batches belong.

It's not as though there is nothing we can say to justify our thought that there is a difference in kind between the 'genuine' and the 'counterfeit', here. We can point to aspects of the legal system, to what happens to the jeans after they leave the factory, etc. The worry is, though, that these factors won't make a difference that will, as it were, show up in a naturalistically-respectable concept of 'reproduction'.⁵³ Insofar as naturalism is a reasonable project, here, I think there is more that can be said in support of the idea that it can respect many of the distinctions we find ourselves wanting to make, such as the distinction between genuine and counterfeit in this example. This looks to be the form that a defence of such distinctions would need to take, if it is to be consistent with my account of artefact kinds as abstract individuals.

This example suggests another interesting question for future enquiry. Imagine the case of the jeans factory but now with a liquid such as Coca Cola. If there could be such a thing as 'counterfeit Coke', qualitatively indistinguishable from real Coke, then it looks like we have something akin to 'created types' of *stuff* here. Coke is not as such countable (though bottles and cans of it are). So the story about how a stuff that is individuated historically might be created will be different from the story we have told about the creation of artefact kinds, conventions, etc., which relied in various ways on the countability of instances of the kind in question. (If we can speak of the 'multiple locatability' of stuffs, it looks like this will be something different from the multiple locatability of biological species or novels.) Would the inclusion of created types of stuff change our picture of the reproduction of kinds? If so, how?

The final case I want to raise runs along the same broad lines as the jeans example, but brings us back closer to conventions and practices. Jonathan Lear recounts part of the history of the Crow people in North America. He is focused on a particular

53 Mohan Matthen describes some cases close, though not identical, to the one I describe here (Matthen 148-149). He ascribes our intuitive individuation of cases like the counterfeit jeans to the role of 'branding' in the individuation of (such) artefact kinds. This seems like a reasonable label. But his use of this example in context implies that he thinks that the role of branding could not be incorporated into an account of the reproduction of (in his example, adapted from one of Millikan's) 1969 Plymouth Valiants. I do not think that anything in his examples show this to be the case.

ritual amongst the Crow of ‘planting a coup-stick’, and the way it structured much of their way of life before U.S. encroachment prevented them from living nomadically (Lear 26-27). Another one of their practices from this time was the Sun Dance. Lear’s claim is that planting the coup-stick lost its meaning once the Crow moved to a reservation, and as a matter of historical record we know that at some point they ceased to perform the Sun Dance. At a later point, some of the Crow wished to revive this dance, but, Lear tells us, “the steps of the Crow version no longer existed in living memory. So the tribal leaders sought out the leaders of the Sun Dance among the Shoshone tribe in Wyoming” (37). But he suggests that, “[o]nce planting a coup-stick loses meaning, so, too, does the Sun Dance. One might still teach people the relevant steps; people might learn how to go through the motions; and they can even call it the “Sun Dance”; but the Sun Dance itself has gone out of existence” (37).

There will of course be room for controversy around Lear’s interpretation of this case. But it suggests something which seems independently plausible for quite a lot of cultural activities, which is that you need to learn it from a ‘living’ model of the thing; a *mere* record of ‘the motions’ through which to go is not sufficient to keep the activity in existence (i.e., to secure the possibility of its being performed again). In all sorts of cultural contexts, there is some intuitive pull to the idea that certain cultural forms exist only as ‘living’ traditions, though this idea is a very slippery one.

This seems to be in some tension with Millikan’s account of conventions. What Millikan’s account would seem to suggest is that when the later Crow perform the steps of the Sun Dance, they are actually following the Shoshone convention. But whilst there may be one sense in which this is true, in another sense it seems false: (perhaps) only the Shoshone can perform the Shoshone Sun Dance. There appears to be a kind of structure to our thought about the life and death of certain cultural forms which is not obviously captured in Millikan’s account of natural conventions. Here we appear to be treating the Dance more like we treat prints made from the artist’s own woodblock.

In case it seems as though this result depends essentially on the ritual nature of the example, it is worth mentioning that similar questions about individuation might arise

in quite different contexts. So, for instance, it seems intelligible to ask (and some academics have asked) whether there is a single practice of ‘Nordic walking’ which is spread across many countries, or whether there are multiple such practices, perhaps one for each country in which it has gained some presence. This question can be raised in spite of the acknowledged fact that all such practices can be traced back historically to a single common (albeit complex) precedent (Shove and Pantzar 61-62).

Others have asked whether acupuncture as performed in Western medicine is the same practice as that traditionally performed in China (Barnes 26-27). The placement of needles in the body in Western acupuncture has uncontroversial historical links with the Chinese tradition, but the ideas and theories surrounding this practice in China have largely been dropped or replaced in Western practice. The associated ideas do in fact play a role in how the needle-placing practice spreads, but how much of a difference do they make to the ‘identity’ of the practice, if the placement of needles remains more-or-less constant through changes in associated ‘theory’?

Once again there may be room for reconciling our thoughts about the persistence-conditions of these sorts of cultural forms with a broadly naturalistic understanding of reproduction. One obvious strategy to try out here is to suggest that the pattern of behaviour being reproduced is broader than it first appears. Instead of understanding some pattern in terms just of what, say, happens before the altar in daily worship, we should understand it as a pattern that begins with the moves made when the worshipper is initially baptised into the church. This would help to explain at least why someone walking into the church off the street and copying the worshippers might not count as doing what they are doing. And in a similar way it might explain why that form of worship could not be resurrected simply by copying the moves seen on a video recording of its performance. This strategy is promising (and something like it has been explored by Millikan in connection with speech acts (2005d)), but there will be a number of questions about how it might work in practice.

The cases I have raised in this section would, I think, present puzzles for most accounts of social practices. The theory of abstract individuals generated through

reproduction sharpens these puzzles, and gives them a particular shape. But sharpness can be an advantage here, as it helps us to know where to look to find an answer. And indeed, in this brief discussion we have seen that there are a number of possible avenues to explore within this theory to give a satisfactory account of these cases.

Conclusion

Over the course of the first three chapters of this thesis, I have been developing and defending an account of the ontology of conventions, along with repeatable artworks and some artefact kinds. I will now pause to summarise the main points of these chapters.

We began Chapter 1 with a number of claims about the view of conventions implicit in much ordinary discourse and social science. I suggested that conventions are generally individuated historically, so that it is possible to have two qualitatively identical but numerically distinct conventions. In addition, particular instances of a convention are such because they bear the right relationships to other instances of the convention. Finally, conventions seem to have some power to explain the occurrence of their instances.

Chapter 1 developed an account of the ontological category of abstract individuals: entities which can be multiply located, but for which the identity of indiscernibles fails. At the most general level, this category secures the possibility of our two points about the individuation of conventions, as well as of repeatable artworks (which have often been described as ‘created types’). Drawing from work on the metaphysics of repeatable artworks, and from Ruth Millikan’s account of natural conventions, we saw that one way to understand the possibility of abstract individuals is through the model of chains of reproductions of their instances, generally tracing back to some exemplar or precedent. Understanding conventions as abstract individuals created through the reproduction of arbitrary patterns of behaviour also sheds light on their explanatory power.

Chapter 2 tackled two broad challenges to the account of Chapter 1, coming from the work of Dan Sperber. I rejected Millikan's characterisation of reproduction in terms of the counterfactual dependence of features of the product on corresponding features of the model. However, I suggested that some such sensitivity is plausibly a *necessary* condition for reproduction in my sense, if it is to have the right kind of explanatory power. Sperber's first challenge, developed in the context of a critique of memetics, was that the normal ways in which novel behaviours spread through human populations do not preserve the right kind of dependence of new behaviours on those they are modelled after. In response, I suggested that the level of dependence one finds even in Sperber's description is sufficient to underwrite my sense of reproduction.

The second challenge from Sperber was more explicitly ontological. He argues for a nominalism about cultural entities which is in tension with my claim that they are abstract individuals. He argued for this position from the heterogeneity of many of the apparent 'tokens' of a cultural type. This argument presupposed, however, that types or abstract individuals collect their instances through qualitative resemblance among those instances. On my account of abstract individuals, they collect their instances instead through chains of reproduction. Finally, I argued against certain other considerations, derived from Richard Wollheim's work in the ontology of art, which might have given us independent reason to think that there could only be little variation among tokens of a created type.

In this chapter I have defended in greater depth the need for a richer understanding of reproduction than Millikan's thin 'official' account in terms of counterfactual dependence. Millikan herself, in practice, seems to acknowledge this, and does provide a richer context, rooted in the biological reproduction of sequences of DNA (and other molecules in organisms without DNA). I argued that this account is more promising than it might appear, but that it nevertheless fails to accommodate the existence of phenomena such as cross-species conventions. Millikan, Sperber, and Dawkins are all pursuing projects which make this sort of consequence look like a reasonable price to pay, in return for a general and simple account of 'reproduction', 'replication', or 'cultural attraction' which could be used as the basis for a new natural science of the social. Perhaps if such projects come to greater fruition, they will

provide us with good reasons to revise the social ontologies implicit in ordinary discourse and existing social science. However, I argued that the mere prospect of such a project is not sufficient to rule out a richer notion of reproduction which is more adequate to existing social ontologies.

The following chapters will extend and apply the account of conventions as abstract individuals which I have developed so far. First I will show in more detail how such conventions can play a role in social explanation. I then turn to their practical or normative role for us.

A final note before moving on. I have focused on conventions in these three chapters, but as I mentioned in the Introduction, and at the end of Chapter 1, I take my account to apply also to rules and norms. So, rules and norms are also abstract individuals, generated through the reproduction of patterns of behaviour. In the next chapter I will talk increasingly of ‘norms’, in particular because I will be concerned with implicit and explicit norms, and ‘implicit norm’ sounds better to me than ‘implicit convention’. But I do not intend to signify any deep distinction between conventions and norms in what occurs below. In the final two chapters I will talk mainly of ‘rules’. In general I reserve this term for norms which agents are characteristically aware of following. This has some important consequences, as we will see, but again the basic claim that rules are abstract individuals created through the reproduction of patterns of behaviour still holds.

Chapter 4. Conventions, Agency, and Social Explanation

Over the previous three chapters I have made a case for the existence of abstract individuals. I have also argued that we can fruitfully understand conventions as being abstract individuals, generated through the reproduction of patterns of behaviour which are importantly arbitrary. I will call conventions, so understood, ‘AI conventions’. ‘AI’ stands for ‘abstract individual’, but it should be remembered that there may be abstract individuals of importantly different kinds.⁵⁴ When I refer to ‘AI conventions’, I mean conventions which are abstract individuals understood through the reproduction of patterns of behaviour, as I described them in Chapters 1-3.

This chapter explores in greater detail how AI conventions can be put to work in projects of social enquiry and explanation. The first question I will address is whether including AI conventions in our ontology is compatible with also including ordinary individual human agents in it. This question gains its relevance in light of a comparison with memetics. Fully embracing the existence of memes in their original sense seems to threaten the existence of human agency as we normally understand it. This is one of the reasons why memetics appears to involve a radical reconceptualisation of the nature of social science. My project has been to provide an ontology which fits better with a more traditional social-scientific approach. Much existing social science, as well as ordinary discourse, acknowledges the existence of *both* ordinary agents, *and* conventions. So for my purposes I need to show that AI conventions are compatible with the existence of ordinary agents.

Next, I develop a more positive picture of how AI conventions fit into projects of social enquiry. I outline a key form of social explanation, which AI conventions are

54 I mentioned some examples which might be understood as abstract individuals through *different* paradigms. These included biological species, certain Searlean social entities, and mere ‘fusions’ of bearers of an impure property.

particularly well-suited to play a role in. In the process I show how my account of AI conventions can also encompass implicit conventions, and conventions governing non-intentional behaviour. This will be accomplished through a case study of a particular such set of conventions: those governing interpersonal distances in social contexts.

These arguments will demonstrate that AI conventions have a role to play in social explanations. But I will end the chapter by asking whether we have reason to think of conventions as ‘social phenomena’ in a more robust sense than this. I will address this question in Chapter 5, which will also lead us into the topic of Chapters 6 and 7: the role of social practices in practical reasoning.

Memes, conventions, and agents

Memetics has often been pitched by its proponents as presenting a radical challenge to existing understandings of the human social-cultural world. This holds in at least two respects. If someone told me that I ought to do something because it will aid the proliferation of a meme which has taken up residence in my mind, I’m not sure that consideration would move me. On the side of explanation and interpretation, it can seem as though viewing human beings as hosts to a population of memes will tend to crowd out many of our more traditional (but far from ineffective) ways of understanding human action. In particular, it is difficult to reconcile with those aspects of folk psychology and folk sociology which view individual humans as relatively unified loci of agency and perspective. I have stressed that AI conventions are not the same as memes. But is my proposal in any better shape than memetics with respect to these two points?

The interpretation of memetics on which it poses a radical challenge to existing approaches to social science is one according to which memes are the *real* agents in social explanation. When we adopt the ‘meme’s-eye view’, individual human agents can appear to dissolve, their ‘actions’ no longer expressions of their agency as enculturated animals, but instead part of the reproductive strategies of some meme or combination of memes which these living bodies play host to. Any residual

activities which cannot be explained by appeal to memes will be fairly uninteresting, and not social in a distinctively human sense.

This picture is theoretically interesting, but difficult to live with. And even those who are keen to push the radical reconceptualisation of the task of social science which is implied by memetics often seek to preserve more of a role for individual human agents than is implied by my sketch in the previous paragraph. Dennett attempts this partly by deflating our sense of what an individual human agent is (e.g. Dennett 2001). But even with a deflated picture of human agency, reconciliation looks difficult. Memes were introduced as entities analogous to genes (Dawkins 1982). It is plausible that there are some quite serious obstacles to properly reconciling the ‘gene’s-eye-view’ with the ‘organism’s-eye-view’. It seems even more likely that, if anything like a majority of our ideas and skills are supposed to be memes, then it will be difficult to reconcile their agency with the agency of the organisms who are supposed to be their hosts, without diluting what is distinctive of memetics.

Are AI conventions in the same boat as memes? They are entities which can supposedly explain patterns of behaviour, but they have an existence which is ‘external’ to particular individual human agents, and as we will see below they can exist without anyone’s having decided to create them. Despite this, I suggest that the existence of AI conventions does not present the same kind of challenge to our ordinary understanding of human agency. The account developed in the previous three chapters held that conventions are abstract individuals, created through the reproduction of patterns of behaviour. The conception of behaviour which was employed throughout those chapters was our ordinary conception of the actions of individual human agents. Examples given included actions like driving on the right-hand side of the road, punting from the deck at the back of a boat, and wearing black to a funeral. These are all the sorts of actions we ordinarily recognise in everyday life, and attribute to individual human agents or sometimes to groups of agents.

Was this just an arbitrary restriction on my part? One might think that the central elements of AI conventions were just those of a pattern, and of the reproduction of

patterns. Might these two concepts not apply equally to things which look nothing like the actions of individual agents?⁵⁵

It is worth emphasising here that not only are AI conventions *created* through the activities of ordinary agents, they are also most directly *manifested* in such actions. In turn, the manifestations of an AI convention are intimately connected to the reproduction of the convention itself. These points played an implicit role in our discussion of the explanatory role of conventions in the previous chapters. They will also be at play in the final chapter, where we will lean on them in making plausible that the evaluation of a convention or rule bears a particularly intimate relationship to the evaluation of the actions which manifest it.

It is worth contemplating whether there are interesting patterns being reproduced in social life which bear a less direct relationship to the particular actions of individuals. (Not only are recessions the work of many hands, but the occurrence of a recession itself cannot be *identified* with any action or collection of actions.) It is also interesting to contemplate the existence of salient properties of events which depend on the actions of many people, where none of the actions, nor their fusion, has those properties. (To slightly mishandle an example of Wittgenstein's – might a festival have a sinister quality, even though nobody involved does anything sinister?) It is difficult, though, to get a handle on questions of explanation with such examples. And practical questions, whilst not irrelevant, are also much more complex (can the sinister quality of the festival influence my participation in it, even if I am doing something that is not in itself sinister?) The *direct* connection between the manifestation of conventions and the actions of individual agents is therefore some reason for treating them as a basic case, given the kind of interest we have in them.

55 The key contrast I have in mind here is between patterns of events which are actions, and patterns of events which are not actions. I refer to the actions of 'individual agents' because it is important that my account of conventions can explain the actions of individual human agents. I don't want to rule out the existence of group or corporate agents, for instance, or the existence of conventions applying to their actions. But a proper discussion of these cases would involve dealing with a number of issues which are tangential to my main concern, here.

I conclude that if we understand conventions, norms, and rules as things which are manifested in the first place by actions, our concern with these entities need not be arbitrary. It is not arbitrary that what we focused on in the first three chapters was the reproduction of *patterns of behaviour*. So the appearance that AI conventions are more compatible with individual human agency than is memetics does not depend on cherry-picking examples that fit my account. There is room for both AI conventions and individual human agents in our picture of the social world.

One task for social enquiry: explaining behavioural regularities

Over the next few sections, I will show in greater detail how we can appeal to conventions, rules, and norms in social explanation in a way that does not crowd out or dissolve individual human agents, as memetics threatened to do. As a preliminary, in this section I will outline one of the possible tasks of social enquiry. It is a fairly traditional, and not especially exciting way of understanding this task. It is not the *only* task of social enquiry. But it is an important task, and one which has a particularly clear connection with rules, norms, and conventions.

One topic of interest in social enquiry will be certain forms of *regularity* in human behaviour. Some regularities seem to be largely explained by general human needs and stable features of the environment, and it might be thought that these fall outside the scope of properly social enquiry. Whiten et al. (1999) describe two populations of chimpanzees who both consume nuts, which they crack using a rock as an anvil. Only in one of the populations is there found an additional technique, that of using a smaller stone as a ‘prop’ to steady the anvil. But the authors do not take this variation as evidence that prop-use is a ‘tradition’ among those who use them, because they find an ‘ecological explanation’ for it, in the fact that the chimpanzee population which does not use props has ready access to suitable anvil-stones which are already embedded in the ground.

I will not try to settle the question whether or not such ‘ecological’ explanation of regularities precludes social-cultural explanation of them, either in humans or non-humans. Either way, it is clear that there are interesting regularities in human life

which do have a more intimate dependence on human activity. An example is provided by the conventions described by Lewis and others, discussed in Chapter 1. The fact that a particular behaviour (say, driving on the right-hand side of the road) has successfully solved a recurring coordination problem in the past allows it to serve as a precedent. Its past success increases its salience as a solution for people who confront the problem in the future, and this helps to explain their choosing to follow this precedent themselves. If all goes well, their choice will reinforce the precedent, which in turn will reinforce the behaviour, and so on.

Not all interesting social explanations for behavioural regularities will fit this model. I emphasised in my discussion of arbitrariness in Chapter 1 that there are many different ways in which precedent can matter in explaining human behaviour. And it is likely that there will be interesting social explanations of regularities which do not appeal to precedent at all. Given this diversity of models, it is worth emphasising that we do in general need *some* such positive story for each case. A behavioural regularity might emerge in some social context merely by coincidence. But it is not the case, as some have apparently thought, that regularity is simply a *default* outcome when human beings live together.

The Victorian anthropologist E. B. Tylor wrote that,

When a custom, an art, or an opinion is fairly started in the world, disturbing influences may long affect it so slightly that it may keep its course from generation to generation, as a stream once settled in its bed will flow on for ages. (63)

There is something compelling about this image. Nevertheless, as a general account of human nature-culture, it is lacking. This can be seen already when we consider another fluid-based folk-sociological image: that of the ‘wave’ of fashion, fad or craze. In the latter metaphor, we are perhaps the water itself, rather than the banks containing it. But the important point is that whilst the wave carries itself forward, its movement seems inherently exhaustible; we do not need to invoke external interference to explain why it eventually crashes and dissipates. Some social-cultural patterns replicate in a way that seems, in principle, open-ended, whereas others do not. This is already enough to show the limitations of the stream picture.

In some respects, a version of this picture continues to be found in certain of the simpler approaches to ‘memetics’ or cultural evolution. That is, to the extent that such approaches treat imitation of observed behaviour as the dominant process in cultural diffusion. We saw in Chapter 2 that Sperber’s critique of such views has a lot going for it. This is not, however, to deny that certain areas of culture might be successfully *modelled* using methods drawn from the modelling of natural selection.⁵⁶ Nor is it to deny that there might be genuine selectional processes in operation shaping culture in particular circumstances.⁵⁷

We should not deny that cultural entities *can* proliferate through imitation, but neither should we assume that this will be their fate unless disturbed by external factors. More generally, the problem with Tylor’s image is that it encourages us to think that the existence of regularity or repetition in human life is a default which needs no special explanation. Whilst there *may* be areas of culture to which we can apply something like the concept of inertia, again this does not seem to be a default condition. So, there is room for distinctively social-cultural (as opposed, say, to ecological) explanations of behavioural regularities. Discovering and describing the reasons for a particular such regularity will be one task for social enquiry.

56 It might be said that selection is really just something that shows up in certain kinds of models, rather than being a specific type of causal process. (For a recent critical discussion of such views, see Otsuka (2016).) Nonetheless, I think we would want to say that the dynamics of a given field were adequately modelled by a selectional process only if the units that figured in the model corresponded to the units which we discern in that field. This is, I think, an underappreciated problem for contemporary approaches to ‘cultural evolution’ in general, and not just selectionist ones: the proper individuation of cultural entities frequently requires knowledge of the *specific* processes through which they proliferate. (This is something I touched on in Chapters 1-3.)

57 A number of references for the contemporary debate concerning the role of selection in the spread of culture were given towards the beginning of Chapter 2.

Explaining regularities by appeal to conventions

Sometimes we can explain a behavioural regularity by pointing to a convention, rule, or norm. In the most straightforward case, the particular instances of the regularity will be manifestations or instantiations of the convention itself. So, for instance, if we wanted to explain why so many people wear black to funerals in the UK, we could point out that it is conventional to do so, and that most of the people who do it are following the convention.

For conventions to *explain* regularities in this way, to say that there is a convention must be saying more than that there is a regularity. What more might we add? One prominent thought is found in the work of H. L. A. Hart. For Hart, what needs to be added to mere regularities is what he called the ‘internal aspect’ distinctive of social rules. The ‘internal point of view’ on a rule is adopted by someone who accepts it as a standard which should govern the behaviour of herself and others (Hart 89). A regularity is not a rule unless at least some people adopt this perspective on it: “[i]f a social rule is to exist some at least must look upon the behaviour in question as a general standard to be followed by the group as a whole” (56).

So, for Hart, the factor that makes the difference between a mere regularity and a rule involves some people adopting an attitude towards the regularity, recognised as such under a relevant description. But this is a demanding requirement. It seems to exclude the possibility of ‘implicit’ rules, where the existence of the regularity, and/or its status as something more than a mere regularity, is not explicitly recognised by the agents who follow the rule. As I have noted before, there may be a linguistic point here: there are reasons for reserving the term ‘rule’ for *explicit* norms. But either way, it is important that we make room also for the existence of *implicit* conventions and norms.

The discovery and careful description of implicit norms is a significant theme in social-scientific work. Implicit norms have tended to receive less attention in philosophical work on the topic. There are at least two important ways in which a norm could deviate from the kind of picture suggested by Hart’s account, each of which might be thought of as rendering norms implicit. (See Table 2.) One occurs

when the relevant pattern of behaviour is recognised by those who engage in it, but its status as a norm is in some important way misunderstood or misrecognised.⁵⁸ So, perhaps, a group of people might recognise that they do things in a certain way, but think that this is because it is in some sense the only option when in fact it is not. Or they might understand it to be a norm, but take it to have a non-conventional source of normativity – perhaps it is viewed as the ‘natural’ thing to do; or perhaps it is taken to be supernaturally commanded. Although certain cases might shade off into the genuinely non-conventional, it seems possible for a norm or convention to be misunderstood by its followers in each of these ways.

Norm Type	Characterisation	Examples
Implicit norm	The pattern of behaviour is not recognised as such by those who follow the norm.	Interpersonal distance norms. Other aspects of bodily comportment.
	The existence of the pattern is recognised, but not regarded as conventional.	A group who think that the meanings of words in their language is fixed by a deity (see Burge 1975). ‘Naturalised’ gendered divisions of labour.
Explicit norm	The norm is not implicit in either of these ways.	Table manners (typically). Cambridge punting convention.

Table 2. Differences between implicit and explicit norms.

The other way a norm can be implicit is when the existence of the pattern of behaviour itself is not recognised by those who follow the norm. Many of the norms governing our bodily comportment in social contexts are of this kind. I will describe one case – that of interpersonal distance norms – in much greater detail in a later section. Again there will be different ways, or different extents to which people can

58 This was the kind of case used by Tyler Burge to criticise Lewis’ account of conventions (Burge 1975; see also Stotts 2017).

fail to recognise a given conventional pattern, although it is less clear how to divide these up in a perspicuous taxonomy. Ruth Millikan's work explicitly aims to account for the existence of norms and conventions which are implicit in this way.

Both dimensions of implicitness fit uneasily within Hart's discussion. Norms which are implicit in the second way seem particularly troublesome. It might, however, be possible to adapt a broadly Hartian notion of the 'internal aspect' to accommodate these conventions. I mentioned Philip Pettit's work briefly in Chapter 1. For Pettit, the question of what makes a norm something more than a behavioural regularity is answered, along the lines of Hart, by appeal to agents' approval of behaviour which conforms to that regularity, and disapproval of behaviour which contravenes it. However, he suggests that these patterns of approval and disapproval might be understood in what, following Lewis, he calls '*sensu diviso*' (1990 729). That is, it might be enough if people adopt the relevant attitudes of approval and disapproval towards each conforming or contravening act on a case by case basis. This pattern of attitudes could be present with respect to the instances of some behavioural regularity, without anyone's putting everything together and realising that they form the pattern they do. So, Pettit's account might handle implicit norms better than Hart's.

Nevertheless, Pettit faces some problems. The clearest case of an implicit norm, within Pettit's framework, would be one where no-one recognises that, say, their approval of S_1 's φ -ing at t_1 , and their disapproval of S_2 's failing to φ at t_2 have a common basis. This may, however, generate a tension with one of the other conditions in Pettit's account of norms. Namely, he requires also that the patterns of approval and disapproval help to *explain* why people conform to the regularity (1990 731). But if no-one recognises that their approval and disapproval form a pattern, we will have lost one obvious route through which these attitudes could explain why people conform to the norm. (That is, no-one will be able to use knowledge of this pattern to try to avoid disapproval by conforming to the regularity.)

That is not to say that there are no options here – perhaps we could invoke a form of conditioning to explain why people stop performing the disapproved actions. However, it points towards a deeper problem with Pettit's account, when interpreted

in sensu diviso so as to apply to implicit norms. When the attitudes of approval exist only *in sensu diviso*, how are we to individuate the norms they generate? What is there in Pettit's conditions to prevent us from saying that there is a norm of stopping-at-red-lights-and-not-castling-with-a-bishop, for instance? In that particular kind of case, where the relevant norms are explicit, we can point to the representations people entertain with respect to these two regularities, which clearly demarcate chess rules from traffic regulations. But what about implicit norms where by hypothesis no such representations exist?

Although it is not clear whether Pettit envisages its role in this way, it might be that we can get some answer from the requirement that the patterns of approval and disapproval explain the behavioural regularities. However, it is not obvious that there will in general be an answer here which reflects our ordinary individuation of norms.

The existence of AI conventions suggests a different approach to implicit norms, which answers both the explanatory question and the individuation question more satisfactorily. (It is not a coincidence that I drew significantly on Millikan's work in the previous chapters, and that she is very interested in implicit norms.)

We have been asking how appeal to a rule, norm, or convention might explain a behavioural regularity. The Hart approach treats rules as behavioural regularities which some people adopt a particular 'internal point of view' on. But the problem was that the most obvious kind of internal point of view is one which involves representation of the regularity itself, and so cannot account for implicit norms.⁵⁹ Subtract that from the picture, and what does one have, over and above the behavioural regularity itself? I do not answer this question, so much as re-frame it. AI conventions are not behavioural regularities, and they are not 'behavioural regularities + ...', either. They are abstract individuals, understood through the notion of the reproduction of patterns of behaviour. The concept of reproduction allows us to explain the occurrence of convention-conforming behaviour, and so to explain the occurrence of conventional regularities, when they exist, by appeal to something distinct from the regularity itself, namely, the AI convention.

59 I will return below to the question of whether *some* kind of 'internal aspect' is a necessary condition even for implicit norms.

The relevant notion of reproduction does not require that the process of reproduction is represented by anyone, nor that it is aimed at by anyone. So AI conventions can be either explicit or implicit norms. The unity of a given convention, and its differentiation from other conventions existing in the same context, is also secured by the concept of reproduction. Roughly, the instances of an AI convention are demarcated by a family of reproductions all tracing back to the same precedent. And if two families do not share a common precedent, then they belong to distinct conventions. Again, this answer to the individuation question does not depend on anyone's *representing* the convention, or the family, or recognising that what is occurring is the reproduction of a pattern of behaviour. (If there are doubts about this, they may be alleviated by the discussion of interpersonal distance norms, below.)

This shows how AI conventions can provide the right principles of individuation. It also shows how they are fitted to provide an answer to the schematic question about explanation: what are conventions, other than regularities, such that they might explain the existence of regularities? In the process, we have brought out how it is possible for there to be implicit AI conventions. In each case, we lean on the concept of reproduction, discussed in the previous chapters. In the case of implicit conventions, we have made use of the fact that reproduction does not require that the reproduction be represented by anyone.

Non-intentional action, and norms governing it

There has been a historical tendency in philosophy, exemplified by Hart and others, to over-intellectualise conventions and norms. We saw above that this tendency overlooks the explanatory role given to implicit norms in social science. Given this role, the fact that the theory of AI conventions can easily accommodate implicit conventions is a benefit of the theory. I think that this would be widely accepted, today. In this section I will argue that we can also accommodate AI conventions which govern *non-intentional action*. The inclusion of such norms will be more controversial, though. In the next few sections, I will introduce a case study which shows that these norms, too, have a place in social enquiry. The details of this case study will also provide a more concrete demonstration of how AI conventions in general can figure in social explanations. But first, I will explain what I mean by non-intentional action, and norms governing it.

A lot of human action, including many of our most interesting actions, is intentional. There is no clear consensus on precisely how to characterise intentional action, but I will outline one fairly common approach, which allows us to get the relevant contrast between intentional and non-intentional behaviour into view. We will see at the end of the next chapter, though, that this characterisation of intentional action may ultimately beg some important questions.

Intentional action is often seen as action which has an *aim*, where this aim is derived in a certain way from a complex structure in the agent's psychology. Although this terminology itself is not always clear, we can understand the relevant psychological configuration in terms of the notions of 'belief' and 'desire'. To take Donald Davidson's formulation, the desire (or 'pro-attitude') will be to perform an action of a given kind *K* (2001a). The belief will be to the effect that such-and-such is a way of performing an action of kind *K*. Put together, this belief and desire could rationalise an agent's doing such-and-such. When an agent does such-and-such *because* she wants to perform an action of kind *K*, and believes that by doing such-and-such she can

perform an action of kind *K*, then she does such-and-such intentionally, and we can think of doing *K* as the aim of her action.⁶⁰

With this understanding of what it is to do something intentionally, I can now explain what I mean by ‘an intentional action’. (See Table 3.) An intentional action will be one which has at least *some* description under which it is performed intentionally, according to the characterisation just given. So, for instance, take a case where I throw a stone at a window to catch my friend’s attention. The stone smashes through the glass, much to my surprise and dismay. So there are at least three things I can be said to have done. I threw a stone, I attracted my friend’s attention, and I broke her window. In the imagined scenario, I did the first two things intentionally, and the third unintentionally. But each of these three report on the same action (Davidson 2001a), and by my characterisation this is an ‘intentional action’, even though it also has *some* descriptions under which it was unintentional.

Action Type	Characterisation	Examples
Intentional action	There is some description under which the action is intentional.	Breaking some eggs in order to make an omelette. Painting a fence because you promised a friend.
Non-intentional action	There is no description under which the action is intentional.	Spontaneously jumping into a pool, fully clothed. Kicking furniture in anger. Startle reflex. Unthinking shifts of posture.

Table 3. The difference between intentional and non-intentional action. (Examples of non-intentional actions are illustrative; actions of some of these kinds may sometimes be intentional.)

60 The only way I could contribute to the literature on how to understand the ‘because’, here, would be to spill more ink over it. See Davidson (2001c) for an important discussion.

Davidson himself seems to have thought that all human actions are what I have characterised as ‘intentional actions’. That is, that there are no human actions which have no descriptions under which they are intentional (Davidson 2001b). I think that this is a mistake. Our capacity to act for the kinds of reasons described by Davidson is tremendously important. But the manifestation of human agency is not restricted to the workings of that capacity. To give a sense of some of the possibilities here, I will list some examples which have been discussed in the literature. I will not be able to make the case for each example in detail here. But in each case we plausibly have something which a) is an instance of human agency, but b) does not fit well with the Davidsonian account of the explanation of intentional actions.

One important class of cases is what Rosalind Hursthouse described as ‘arational’ actions (1991). Her examples are typically actions which are expressive of some emotion, but which cannot easily be interpreted in the light of the belief-desire structure described above. These include spontaneously tousling someone’s hair out of affection, and kicking inanimate objects out of anger.

Certain other kinds of spontaneous action may also escape Davidson’s framework. It may be, as Brian O’Shaughnessy claims, that some spontaneous actions are also intentional. In his example, “you turn your head casually, and there, three or four yards away and bearing down fast, is a bus, and you hurl yourself out of the way!” (244) However, there are other spontaneous actions, such perhaps as suddenly leaping, fully-clothed, into a swimming pool, which do not appear to be intentional in the sense of being geared towards an identifiable goal.⁶¹

Another important class of cases was brought together by O’Shaughnessy under the term ‘sub-intentional actions’. These include many of the movements of one’s

61 I believe that this example is due to Paul Snowdon. I do not mean to suggest that jumping in a pool fully-clothed is something one could not have a reason for doing – it might for instance be part of a swimming proficiency test – but rather that when we view the action as spontaneous we think the agent was not acting on such a reason. There is some reason to think that Davidson would want to construe this as a limiting case of intentional action (2001a 6), but I am not convinced that this is plausible for all cases of spontaneous action.

tongue made whilst one's attention is focused elsewhere, idly tapping one's feet to music, and presumably many of the near-constant shifts of posture we make whilst standing or sitting (cf. O'Shaughnessy Ch.10). The latter may be prompted in some sense by feelings of discomfort which are then alleviated for a time by the shift in posture. However this is, I think, different from those cases where one takes note of discomfort and decides to move in order to alleviate it.

Finally, we can consider 'involuntary' actions. On some conceptions of the involuntary, such as those which emphasise the role of external threats, it is clear that it is possible for an action to be both voluntary and intentional (and, as it were, both 'under the same description'). John Hyman's account of the involuntary is one which makes room for this possibility (Hyman Ch.4). There is, however, a sense of 'involuntary' in which the fact that an action is involuntary seems to be incompatible with its being intentional under any description. Jumping in surprise, crying out in pain, or flinching away from a perceived danger might be examples. This sense of 'involuntary' is, I think, closer to J. L. Austin's. Austin gives as examples of potentially involuntary actions 'hiccoughing' or 'making a small gesture' (17). He suggests that 'involuntary' is not, as sometimes supposed, the 'opposite' of 'voluntary' (though I believe he thinks that they are mutually *exclusive*). The opposites of 'voluntarily' are terms like 'under duress', whereas the opposites of 'involuntarily' are terms like 'deliberately' (17).

One paradigm of involuntary behaviour in this sense may be certain kinds of reflex action. I want to say that such involuntary action can still be a manifestation of human agency, although I recognise that this will be controversial. The inclusion of reflex behaviour may indeed be thought to provide a *reductio* of this claim. However, there is scope for distinguishing amongst reflex reactions between those which are and those which are not manifestations of human agency. I cannot provide a proper argument to this effect here, but there seems to be something to the idea that in startle, for instance, the agent as a whole is (albeit involuntarily) doing something, whereas with the knee-jerk reflex it is just her leg that is doing something. I assume here, then, that involuntary action is a manifestation of human agency proper, which either can be or perhaps necessarily is non-intentional.

Given the existence of non-intentional actions, it is natural to think that there can be AI conventions and norms which govern such actions. In principle, this extension should be straightforward. I have already argued that ‘reproduction’ in the sense relevant to the reproduction of conventions does not require that agents *aim* at reproducing a conventional pattern of behaviour. So the fact that there are no aims of the relevant sort behind non-intentional actions does not prevent non-intentional actions from reproducing conventional patterns in the right kind of way. Nevertheless, norms governing non-intentional action have been largely overlooked and sometimes their existence explicitly denied in the philosophical literature. Some of this may stem from a general rejection of the concept of non-intentional *agency*, such as we saw in Davidson’s work. However, it may also be due to a suspicion that there is just nothing of interest happening at this level of action. The next few sections aim to undermine this suspicion through a case study. A separate suspicion will be raised and addressed in Chapter 5.

Norms of interpersonal distance

There are norms governing how humans take up space in relation to each other. Yet these norms seem, in many cultural contexts, to be largely implicit. They are often spread and manifested without their existence being explicitly recognised by many of those who follow them. It seems likely that many do not even recognise the existence of the relevant patterns of behaviour, let alone that these patterns are conventional. The most easily observable of these norms space concerns the distances at which people stand when engaged in general social interaction. There appear also to be norms concerning, for instance, bodily orientation and other aspects of posture, and these interact in interesting ways, but here I focus on interpersonal distances. In this section I will outline the literature on interpersonal distance regulation, showing that there is good reason to think that this activity is shaped by local AI norms in many social and cultural contexts. In the next section I show how interpersonal distance norms can reproduce implicitly, and through non-intentional action.

In part because many cultures lack explicit articulations of these norms, it is difficult to describe them clearly and precisely. I will describe some of the complexities here, although this summary will not be as systematic as I would like. Despite being a topic of study since the middle of the 20th Century, there is no clear theoretical framework or frameworks through which to approach the topic as a whole (as opposed to particular facets or applications). This is due in part to the fact – itself of some significance – that human space-occupying behaviour is a topic of concern for so many different disciplines. Early discussions drew from ethological work on physical contact and distance regulation in non-human animals (e.g. Hediger 1950). Interpersonal distance has since been taken up by anthropologists, sociologists, psychiatrists, researchers in a variety of branches of psychology, and robotics engineers.

There is evidence that, in those cultures which have been studied, the distances people maintain and/or prefer tend to vary according to the age, gender, status, and personal relationship of the parties, as well as being influenced by various physical features of the parties and of their environment. There is a helpful summary of much of the empirical literature in Sommer (2002). Despite these apparent cross-cultural similarities,⁶² and interesting continuities between human and non-human animals, distance regulation in humans appears to be significantly shaped by local cultural context.

E. T. Hall was among the first to emphasise the cultural diversity of distance regulation in humans. For Hall, interpersonal distance was structured. At least for North Americans, he posited four distinct ‘zones’, possessing different features and reserved for different kinds of activity and relationship (he labelled these ‘intimate’, ‘personal’, ‘social’, and ‘public’). He charted a number of relationships between

62 The similarities in the structure of interpersonal distancing across cultures should not be over-stated. Sorokowska et al. find that gender has a greater effect in some countries than in others. They also find that those who prefer a relatively greater ‘social’ distance do not necessarily prefer a greater ‘personal’ and/or ‘intimate’ distance. A striking example is Romania, where subjects displayed a preference for the greatest distance from strangers of all countries studied, but one of the smallest distances from those with whom subjects were more intimate (Sorokowska et al., 585).

interpersonal distances and the sensory qualities of interaction at those distances. Perhaps most obvious, the further one stands from another, the louder one must normally speak to be heard by them. More surprising was the existence of interesting clusters of distances at which, for instance, fine details of the face, bodily odours, and radiant heat become perceptible. These perceptual thresholds might be thought to depend on fairly universal human endowments, which may therefore tend to encourage a certain kind of structure for the regulation of interpersonal space. (Though we should not forget that the cultural significance attached to, e.g. the possibility of smelling the bodily odours of those with whom one is interacting will vary in complex ways. In addition, the perceptual thresholds will be affected by culturally-variable factors such as clothing and the use of perfumes.)

Despite these perceptual clusterings, Hall saw the regulation of interpersonal distance in humans as an intensely cultural phenomenon. He emphasised that the general structure, the average distances, and also various more specific rules for taking up space vary between cultures (1966 128-129; Ch.XI-XII). In *explaining* cross-cultural differences, he put some store by a broad distinction between ‘contact’ and ‘non-contact’ cultures. Although some of the anecdotal observations he used this distinction to explain appear to have been supported by subsequent studies (Sorokowska et al. 579-580), his division seems too simple, and its explanatory purport as a construct is unclear. However, the most easily testable hypothesis emerging from Hall’s work – that actual and preferred interpersonal distances between a given category of interactants varies across cultures – has been well-supported in the subsequent literature.

It is important to note that intercultural variation is not, in itself, sufficient evidence for the effect of norms. The mere fact of intercultural variation leaves open the question of whether this variation is to be explained by the existence of varying local norms. In addition, where variation is explained by norms, it may not be straightforward to read the content of the norms off from the observed behavioural data.⁶³ In connection with the first point, the authors of a large recent study which

63 I will briefly illustrate one such difficulty, which I have not seen properly acknowledged in the literature. We have noted that interpersonal distance often varies according to

supported the finding of intercultural difference nevertheless express some scepticism about the role of norms in explaining these differences, suggesting that, “[i]t is likely that what has been explained in terms of vaguely defined cultural norms is underpinned by some psychological and ecological variables” (Sorokowska et al. 580).

These researchers tested for various ecological variables and found a number which appear to have significant effects. One such variable is temperature: “The higher the annual temperature of a country, the closer was the preferred distance to strangers” (Sorokowska et al. 585). It does seem quite plausible that there are some environmental variables which directly affect interpersonal distance in a way that is not significantly mediated by social norms. For instance, universal features of human psychology might be such that the temperature at the time of interaction affects people’s preferences for proximity, which are then realised in the proximity they actually adopt in interaction. (Compare: it does not seem as though we need a special explanation to understand why people would stand closer together when they are conversing in a loud environment.) Cultural groupings map, to a greater or lesser extent, onto environments which reliably possess certain features which differ from those of the environments associated with other cultures. In this way, following Sorokowska et al.’s suggestion, a combination of psychological and ecological factors might account for some intercultural variation in interpersonal distances. If *all*

factors like participants’ gender, age, and ‘status’. When we see a given distance frequently maintained between e.g. bosses and employees, or men and women, it can be tempting to read this as being the normative distance for interactants with those properties. However, the picture might be more complicated. In some contexts, it seems plausible that one of these parties will typically have more social power to determine the interpersonal distance being maintained. If so, what appeared to be *the* ‘normative distance’ between e.g. bosses and employees might turn out to be more reflective of the bosses’ sense of the appropriate distance, which is regularly manifested in interaction in spite of their employees’ desire to maintain a different distance. (One study which found that men approaching women tended to stand closer than women approaching men is (Kaya and Erkip 1999).) Acknowledging this seems to be perfectly compatible with saying that interpersonal distance is being governed by norms, but it suggests that we should be careful about how we describe those norms.

intercultural differences could be explained in this way, then there would be little room for local norms to do any explanatory work here. I take it that these are the grounds for their scepticism.

However, it is not the case that those researchers who invoke social norms in their explanations of interpersonal distance simply overlook this possibility. So, for instance, Aiello and Aiello studied the development of interpersonal distancing in children interacting with other children of the same school grade and gender, from the ages of 6 to 16. They found significant differences in the interpersonal distances maintained depending on age, as well as significant differences between the distances maintained by boys and girls at various ages. The ecology plus psychology approach suggests certain hypotheses for explaining these results. It might suggest that the first results were due to the differences in height, limb length, etc. between younger and older children, or possibly to some process of cognitive development which occurs during that time. The second set of results might be explained by some general differences (i.e. not specific to interpersonal distance regulation) between boys and girls.

However, the authors of the study took these possibilities into account, and do not think that they apply here.⁶⁴ Their own interpretation is that the gender differences are explained in terms of the fact that the local norms of interpersonal distance for boys and for girls are different (Aiello and Aiello 186). Gender is making a difference, but it is making a difference because it is a factor which figures into the *content* of the local norms.

The fact that the authors make this distinction speaks in favour of the idea that when they talk about ‘norms’, here, they are invoking something belonging to the category I have described as AI conventions. We have seen that they are appealing to

64 They control for changes in height, limb length, etc. across the ages. They also note that behavioural measures (observed distances in actual interaction) and ‘projective’ measures (e.g. placement of dolls in a modelled scenario) appear to develop in sync. With respect to the effect of gender, they note there is little difference for young children, but boys and girls then follow separate but *parallel* developmental trajectories which lead them to the normative distances observed in adults (Aiello and Aiello 186).

something other than Sorokowska et al.'s ecological and psychological variables. And so, in particular, they are not using the term 'norm' to refer to some hypothesised combination of ecological and psychological variables, as yet unknown. Norms are being appealed to as explanatory entities in their own right, rather than as a way of gesturing at an unknown cause.

'Norm' is not just a name for a behavioural regularity here, either.⁶⁵ We can see this from the authors' explanation of the developmental trajectory for children of both of the genders studied. Their suggestion is that young children do not manifest the distances regularly manifested by adults in their culture, because the norms which account for the adults' behaviour are not applied to the children until they are older (Aiello and Aiello 187). The scope of the norm is being invoked to *explain* the regularity in adult behaviour, and the age at which children begin to manifest that regularity, too. (The same line of thought suggests that they are not using 'norm' to refer to one of Sperber's fuzzy sets of particular actions and representations, either.)

So, when these authors talk about 'norms', they are not talking about some prevalent combination of ecological and psychological factors, and they are not talking about mere behavioural regularities. They are invoking something that exists in a particular social context, is manifested in the behaviour of many distinct agents over a significant period of time, and helps to explain that behaviour. In short, they are

65 I have said that regularities in social behaviour are not, just as such, explanatory. However, the existence of a regularity might explain a piece of behaviour without the involvement of a norm. If I want to avoid tripping over other people whilst walking in a crowd, I do well to match my step to theirs. There can be other reasons for trying to 'fit in' with a regular pattern found in one's social context, which do not depend on that pattern being normative. However, the developmental story invoked in this paragraph is not the story of how children come to have reasons for doing what others around them happen to be doing, but rather the story of how children come to have the local norms applied to them. I should also mention that it is consistent with the story developed in the text that there might also be child-specific norms which partly explain how children regulate their distances. However, the main point is that there can in principle be an asymmetry between the regularities observed in the children's behaviour, and in the adults' behaviour, such that only the latter is explained by a norm.

talking about something like AI conventions as I described them in the previous three chapters. I suggest, then, that we should understand the norms that these and other authors invoke as belonging to the category of AI conventions.

I will sum up the argument of the last few paragraphs. Many of the researchers working on this topic appeal to local norms as a partial explanation of the regularities found in interpersonal distancing in a given cultural context. Far from being a term used as a place-holder for an unknown combination of ecological and psychological factors, we have seen that the norms they talk about fit well with our understanding of AI conventions. In the next section I will show in greater detail *how* such norms get reproduced, and in particular how they can reproduce implicitly, and through non-intentional action. In the rest of this section I will elaborate on how accounts of interpersonal distance which invoke norms can respond to Sorokowska et al.'s scepticism.

It would be foolish to defend every claim made for norms in the literature on interpersonal distance. For one thing, they form an inconsistent set. But, more importantly, I do not need to for my purpose during the last few paragraphs, which was to show that those theories of interpersonal distance which invoke norms are often talking about AI conventions. Nevertheless, the Sorokowska et al. thought might not seem to have been adequately addressed yet, and so it may seem to undermine any such theory.

It is important to note, first, that the ecological factors for which Sorokowska et al. found a significant effect nevertheless did not suffice to account for all of the intercultural variation that they observed. The second point is that there need not be *competition* between norms and ecological variables as explanatory factors, here. As we saw above, one of their findings was that, “[t]he higher the annual temperature of a country, the closer was the preferred distance to strangers” (Sorokowska et al. 585). But this correlation does not tell us anything about the mechanism through which temperature might have this effect. It is implausible that individual agents are *directly* responding to ‘annual temperature of a country’ when determining their preferred distances. It is possible, as sketched above, that the effect of annual temperature can be explained by the aggregate effects of the temperature in particular interactions, as

in the ‘ecology plus psychology’ model. However, it is also consistent with the current findings that what different local average temperatures help to explain is differences between local *norms*, and that local norms explain preferred and observed interpersonal distances. The fact that clashes in interpersonal distancing between people from different cultures can persist in spite of extended residence in the same environment would also seem to speak against the ecology plus psychology suggestion (Hall 1966 Ch.XI).

To sum up this point, the existing data do not establish the view that intercultural differences in interpersonal distance regulation can be explained without appeal to norms, just in terms of ecological and psychological factors.

How interpersonal distance norms get reproduced

In the previous section I argued that the norms invoked in explanations of interpersonal distance regulation belong in the category of AI conventions. But AI conventions are created and spread through the reproduction of patterns of behaviour. So far I have not said anything about how interpersonal distance norms might be reproduced. In this section I will provide some details. This discussion helps to fill out the suggestion that these norms are indeed AI conventions. It also demonstrates how norms can exist and proliferate implicitly, and through non-intentional action.

The fact that observation of clashes in interpersonal distancing between people from different cultures has played a crucial role in bringing researchers’ attention to interpersonal distances itself speaks to the implicit nature of many of the norms regulating such distances. It is no part of my claim that this need be universally true, however. So, for instance, there may be cultures where such norms are much more explicitly thematised than in contemporary North America and Europe. In addition, there are particular contexts where interpersonal distances are explicitly addressed. At a reception desk in a doctors’ surgery in England, for instance, one might find a sign saying, “Whilst the person in front of you is being served, please stand behind this

line.”⁶⁶ However, I assume here that interpersonal distance norms are often largely implicit.

If these norms are generally not explicitly recognised as such, then they will not be proliferated through explicit instruction. How do they spread? There is surprisingly little empirical work directly addressing the mechanisms through which distance norms proliferate. However, Ruth Millikan’s suggestion here seems very plausible, and it appears to be supported by what empirical work I have been able to find on the topic. Millikan suggests that, “[i]f you are standing at the wrong social distance, the person to whom you are talking will move; so to prevent slow circling around the room as you talk, you unconsciously reproduce the conventional social distance [...]” (2005a 5). That is to say that the convention is learned by being instantiated. Those who have not yet developed a ‘feel’ for the norm will tend to remain still at some point, to avoid circling around the room, and through experience of conversations carried out at this distance, they can be expected to learn to maintain it themselves.⁶⁷

A possible objection to this story of how interpersonal distance norms spread might highlight how long it typically takes children to learn to adopt the locally normative distances. Very often young children will stand much closer both to other children and to adults, than adults stand to each other. Over time they tend to stand further apart, eventually converging on the distances maintained by adults, but as we have seen this is not simply explained by an increase in their height or other bodily dimensions (Aiello and Aiello 187). It might seem odd that if interpersonal distances are learned in the relatively simple way described by Millikan, it would take many years for a typical child to learn them.

66 The ‘proper’ interpersonal distances are also an explicit theme of much homophobic bullying, largely among adolescent boys. It is not always clear whether the distance norms invoked here are those present in the wider social context, or rather a semi-independent set of norms created by the commentary itself.

67 It is difficult to avoid purposive phrasing here. However, if we’re willing to help ourselves to Millikan’s notion of ‘unexpressed purposes’ then purposive behaviour need not be intentional behaviour (e.g. Millikan 1990).

However, we have already encountered the suggestion that at least part of the explanation for this developmental trajectory lies in the fact that adult distance norms are not applied to young children. Adults do not tend to move away from young children so as to maintain conventional distance, and so children do not become involved in the dance described by Millikan until later on. One study asked participating children of various ages to stand very close to adult strangers in a queue for the cinema. Adults frequently engaged in a friendly way with the 5-year-old children, and 8-year-old children were largely ignored. When 10-year-olds approached, however, many adults stepped back, leaned away, or showed signs of discomfort (Fry and Willis 386-387). Interestingly, the authors do not record any cases where children were explicitly told to stand further away. The cinema queue cannot stand in for all child-adult interactions, but this study implies that adults do not treat children as subject to distance norms, or regulate their interpersonal distances accordingly, until the children reach a relatively late age. If so, the fact that children do not learn these norms properly for many years does not speak against Millikan's account of how these norms are learned.

In the scenario described above, distance norms are being imparted without explicit instruction. Moreover, Millikan suggests that this behaviour might occur without anyone's entertaining thoughts of the norms at all. (This is also supported by anecdotal evidence that people are frequently unaware of the existence of these norms.) If this is correct, then the norms will be implicit in the two ways I described above.

Interpersonal distance norms also govern and are frequently manifested in non-intentional action. Not only do we not think explicitly about the norm itself, we also do not tend to think of many of the actions by which we regulate interpersonal distances as *means* for doing so, or for doing anything else. Indeed, from experience and observation it seems as though we are frequently unaware of making many of the movements – the subtle shifts in posture and position, etc. – which regulate these distances. I suggest, then, that much of the activity which creates and sustains these norms is non-intentional. True, we might be aware of a sense of discomfort if someone stands 'too close', and of its reduction when we move away, but as I

mentioned above, something very similar is true in the case of much fidgeting and shifting of position, which again we may be hardly aware of doing, and which often does not count as intentional action.

We can envisage other processes by which norms governing interpersonal distances might be manifested, spread, and reinforced implicitly. (Although it is more likely here than above that the processes will involve *some* intentional actions.)

A number of possible routes proceed through material culture and the configuration of the environment. Sometimes our activity leaves traces on the environment which will tend to reinforce occurrence of the same activity in the future. A paradigm here is the path. (Here, not coincidentally, Tylor's stream metaphor appears quite apt, although perhaps in ways he did not intend.) In a similar way the natural environment might register where we sat to talk, with patterns of wear on fallen logs or grassy slopes used as seats. Looking for somewhere to sit, others may then fall into these grooves, and their spatial configuration relative to each other will reproduce that adopted by those who sat there before them.

More actively, and bringing ourselves into the human-built environment, we might adjust chairs (perhaps unthinkingly, as in Millikan's story) until we are comfortable in our distance from each other. Unless they are cleared away, the next people to use the space will find a default configuration shaped by the previous users, remaining after their conversation has finished. Where the material environment is less yielding (as many benches are fixed to the spot, or too heavy to move), a form of 'selection' may favour configurations which fit with local norms. Let's imagine that, not having read Hall, our designers place such furniture with no particular regard to the interpersonal distancing which it facilitates. We might expect that, all else being equal, those configurations which fit most closely with local distance norms will prove most popular, because most 'comfortable' (though no-one puts their finger on why). If the popular designs are then copied, the resulting environments will reproduce those same norms, but once again no-one need have recognised this.

As before, similar processes might also operate more explicitly. An architect may consult her sense of what distances are 'fitting' to maintain between conversational

partners, and design a space accordingly. Or, having read Hall, she might use the measurements he describes for local interaction distances, either to determine the ‘proper’ way to do things here, or else out of a more utilitarian concern with the comfort of her clients.

In this section and the previous one I have described the research on interpersonal distance regulation in humans. This research suggests that robust (albeit complex) patterns of interpersonal distance regulation are explained in part by local norms. We saw in the last section that these norms fit the profile of AI conventions. In this section we have seen in greater detail how AI conventions governing interpersonal distance can be reproduced implicitly, and through non-intentional action.

Connections to broader projects

It might be thought that non-intentional action forms a largely encapsulated arena in human behaviour. These are actions which we often fail to notice either when they are performed by others or when they are performed by ourselves. The fact that there are norms which govern such behaviour might be of some idle interest, but do these norms, or this behaviour, *matter*? Or are they just something that occurs in the background? In the next chapter I will defend the claim that non-intentional action can have what I call ‘practical significance’ for us. In this section I will briefly show some of the ways in which non-intentional action, and the norms governing it, can be of broader *theoretical* significance than we have seen so far.

As I mentioned right at the beginning of this thesis, interpersonal distance norms may also help to explain certain patterns at a more ‘macro’ level, which are not *directly* governed by norms. Consider behaviour in fixed, restricted spaces where many people find themselves temporarily thrown together, such as elevators. (I confine myself here to observations made in the USA and Australia.) In such spaces, we find regularities in the overall pattern of distribution of bodies (e.g. Hall 1955 86). We also see regularities in the particular positions people take up, for instance more senior men tend to stand against the back wall (Rousi). Finally, people often act in unusual ways when they find themselves temporarily ‘too close’ to sustain a normal

interaction (Goffman 137-138). The appeal to interpersonal distance norms helps us to make sense of each of these phenomena.

Interpersonal distance regulation also relates to other forms of behaviour, some of which are much more obviously intentional. One fairly dramatic example can be found in a study of assaults on a dementia ward. Researchers noticed that assaults were more frequent on a dementia ward at meal times, meals being taken on a different floor (Negley and Manley 1990). They conjectured that this might be explained in part by the fact that a number of patients struggle to adjust their behaviour ‘normally’ to the crowded conditions in the elevator to the dining room. Standing ‘too close’, or orienting face-to-face at close proximity, were treated as provocations (Negley and Manley 31). For the most part, elevator behaviour is not directly reproduced from one occasion to another; it is doubtful that there are many norms *directly* governing such behaviour. But understanding the norms which *do* exist concerning interpersonal distance in general allowed the researchers to understand the conditions which led to assaults, and to improve the problem by having meals be taken on the same floor as the ward.

As a final case, we can look at Perry et al.’s study of people with damage to their orbitofrontal cortex (OFC). The authors report that such people often do not “conform to social norms”, and in particular they frequently stand much closer to others than is standard in their cultural context (Perry et al. 1898-1899). They suggest that distance-regulation might normally be underpinned by an ‘automatic feeling’ that one is, say, too close to another person, but that people with OFC damage do not have this feeling.⁶⁸ Interestingly, they note that the patient whose behaviour most closely resembled that of control subjects was making use of an explicitly formulated norm, remarking, “I *know* that this is the distance I *should* choose to stand from a stranger” (Perry et al. 1899).

68 To say that the feeling is ‘automatic’ is not, as I understand it, to say that it is not learned, so this view of the mechanics of distance-regulation need not undermine my suggestion that distance regulation depends in part on local norms. Perry et al. themselves make free appeal to social norms in their discussion.

This person had to explicitly learn the norm, and use it to consciously guide her behaviour. Doing so allowed her to maintain more ‘normal’ interpersonal distances than the other people with OFC damage. This helps to highlight the work that the norm is generally doing in the background. But it is also striking how different that subject’s experience is from mine, at least. Reading this description I am struck by how much effort she has to put in to doing something that I am generally hardly aware of doing, unless something specific about the context brings it to mind. This case therefore highlights that even if the non-intentional behaviour regulating interpersonal distance usually resides in the ‘background’ in some sense, this may be due more to our taking it for granted, than to its being inconsequential.

We can widen our view out from the case of interpersonal distance, to look at how other kinds of non-intentional action figure in social enquiry. One place to look to find study of non-intentional action in contemporary social science is in work influenced by what is sometimes called ‘practice theory’.⁶⁹ For example, Pierre Bourdieu describes the ways in which gender roles amongst the Kabyle were reinforced not only through explicit formulas, but also through the physical layout of the house, and through habitual bodily orientations and postures (1977 90-95). A common thread running through much other work in practice theory is the thought that close attention needs to be paid to the material environment and to ‘habits’ (Schatzki 17), or “routine activities (rather than consciously chosen actions) notable for their unconscious, automatic, un-thought character” (Swidler 83).

These examples give a sense of how non-intentional behaviour can have interesting links with other kinds of behaviour. It is not something that simply occurs in the background of human lives, like, perhaps, digestion.

69 The usual caveats apply about labels for trends, paradigms, or research programmes in academic thought. However, this label does seem to serve a useful function for my purposes, so I will use it.

Conclusion

This chapter has applied and built upon the account of AI conventions developed in the first three chapters, focusing on their role in projects of social explanation. This has involved three tasks. One was to provide a more positive account of the explanatory role of AI conventions. A second was to explore their scope, and in particular to argue that there can be implicit AI conventions, and conventions governing non-intentional action. A third was to defend these points from certain objections.

The first task had two main parts. First was to answer the schematic question of how AI conventions could figure in explaining significant social regularities in behaviour. The answer here was that AI conventions are distinct from regularities, and hence can explain their occurrence. The second was to show that the entities appealed to by social scientists in their explanations of behaviour do in fact fit the profile of AI conventions. This was achieved through criticism of the norm-scepticism of Sorokowska et al.

The second task had a similar structure. First we saw that understanding AI conventions through the concept of reproduction allows us to see how they could be a) created, and b) individuated, without appealing to agents' representation of the convention, or their aims in acting. So, unlike the work of Hart and Pettit, the theory of AI conventions allows us to see in a schematic sense how there could be implicit norms, and norms governing non-intentional action. The second part involved showing that there are indeed such norms. We saw this through the case study of interpersonal distance norms. This revealed that such norms tend to be implicit. It also revealed in some detail *how* they can reproduce implicitly, and through non-intentional actions.

Two objections were also addressed. The first was an objection to the general attempt to use AI conventions in social explanation. My aim in Chapters 1-3 was to develop an ontology of conventions which helps to make sense of much of our existing discourse in social science and everyday life. That discourse accommodates the existence of both conventions and ordinary human agents. The worry I

addressed was that the theory of AI conventions would be as revisionary of our ordinary understanding of human agency as memetics threatens to be.

The second objection was more specifically to the claim that AI conventions can explain *non-intentional* action. It might be thought that non-intentional action just quite generally lacks any social significance. However, in the final section I described a number of ways in which the investigation of interpersonal distance norms in fact feeds into other areas of social enquiry. For each of these broader projects, it does not matter if distance regulation is often achieved largely through non-intentional action.

However, this last point suggests an ingredient which seems to be missing from our picture of how AI conventions fit into the social world. We think of conventions, rules, and norms as ‘social phenomena’ in a way that goes beyond anything that is captured just by pointing out that they figure in social explanations. We can see this point by contrasting them with, for instance, purely geographical entities. It has been suggested, for example, that early cities or proto-cities which grow up on the edge of marshes or other wetlands tend to be more resilient than those which do not (Pournelle and Algaze 2010). This could be considered as a social-explanatory hypothesis. But if it turns out to be true, it does not suffice to show that marshes or river deltas are themselves social phenomena in any interesting sense. Might AI conventions be in the same boat?

I will address this worry at the end of the next chapter. My answer will draw upon the claim that AI conventions can have *practical significance* for us. That claim is also a presupposition of the arguments of Chapters 6 and 7, where I will apply the ontology of AI conventions to shed light on the role of social practices in practical reasoning. The claim that AI conventions have practical significance therefore provides a bridge between the explanatory questions we have just been considering, and the practical questions of the final chapters. Establishing that key claim is the task of the next chapter.

Chapter 5. The Practical Role of Conventions

The preceding chapter showed that AI conventions can be implicit. It also showed that they can be generated through, and in turn govern, non-intentional actions. From the point of view of social explanation, this is a positive result. We have seen that implicit norms and norms governing non-intentional behaviour are an important part of the social-scientific literature on interpersonal distance and a range of other topics.

But conventions, rules, and norms are not invoked only in social explanation. They are also supposed, at least given the right circumstances, to be *norms* in the sense that they provide reasons for action, and standards by which action can be judged. This is an assumption of the final two chapters, where I will show how the ontology of AI conventions can shed light on their role in practical reasoning. However, the arguments of Chapter 4 might appear to threaten the use of AI conventions for this purpose.

It is possible to generate a quite intuitive sense of unease about the alleged practical significance of those AI conventions which are *both* implicit *and* govern non-intentional action. Because what we are saying when we say that a norm is implicit and governs non-intentional action is that it is not something we designed, or chose to install. Further, although it is in some sense the upshot of our actions, these actions themselves are also things we did not choose in the sense of their being intentional. Finally, it is a thing which, once installed, purports to govern further actions which we still do not choose in this way. Can we make any sense at all of the idea that entities which apparently arise in this way could ever make a serious claim on us?

There are two related worries here. One is that it quite generally doesn't make sense to apply notions like agential responsibility or socially-determined standards of

correctness to non-intentional behaviour. The second is that even if we can apply these concepts to non-intentional action, nevertheless *implicit AI conventions* cannot generate standards of correctness for such action. If either of these thoughts are correct, then this would seem to undermine the claim that AI conventions as I have described them are fit to provide a bridge between social explanation and practical philosophy.

I will address these two concerns in this chapter. Along the way I will explore the relationship between implicit and explicit norms, and the way in which reflection on a previously implicit norm can transform our practical relationship with it. By the end of the chapter, we will be in a position to provide a more positive characterisation of the sense in which AI conventions are ‘social phenomena’, in a way that contrasts with deltas and marshes.

The practical significance of non-intentional action

In this section I will address the concern that it doesn’t make sense to talk of norms in any genuinely ‘normative’ sense, in connection with non-intentional action. This concern is motivated by the thought that non-intentional action, quite generally, lacks what we might call ‘practical significance’. It is uncontroversial that we can apply notions like agential responsibility or socially-determined standards of ‘correctness’ to intentional actions. But it would widely be thought that such notions are out of place regarding certain other bodily events and processes, such as digestion. Philosophers often write about non-intentional action as though it falls on the side of digestion, in this respect. In this section I will describe some cases which suggest that, to the contrary, we intelligibly treat non-intentional action as having practical significance in many social contexts.

Clifford Geertz, following Gilbert Ryle, reminds us that from the point of view of ethnographic interpretation there is a world of difference between a wink and an involuntary twitch of the eyelid (C. Geertz 1973). As we will see below, from the point of view of interpretation the gap is sometimes thought to be such that interpretation is in fact entirely out of the question in the case of the mere twitch.

We can ask a similar question about the difference between the two from a practical perspective. One might blame or resent the winker; or her winking might be a signal to you within some joint project that you are pursuing together. But a mere twitch of the eyelid?⁷⁰

What do we mean by ‘practical significance’, here? By definition, we do not deliberate about whether or not to perform a non-intentional action. But the practical perspective is broader than the deliberative perspective, and there are a number of other ways in which non-intentional action might have a kind of practical significance which events in the inanimate world, and *merely* bodily processes like digestion, lack.

Some of the most well-trodden ground in analytic philosophy, here, concerns what we might term ‘reaction’. Practical questions here concern the attribution of responsibility, the allocation of praise, blame, guilt and pride, the seeking of revenge, and so on. But there are other facets of the practical perspective beyond reaction. So, for instance, there is acting *jointly* with others. There are also ways of trying to *oppose* another’s action, which differ from the ways one might oppose other kinds of development in the physical world (cf. Haase 2014).

For each of these phenomena, it is sometimes thought that the real difference is between intentional actions and all other kinds of events or processes. So, for instance, it is sometimes thought that it is only legitimate to praise people for the results of their intentional actions.⁷¹ That it makes no more sense to blame them for their non-intentional behaviour than it does to blame the sea for eroding the cliff. Or that opposing someone’s non-intentional behaviour is just like opposing the falling of a tree towards the ground.

70 An involuntary twitch of the eyelid may not be an action at all. Sometimes it really will fall on the side of digestion. But in some cases it will be a non-intentional action, as when it forms part of a startle reaction. I am considering cases where it is an action.

71 Some have thought that praise or blame are only appropriate for the *intended* outcomes of their actions. This is a stronger and to my mind less plausible view, though I wouldn’t deny that whether a given outcome is intended, or just a result of an intentional action, can make a difference.

There is no need to deny that whether or not an action is intentional, and under what description, can make an important difference from the point of view of reaction, as well as from that of joint action (e.g. Gilbert 1990; Velleman 1997), and the opposition of action (Haase 133-141). There are many complex questions concerning the precise role of intentions, here. I won't try to provide a full theory of, for instance, reaction. My aim here is just to show that there are some interesting cases where people appear to take up the non-intentional behaviour of themselves and others into the practical perspective in a way which suggests that it has the kind of significance required for my account of norms governing non-intentional behaviour. That it is taken up in this way by some people does not, in itself, prove that it has practical significance. It is possible that they are making a mistake. However, I hope that the examples will begin to show that these practices are at least intelligible, and that if there is a mistake here it is not an obvious one.

Such cases can be troubling. It can feel unfair to attach much significance to people's involuntary behaviour, or to police it in the ways that we often do. I think that this is true, but it is equally true that we are often unreasonably concerned with controlling and judging the *intentional* actions of other people. The point that we are often over-zealous in this respect should, I think, be held separate from the claim that it is, say, not even coherent to apply the notion of personal responsibility to individual intentional action. Likewise, whilst we might think that many of the ways in which we hold people responsible for their involuntary behaviour or try to shape it are mistaken, unfair, or damaging, this need not depend on the idea that it is in principle incoherent to apply the notion of agential responsibility to such behaviour.

The first case is drawn from a story which circulated about Queen Victoria during her reign. As a young woman, she was visiting the backstage of a circus when a caged tiger sprang towards her. Her retinue, it is said, jumped back in fright, but "the youthful Queen never moved face nor foot. With look undiverted, and still more deeply-riveted, she continued to gaze on the novel and moving spectacle" (Hobhouse, qtd. Watt-Smith 3). Although it seems likely to be fictional, we can assume that this detail was not invented at random. We feel that it was included because it seemed to the teller and their audience that it said something of

significance about the Queen. According to Tiffany Watt-Smith's interpretation, this story was thought to say something about Victoria's 'character' (16):

In her unflinching demeanour—which was almost certainly more rhetorical than real—the monarch was attributed with an emotional self-mastery usually only associated with men in the Victorian period. Jumping startled into the air, retreating and turning aside, or other forms of recoiling, shrinking, wincing, and flinching were habitually linked to intrinsic cowardice, feminized sensitivity, or attention-seeking theatrical performers. (3)

Yet in the story, Victoria does not *do* anything, and her inactivity gains its significance through contrast with a set of actions (jumping back in fear) which we can presume to have been *involuntary* on the part of her retinue.

Common views about the particular significance of flinching have changed somewhat since Victorian times. (Though perhaps not as much as we might like – for instance, visibly reacting to startling events is still frequently coded as 'feminine' behaviour for both men and women, and the supposed incongruity of men doing this is played for laughs in TV and film.) However, even if we no longer share the view that flinching has such deep moral significance, the basic form of thought described by Watt-Smith is nevertheless intelligible to us. It seems that we still, under certain circumstances, treat involuntary behaviour as manifesting something interesting about the person whose behaviour it is. Perhaps – picking up Watt-Smith's term – about the person's 'character'. As just mentioned, such involuntary behaviour as flinching continues to be employed in significant ways in fiction, as well as in a number of (generally informal) games, where it can be tied up with notions of personal courage or interpersonal trust.

The point I wish to take from this example is that we have access to fairly rich vocabulary for talking about human agency which we sometimes overlook in philosophy. Watt-Smith's term 'character' demonstrates this fact, though I do not want to put too much weight on that particular concept. (We are often wary of it, today, and I do think it carries around some unhelpful baggage.) The term points to something more than a set of intentions, or even dispositions to form intentions of a

given kind. And because it is more than this, it is something which we can imagine we see manifested in spontaneous or involuntary reactions or, as perhaps in Victoria's case, control over or suppression of such reactions. Yet in this usage, 'character' is clearly not just a name for whatever sub-personal mechanisms might inhibit the startle reflex, for instance. If it has any use here, character is a concept which exhibits a rich set of connections not just with non-intentional action, but also with intentional action, evaluation, reaction, and so on.

I will now move on to a form of behaviour which appears similar on one level, but which has been made into something very different by the local context. *Latab* is a term used in Malaysia and Indonesia to refer to a particular condition, as well as to those who have it. *Latab* episodes are often marked by the repeating of words or gestures which the *latab* observes, by profanity or rude comments, and by what is often called by researchers 'automatic obedience' (Yap; Simons; Lanska). In automatic obedience, *latabs* will carry out orders which are given to them. The orders are often given in a teasing spirit, such as to do something silly or to hit someone else who is present (Simons 165). There are also reports of more fraught cases, such as *latabs* being ordered to take off their clothing in public (H. Geertz 94), or to kill others (Simons 167). (It is not clear to me, though, whether the latter has ever actually occurred, as opposed to being a feature of stories and myths.)

The interpretation and explanation of *latab* remains a matter of significant controversy in debates spanning anthropology and psychiatry. Interestingly, on one prominent interpretation, *latab* is a condition which is built around involuntary behaviour on the part of the person with *latab*. I will provide a brief outline of this interpretation, drawing largely on the work of Ronald Simons, who has developed it in greatest depth (Simons 1996; see also Bakker et al. 2013). If this interpretation is correct, then a significant kind of social identity or way of being has been constructed in Malaysia and Indonesia around involuntary behaviour. However, for my purposes it is not essential that Simons' interpretation be the correct one. Rather, the point is that it appears to be perfectly *intelligible*. If there are problems with his account, they do not stem simply from the fact that he is suggesting that involuntary reflex behaviour lies at the core of this cultural phenomenon.

According to Simons, the characteristic elements of *latab* behaviour should be understood as manifestations of a ‘hyperstartle’ reaction. The startle reflex is fairly universal in humans, but some people startle more easily and more strongly than others. At a certain level of sensitivity (Simons suggests that we should think in terms of a continuum with ‘normal’ startle responses) they can be considered hyperstartlers (Simons 26). The typical *latab* reactions can be observed very occasionally among people with normal startle responses, as a reaction to being startled, and more frequently among hyperstartlers. Amongst hyperstartlers, being startled regularly can apparently increase their sensitivity and also the extent of their startle responses. Simons’ view is that *latab* (which has often been approached as a ‘culture-bound syndrome’) is largely the result of widespread practices of playful (and sometimes less playful) startling in Malaysia and Indonesia, which becomes more targeted once it becomes thought, on the basis of their extreme reactions, that certain people are *latabs*.

According to Simons’ interpretation, the concentration of hyperstartlers in Malaysia and in a few other places and times (one famous example is the so-called ‘Jumpers’ of 19th Century Maine) is due to local practices of startling. Its incidence *within* the relevant populations also appears to be shaped by social-cultural factors. So, for instance, it may be that the relatively lower incidence of *latab* in high status men is explained by the fact that it is less permissible for others to intentionally startle such men (hence even if they were susceptible to hyperstartling, this would be less likely to become known, or to become exacerbated through regular “startle-teasing”) (Simons 232). But the whole ‘condition’, according to this interpretation, is built around reflex and/or involuntary behaviour on the part of *latabs*.⁷²

Nor does it appear, from what I can tell, that locals regard *latab* behaviour as intentional or voluntary. Interestingly, Simons claims that whilst *latabs* may be embarrassed of some of the things they do after being startled, they are not “considered either legally or morally responsible” for such actions (167). These

72 This is, I take it, consistent with the claim that *latabs* may base some of their intentional action on their identification as a *latab*, and that some of what gets classified as *latab* behaviour is not involuntary.

remarks suggest an interesting and complex sense of responsibility and ownership over these actions. However, I am not aware of ethnographic sources which would allow for a more detailed exploration of this (though there are some interesting remarks on the matter in Hildred Geertz's paper on *latab* in Java) (H. Geertz 98-101). Nevertheless, according to Simons' interpretation, in Malaysia, Indonesia and perhaps in other places and times, what we might think of as a 'kind of person' has been created around these reflex reactions.⁷³

A third case brings us closer to our central topic. Erving Goffman is describing local norms governing behaviour in contexts where a number of people are present in the same place, but there is no active engagement between all of them at the same time. In particular here he is concerned with how we should behave when we are clearly in a position to overhear what is going on in an interaction that we are not involved with. He suggests that,

[b]ystanders are obliged to refrain from exploiting the communication position in which they find themselves, and to give visible expression to the participants of the gathering that they are focusing their attention elsewhere. (1966 156)

However, he notes that in practice this is often difficult to achieve. We do not have complete control either over the direction of our attention, or over how this is expressed in our behaviour:

Here I do not want to overstress rational intent in situational behavior. [...] [E]ven the individual who wants to follow this rule cannot completely control the expressed direction of his attention. If his attention is attracted to an accessible encounter, then his attempt to conceal the fact is likely to be visible both to those with whom he ought to be participating and to those whom he ought to be disattending. (156 n.4)

73 By 'kind of person' here I mean that *latab* is something deeper than a mere feature of some people which could be used to pick out a proper subset of all humans, for instance. For an interesting discussion of what more could be meant here, see Hacking (1996), although it should be borne in mind that the kind of 'kind' Hacking focuses on is largely the object and to some extent creation of Western medical and sociological enquiry over the last couple of centuries. It may well be that there are importantly different 'kinds of kinds' in different historical and cultural contexts.

It is unclear whether we should understand the norm here to concern primarily the allocation of attention, or rather the ‘expression’ of the allocation of attention (where it is allowed that what is expressed may not actually be the case). Either way, it seems right to say that one’s attention can be drawn against one’s will in ways that contravene the norm Goffman describes. (It is important here that it would also be possible to contravene the norm through intentional behaviour.) For someone who subscribes to that norm, that a breach was involuntary may provide some mitigation, but it need not remove the sense that *one has breached the norm* with one’s action.

Once again, whether or not we *should* police our own and others’ involuntary behaviour in this way, it nevertheless seems that we often do, and that our doing so is intelligible. Non-intentional actions are possible objects of responsibility and embarrassment, and they can intelligibly be held up to standards of ‘correctness’. Moreover, even when they have not involved *norms*, each example discussed here is affected by features of the local social-cultural context. The worry was that there could not be norms governing non-intentional behaviour, because such behaviour *in principle* lacks practical significance. If any of these examples have been convincing, then this shows that worry to be mistaken.

The ‘internal aspect’, revisited

In the previous chapter, I introduced Hart’s notion of the ‘internal aspect’ of social rules. I argued that his account of the internal point of view on a rule would rule out the existence of at least some kinds of implicit norm or convention. This is because his account required agents to represent the rule itself in taking up a normative attitude towards it. The lack of such representation is characteristic of one of the ways in which a norm can be implicit.

In my initial discussion of Hart, I treated his talk of the internal aspect as the element in his account which was supposed to differentiate a rule from a mere behavioural regularity. I then argued that AI conventions, based as they are on the reproduction of patterns of behaviour, differ from regularities in a way that allows for the existence of both implicit and explicit norms. However, it is reasonable to

think that Hart's talk of the internal aspect was supposed to be doing more than *just* differentiating rules from regularities.

In this chapter I am addressing the question, "What are these things, to us?", with respect to AI conventions as I described them in Chapters 1-4. One way to interpret Hart's talk of the internal aspect of a rule is as getting at the sense in which a rule is something that is, or can be, something 'to us'. It is not straightforward to characterise the requirement that a norm be something to us without presupposing too much. However, it is easy to feel that for norms which are *both* implicit, *and* govern non-intentional behaviour, there *could be* no analogue of an 'internal aspect' to them. That there is simply nothing to go on to characterise an 'internal perspective' on such putative norms; that even when they are our norms, we can only treat them as something 'external' to us, unless we change them by rendering them explicit, for instance. The relevance of this thought is that if AI conventions were always 'external' in this way, then they could not meaningfully provide standards of correctness for our actions. In this section I will briefly sketch why it would be too hasty to conclude that there is nothing to go on in characterising the internal aspect of implicit AI conventions.

What is it *like* to follow local interpersonal distance norms? We have already seen that there will not be a single answer here. Some people, such as the person with OFC damage mentioned in Chapter 4, have to make a conscious effort to adopt the locally normative distances. Others will, for a variety of reasons, be uncomfortable with the distances typically prescribed for a given kind of social interaction. This might be for individual psychological reasons (and if so, either general or quite context-specific), or it might for instance have to do with asymmetric expectations reinforced by power imbalances, as noted in Chapter 4. There is also the rather different sort of case where one has gotten used to the local norms in one place, but then travels to a society or culture where other norms are in force. Finally, anyone might be prompted by a variety of circumstances to doubt or to dislike the proximity or distance of others on a particular occasion.

Each of these situations will typically feel quite different to those who find themselves in them. I will try, though, to describe what I take to be a fairly standard

experience for those who are not alienated from local interpersonal distance norms in one or more of the ways mentioned above.

What it is like is that for the most part we do not have to actively think about where to stand in relation to others, in order to maintain normative distances. Because a large range of non-normative distances is already excluded from consideration for us. It seems correct to me to say that I *would not even think* of standing very close to a stranger in a large, empty room, unless some exceptional circumstances prompted me to. And if it did occur to me as an idle thought, I would find it very difficult to entertain it seriously, or bring myself to do it – again, unless there was some exceptional reason. Something like this seems true of many bodily movements which we are physically capable of performing, but which are excluded from our local cultural repertoires.

If this description is along the right lines, then it suggests a way in which we might make sense of the ‘internal aspect’ even of implicit norms governing non-intentional behaviour. The description I just gave is reminiscent in many ways of what Bernard Williams says about ‘moral incapacities’. For Williams, a moral incapacity can be realised in the thought “I can’t do that”, but it can also be realised through limits on the thoughts which one will entertain, or entertain seriously (1993 60-62). Despite this, he thinks that some such incapacities can be genuinely ‘moral’ incapacities, rather than for instance cognitive incapacities. One of the things that makes it a moral incapacity is the fact that an agent can in some sense be *identified* with her incapacity, which is something that doesn’t apply to merely cognitive or physical incapacities.

For Williams, talk of moral incapacity is not just a way of reporting on the subject’s attitudes towards doing something. Rather, it is supposed to be a genuine incapacity – which can hence *explain* her not doing something – but one which is intimately tied up with her identification with the incapacity itself. But as we have seen, the incapacity may manifest not just in thought, but also in certain *absences* of thought. Furthermore, identification with a moral incapacity is not a matter of mere decision. If it were, it would not be a true incapacity. So, moral incapacity, and the related notion of practical necessity, are “neither simply external to the self, nor yet a

product of the will” (Williams 1981c 131). If this is all correct, then moral incapacities appear to open up the possibility of explaining certain behaviour without requiring that either the *behaviour* itself is represented by the agent (as in intentional action) or that the *incapacity* is explicitly represented by the agent (as with explicit norms).

Interpersonal distance norms are different in important ways from the examples which Williams has in mind. But my suggestion is that perhaps we can generalise his concept of moral incapacity to provide something like an account of ‘practical incapacity’ more broadly. If so, identification with a practical incapacity might allow us to make sense of how an implicit norm governing non-intentional behaviour can nevertheless have an ‘internal aspect’, and thereby provide standards of correctness for our actions. More work would be required to flesh this suggestion out fully, but it presents a possible avenue for making sense of an ‘internal point of view’ on implicit norms governing non-intentional behaviour.

Reflection and reform

Marcel Mauss reports that one of his school teachers used to shout at him: “Idiot! why do you walk around the whole time with your hands flapping wide open?” (72). It is not clear from the text what alternative the teacher would have preferred, but I believe it was to close his hands into fists. Mauss concludes that, “[t]hus there exists an education in walking, too” (72). Walking itself is typically an intentional activity, but what about those movements which maintain a culturally-determined disposition of the hands? The teacher’s invective will have been designed to get Mauss to reflect on the position of his hands, and to intentionally modify it. But it seems fair to assume that the long-term aim was to inculcate a bodily discipline through which he would automatically adopt the favoured disposition of his hands.

I was concerned in the previous section with implicit norms governing non-intentional behaviour. But it seems likely, as in the case reported by Mauss, that there can be *explicit* norms governing non-intentional behaviour, too. For such norms it will be easier to give an account of their ‘internal aspect’. The norm itself is

represented by Mauss and his teacher, and can be a direct object of approval or disapproval, for instance.

The personal cruelty of Mauss' teacher, and the absence from Mauss' text of a sense of what might justify the norm itself, can be distracting here. So I will return to the case of interpersonal distance norms. I suggested in the previous section that the idea of identification with a practical incapacity might point towards a way of making sense of implicit identification with a norm governing non-intentional behaviour. But what happens when the norm becomes an object of explicit reflection?

One might think that explicit reflective endorsement of a norm would render the behaviour which it governs intentional. I do not think that this follows, however. It may be, as in the Mauss case, that in order to gain a habit on purpose one must do certain things intentionally. But it does not follow from this that once the habit has been developed, its manifestations must be intentional actions. And the presence of explicit endorsement of the habit does not change this. (There is no obvious reason to treat endorsement as a 'standing intention' to act in a certain way, for instance.)

What this shows is that there are at least two levels at which reflection can be brought to bear, when following a convention. There is the convention itself, and there are the particular actions which manifest, or perhaps go against, the convention. The nature of AI conventions makes it clear that these two can come apart. The AI convention is a real entity, and so a possible object of reflection, although as we have seen, they do not in general *need* to be represented or reflected upon to exist. It is distinct from the particular actions which manifest it, and so can be reflected upon independently of reflection on those actions. The particular actions which manifest or contravene the convention are also possible objects of reflection. Again, we have seen that a particular action need not be reflected upon in order to count as a manifestation of an AI convention. Non-intentional convention-following actions are a prominent example of actions which are unreflective in this second way, though there will also be intentional actions where we don't reflect on the fact that in performing them we are following a convention.

Having a robust distinction between the AI convention itself and the actions which manifest it is crucial for capturing important aspects of their normative role. The situation in which an agent reflectively endorses a rule, and takes this as a basis for reflectively endorsing her rule-following action as well, will be the focus of Chapter 7. Here I will briefly explore some of the possibilities which arise when an agent comes to reflect on a norm which has previously been implicit, and which she has been conforming to through behaviour which has been non-reflective in the particular case.

One possibility here is that reflection on the norm itself will leave everything else more-or-less as it was before. Whilst writing this thesis I have had a number of conversations about interpersonal distance norms with people who did not seem to have given them any thought before. Reactions to the realisation that they exist varied. One fairly common reaction, though, has been to treat this as an interesting but fundamentally untroubling piece of news. The realisation that their behaviour has been shaped by these norms does not provoke a radical reassessment of their past behaviour, or a commitment to do anything differently in the future. Nor, from what I can tell, has it prompted them to try to be more conscious of how they are regulating their distances from others in particular situations. That is, they seem to have found that they can endorse the existing norms, and continue to follow them unreflectively in the particular case.⁷⁴

There are different ways in which this kind of reflection might leave everything as it was. Mere acquiescence need not be a sign of having come to reflectively endorse a formerly implicit norm. If one learned that one had a habit that was mostly benign but sometimes foolish, this need not provoke any attempt to change the habit, if doing so appears to be more trouble than it's worth. There are other reactions,

74 If they think about it a bit more, it might cause them to re-evaluate certain particular situations from the past. Perhaps they now understand that the person from another country who stood 'too close' was not being rude, but rather was used to a different set of norms. Their own behaviour in different cultural contexts may also appear in a new light. If they now divide their behaviour in these situations into (albeit perhaps in a very thin sense) 'correct' when they were at home and 'incorrect' when they were away, this is evidence that they are regarding the norms they have just learned about *as norms*.

though, which appear to differentiate the two cases. It is possible that one will come to see one's past behaviour in a new light as being in some sense 'correct', even when it might have appeared to have nothing going for it considered independently of the existence of the norm. *This* reaction would not make sense if one regarded the past behaviour as the manifestations of a mere habit. The latter reaction is in fact indicative that it is the AI convention itself that one has come to reflectively endorse. (We will explore this connection in greater detail in Chapter 7.) This kind of reaction can also provide evidence that a practical incapacity which one possessed *before* becoming explicitly aware of a norm consisted in implicit identification with the norm itself, as opposed to being a merely individual psychological incapacity.

So the case where reflection leaves things largely as they were is an important one. But it is not the only possibility. One might, of course, reject a formerly implicit norm out of hand, once one becomes explicitly aware of its existence. Slightly more interesting will be the case where one finds something unsatisfactory about the existing implicit norm, which suggests reform, rather than outright rejection. I will highlight two interesting possibilities.

One reason for discussing interpersonal distance norms is that they are *relatively* innocuous. By this I mean that there seem to be good reasons for having *some* norms governing interpersonal distancing. They can help to regulate intimacy, and to preserve a dependable sphere of privacy and bodily autonomy. But on the face of it there are many different ways in which these functions could be fulfilled, depending in part on broader aspects of the local culture. This is, I take it, part of the reason why we have local norms governing distances.⁷⁵ However, this is *not* to say that any old set of interpersonal distance norms which manages to become established will be good, or be binding on those who fall under them.

The most obvious ways in which a given set of distance norms might be unacceptable involve problems with the overall pattern prescribed by the norms. Norms which can emerge and proliferate implicitly and through non-intentional behaviour will tend to take a certain kind of shape. They may have a complex structure, but there will be limits to this complexity, and to the kind of flexibility they

⁷⁵ cf. the discussion of 'arbitrariness' in Chapter 1.

can have when applied to particular cases. Explicitly reflecting on a norm which had been implicit might allow us to design a better norm for the purposes. (The examples involving an asymmetry of social power, touched on above, suggest a different kind of way in which things might go wrong on this level.⁷⁶) Actually *changing* the norm to fit the new design may require various actions which are reflective in the particular case. However, once it is in place, in many cases it will be possible to let routine take over and follow the new, reflectively endorsed norm in an unreflective way from one occasion to another.

But there is another, more subtle way in which norms might be found unacceptable, particularly once their existence has been explicitly recognised by those who follow them. We have already seen that there might be particular people for whom, or particular contexts in which, an *otherwise reasonable* set of norms will cause problems. As has been hinted at already, we can expect that if a norm or convention is doing any real practical work, then it will justify some actions that would not have much to recommend them on independent grounds.⁷⁷ So pointing out that following a norm will lead us in certain particular cases to what would otherwise seem non-optimal behaviour need not be a reason to reject that norm. However, some of the problems we have encountered with existing interpersonal distance norms in the contexts I have described seem to be more than just coincidental negative outcomes. Moreover,

76 Further, related but distinct, options come to mind, too. It might turn out, as it often has in other cases, that contradictory norms are applied to members of a particular group, putting them in a ‘double bind’ where nothing they do will be seen as correct. I mentioned in passing that it seems possible that this would apply to distance norms for women in a Western European context – perhaps any distance they maintain from an unfamiliar man in certain contexts is liable to be judged either standoffish or inappropriately intimate. I have not been able to find much work on this specific question, but the broad pattern of the double bind would be a familiar one. I take it that if this were the case, it would provide grounds for modifying our norms. Consciously modifying our norms will require some reflection, but it seems that the modification required in this kind of case would be compatible with following the *new* norms in an unreflective way in day-to-day interaction.

77 This point is discussed at greater length in Chapter 7 below.

they each seem to indicate that the source of the problems is the fact that whether or not we are explicitly aware of the norms themselves, we tend to *apply* them unreflectively from case to case.

I have three issues in mind, each of which has been touched upon in the preceding discussion. One is that some people will find it difficult or impossible to learn to follow their local distance norms in the ‘unreflective’ way we have been describing. We saw this in the case of people with OFC damage, but it also appears to be true of many people who are considered to be on the autistic spectrum, and quite likely of other groups, too. Despite this explanation, such individuals may be perceived by others as behaving ‘inappropriately’ in their distance-regulation behaviour. The second issue is that people will have different individual requirements in terms of personal space. For instance some people will, for any number of reasons, require more space to be comfortable than is typically allocated to them by others. Finally, we saw in the case of the assaults on the dementia ward that perceived invasions of personal space can be interpreted as threatening (sparking ‘retaliation’) even when no threat is presented by the person who ‘invades’ the space of the other.

These are the kinds of consequences that make a claim on our efforts to justify a norm. And, importantly, each looks as though it would be best addressed by our being more reflective about our distance-regulation behaviour *in particular cases*. In the first two cases, we need to try to take individual differences into account in various ways. In the third, we need to introduce some space to be able to ask, “is there actually a threat here?” rather than letting our reactions be dictated by mere feelings of threat.⁷⁸

If we find ourselves following this line of thought, reflection on the norm itself seems to introduce a kind of instability in our normative situation. We are thinking that there is good reason to have some norms governing interpersonal distance. We can also see, as comparison with the person with OFC damage shows, that it is very

78 I don’t want to imply that pre-emptive assault would be an appropriate response even if the other person *did* represent some kind of threat. However, assault is, here, at one extreme of a range of possible responses to personal space invasions. Other, less violent, responses might be more justified when a threat is actually posed.

difficult to be fully reflective from occasion to occasion about how we are regulating our interpersonal distances. But we are also starting to feel that once the question has been raised, we cannot endorse a set of norms for interpersonal distance which *can* be manifested by actions which are themselves carried out unreflectively – we feel the need for more flexibility than unreflective norm-following can provide. Whether or not one shares these concerns about this particular example, the question arises: what can we do in a situation like that?

I do not know what the answer to this question is in general. I find it difficult to imagine how we could proceed if we tried to bring interpersonal distance regulation under active conscious control in *each* particular case. Perhaps instead it would be possible to become more sensitive to appropriate cues – such as discomfort in our interlocutors – that will prompt us to reflect in those cases where it might be needed. This sort of sensitivity looks like it would be compatible with following the familiar norms unreflectively in cases where the cues are not present. Such a balance would not look quite like our model of fully unreflective norm-following, nor like the model of reflective rule-following discussed in the next two chapters. How to understand this kind of case will have to remain as a question for future work.

Social phenomena and formal individualism

In the previous chapter I suggested that there is a sense in which figuring in social explanations is not sufficient for an entity to count as a ‘social phenomenon’. I illustrated this through a comparison between conventions and river deltas. Both can figure in social-scientific explanations. But this does not show that deltas are ‘social phenomena’ in an interesting sense. By contrast, we do tend to think that conventions are social phenomena. How can we understand this contrast? What relevant property do conventions have that deltas lack?

This is not the place to try to give a general account of what makes a ‘social phenomenon’ in the relevant sense. Instead, I will just outline what makes it plausible that AI conventions count.

One way of demarcating the realm of social phenomena is through a kind of *significance* which can be imbued into an entity by its origins. This approach seems to be at play in certain work by Max Weber and Charles Taylor. This work suggests that the genuinely social entities are those which are the proper objects of ‘subjective understanding’ or ‘interpretation’ (Weber 88; Taylor 12-14). In describing the objects of interpretation, both give a special role to *intentional action*. The thought seems to be that intentional action and its upshots are imbued with significance or ‘subjective meaning’ through the distinctive mental/agential capacities which give rise to it. This is what makes them appropriate objects of interpretation, and hence genuinely social entities.

Whatever exactly we make of the notion of ‘interpretation’, and its relationship to the human sciences, we have already seen in Chapter 4 that social enquiry need not mark the kind of sharp distinction between intentional and non-intentional action which seems to be implied by this line of thought in Weber and Taylor.⁷⁹ I am interested here in a different approach to demarcating the realm of social phenomena.

Instead of seeking the difference from the interpretive perspective, we can adopt a practical perspective. We have seen during the course of this chapter that AI conventions – even the implicit ones which govern non-intentional behaviour – can have an ‘internal aspect’. They are a possible object of identification (or alienation) for individual human agents. *This* is the key feature that distinguishes them from deltas.⁸⁰

Being the object of this kind of identification does not seem to be *sufficient* to count as being a social phenomenon, though. One might equally be identified with a purely

79 I don’t pretend to have done proper justice to the work of either Weber or Taylor here. But a full discussion of their work on this topic would take us far beyond the scope of this thesis.

80 That is, according to many understandings of deltas and our relations to them. Some people speak of the forest in which they live as a parent to them, and I do not think this is easily diagnosed as a mere mistake or metaphor (cf. Ingold 47). I believe that such cases make a serious claim for our understanding, but I cannot explore this further here.

personal, psychological disposition, for instance. However, we saw in the discussion of interpersonal distance in the previous chapter that when researchers talk about norms they are (at least some of the time) talking not about psychological states or dispositions which happen to be possessed by a number of people, but rather about public entities which govern the behaviour of a multiplicity of agents. The same distinction applies when it comes to the objects of an agent's identification or alienation, and we have seen that an agent can become identified with *the norm itself*. I will say a little more about what this looks like in Chapter 7.⁸¹

If all of this is correct, then we can give a more positive characterisation of why a given AI convention governing interpersonal distances in a given place and time counts as a 'social phenomenon' in an interesting sense. First, it is because unlike the delta this AI convention is a possible object of identification or alienation for agents in that place and time. Second, it is because it is *the AI convention itself* that is the object of identification. And we know from what has gone above that these AI conventions are 'public' entities, they are the creation of many hands, and they govern the spatial relationships of multiple agents. When you and I maintain a given distance in conversation because we are following the same norms, our behaviour therefore has a common source in a way that differs from the kind of 'common source' that might be found in a harmony of our personal psychological states. This is why the AI convention is a 'social phenomenon'.

With this answer in place, I want to pause to bring out one of its further consequences. Bernard Williams, in his discussions of 'formal individualism', can also be seen as pointing towards a way of understanding social phenomena in terms of their practical import for us (1995a; 1995b). He seems to have thought, in those papers, that social phenomena could only have practical import for us by shaping the *intentions* on which we act. We have seen that this is not correct. Social factors can

81 The matter is a bit more complicated in the case of practical incapacities, where there may not be much to go on to differentiate identification with a norm from identification with a personal, psychological incapacity. There is undoubtedly more work to be done here. However, we saw above that certain reactions to becoming aware of the norm will indicate that it, rather than something more personal, was the real object of the previously implicit identification.

shape our *non*-intentional actions, too. And in fact we drew on some other work of Williams' – his discussion of moral incapacity – to see how this can be possible. So there is good reason to think that he could modify his position to allow for a broader understanding of practical import.

However, the picture we have developed in this chapter seems to provide a much more fundamental challenge to the core idea of formal individualism. This is a commitment which Williams suggests anyone should accept if they find room in their picture of the world for individual human agency at all. (Williams contrasts such pictures with those which subsume individual agents entirely under macro phenomena like 'structure' (cf. James 1984, esp. Ch.IV-V). He could also, as we have seen, have contrasted them with pictures that dissolve human agents into collections of memes and genes.) The commitment which Williams thinks comes along with commitment to the existence of human agents is a commitment to the claim that "the actions of an individual are explained in the first place by the psychology of that individual" (1995a 86).

This commitment is interpreted by Williams as showing that if a social phenomenon explains an individual's actions, then it must do so *by* explaining the intentions with which she acts (1995a 86; cf. 1981b). We have seen that he can and should drop the restriction to intentions here, so as to allow room for non-intentional action. But the core claim would remain that individual action is explained in the first place by psychology, now with a broader understanding of what the relevant psychology might involve. This claim Williams treats as "inoffensive to the point of triviality" (1995a 86). However, the picture of norm-following we have developed in this chapter and the previous one seems to provide a radical challenge to Williams' core claim. If we explain an individual's action by reference to her identification with an AI convention, then there is a psychological element to our explanation, but there also seems to be an irreducible reference to something non-psychological, too: the AI convention itself.

Conclusion

Chapters 4 and 5 have built upon and begun to apply the ontology of AI conventions developed in Chapters 1-3. In doing so they have also built a bridge to Chapters 6 and 7, which are concerned with the role of social practices in practical reasoning.

In this chapter I have defended the claim that it is possible for AI conventions to have a practical role for us at all. The main objections to this claim came from the inclusion of implicit norms, and norms governing non-intentional behaviour, in Chapter 4. The worries were that non-intentional action could not be subject to socially-determined standards of ‘correctness’ at all, and that even if it could they could not come from implicit norms. In response to the first worry, I described in some detail a number of cases where we seem quite intelligibly to treat non-intentional actions as having a practical significance which differs from any possessed by mere bodily processes such as digestion. In response to the second worry I sketched how we might make sense of questions of identification and alienation even in relation to implicit norms governing non-intentional behaviour. This shows that such norms can possess an ‘internal aspect’.

Chapter 4 was concerned with the role of AI conventions in social explanation. Chapter 5 began to investigate their practical role. But the answer to the question of how AI conventions, and not deltas, count as ‘social phenomena’ began to show that the division between these two applications may not be quite as neat as we sometimes imagine. We said that AI conventions are social phenomena because unlike deltas they are possible objects of identification and alienation for individual agents, but unlike personal psychological states they are importantly *public* objects. This answer drew upon the discussion of the internal aspect in this chapter. But it also drew upon the arguments of Chapter 4, which showed that there is an explanatory role for *the convention itself*, distinct from that of a collection of psychological and ecological factors.

This combination – whereby conventions are in some sense external to us, but also, given the right circumstances, are things we identify with – underpinned our

challenge to the primacy of psychology found in Bernard Williams' 'formal individualism'. It will also, as the arguments of Chapter 7 will show, be crucial to understanding the distinctive sense of 'requirement' that we can come to associate with the demands of a social practice. I begin to describe this sense of requirement in the next chapter.

Chapter 6. Rules and the Sense of Requirement

The preceding chapters have focused on conventions and norms. I suggested that these form a family, along with rules. Each is a kind of abstract individual, created through the reproduction of patterns of behaviour. I will not try to provide a full characterisation of the possible differences between them, here. However, one important difference is that rules seem to have a much more intimate connection with reflective awareness than norms or conventions. Whilst it is possible for agents to be aware of the conventions and norms they follow, we have seen that it is also possible for these to be quite radically implicit. By contrast, it seems plausible that the paradigmatic rule-follower is aware of the rules, *and* aware that she is following them. This will be an assumption in what follows.

In this chapter and the next I turn to the practical role of the rules of ordinary practices, so understood. I will describe what I take to be an important part of their normative profile – namely, the fact that rules are typically associated by those who follow them with a sense of ‘requirement’ to do what the rule specifies. In this chapter I will develop an account of what this sense of requirement consists in. I argue that this description of the sense of requirement speaks in favour of interpreting this sense as reflecting something distinctive about the *structure* of reasons facing agents who fall under rules.

I will then argue against a prominent set of attempts to explain the sense of requirement without appealing to anything distinctive on the level of reasons. The views I focus on try to explain the practical reasons generated by the rules of ordinary practices in terms of an agent’s ‘ends’, plus a minimal social ontology of so-called ‘constitutive rules’. (Constitutive rules here are conceived of as lacking any direct normative or practical import.) The rejection of this view will set the scene for

my positive account of the sense of requirement in the final chapter, which will draw on the ontology of rules developed in the first three chapters above.

Rules in practical reasoning and practical philosophy

‘Rule’ is a relatively rare example of a normative concept which figures explicitly in our thought and discourse, perhaps not every day, but at least in everyday contexts. Not just when playing a game, but also when joining a new school, university, workplace, or campaign group, we ask ourselves or others: what are the rules, here? And many of us, much of the time, treat these rules as reasons, or as sources of reasons, for action. In doing so, I suggest, we typically treat them as reasons of a distinctive kind, which I describe as being associated with a sense of ‘requirement’. I will argue below that this sense of requirement is naturally understood as reflecting a feature of the *structure* of the practical reasons which confront those who find themselves in the scope of a valid rule.

This is, at least, how many of us ordinarily regard the rules of ordinary, profane practices and institutions, such as the ones mentioned above. Philosophically, matters have seemed less clear. The idea that rules are sources of requirement can easily seem – wrongly, in my view – to be in tension with the conviction that sometimes we ought to break the rules. It has widely, and correctly, been thought that it *must* sometimes be right to break the rules of ordinary practices. The view that rules must *never* be broken seems applicable, if at all, only to the rarefied practices associated with justice, such as promising.⁸² Those who cannot accept this even for moral rules have tended to be sceptical of rules in practical reasoning in general. Whereas those who have accepted it have tended to bifurcate their treatment of rules in such a way as to provide an account of ordinary practices which fails to capture the sense of requirement. I will argue that, properly understood, the sense of requirement is

82 I have mentioned games on the one hand, and promising on the other. In my concern with ‘ordinary’ practices I will stick mainly to games, schools, etc. I will not discuss good systems of law, which have seemed to many to be a special case, somewhere between the sacred and the profane.

compatible with the view that rules can be broken, meaning we can resist both of these revisionary positions, without landing in an untenable absolutism about the demands of ordinary rules.

This chapter will describe in greater detail the association of rules with requirement which is implicit in much of our everyday practical thought. I will make a preliminary case for thinking that we should understand this association in terms of the distinctive *structure* of practical reasons associated with rules. This suggests a problem: how can we talk about the structural peculiarity of practical reasons in the context of rules, prior to having a theory of practical reasons in general? The literature contains a variety of accounts of reasons, which will give widely different senses to talk of ‘the structure of practical reasons’. This makes it difficult to even raise the question of the structural peculiarity of rules implicit in *everyday* practical thought.

This problem shapes the account I offer. In describing our everyday conception of rules, I aim to give some content to the claim that rules are structurally distinctive, without presupposing any particular theory of practical reasons in general. This picture can then be used to motivate a more theoretical account of practical reasoning in the context of rules, or else as a standard to judge competing accounts of reasons, as a datum to be explained, or explained away.

This problem also limits the scope of my claims. In claiming that rules are associated with a distinctive structure of practical reasoning when compared to a range of other reasons, I do not suggest that this structure is *unique* to the case of rules. In this respect I take my position to be analogous to Hume’s broad distinction between natural and artificial virtues, or to Anscombe’s grouping together of rules, rights, and promises (Anscombe 1981f). I also do not claim that the distinction I describe is *exhaustive* of the realm of practical reasons.

Chapter 7 develops a more positive account of the practical role of ordinary rules, and an explanation of the sense of requirement attaching to them. It introduces this through a critical discussion of Michael Thompson’s work on ‘practical generality’. Thompson explains the distinction between instrumental and practice-based

explanation of actions in terms of the metaphysical distinction between particular and general. I suggest that the metaphysical generality of rules – which is supported by the account I developed in Chapters 1-3 above – can help to explain the sense of requirement attaching to them. The limitation of Thompson’s account is that it only applies to rarefied practices, such as the practice of promising. I will argue that we can modify certain of his key commitments in order to provide an account of the practical role of rules which includes the rules of ordinary practices.

The basic picture

Philippa Foot describes two distinct sets of norms attaching to the use of words like ‘should’ and ‘ought’ in advice-giving scenarios. In ‘Morality as a System of Hypothetical Imperatives’, she claims that these norms attach to two different ‘uses’ of these words in such contexts. What Foot calls the ‘hypothetical use’ of ‘should’ or ‘ought’ is answerable to the *ends* of the agent to whom the advice is given. ‘Ends’ are characterised as those things which an agent wants or desires, and also what is in her self-interest (1978a 158). The diagnostic mark of a hypothetical use of ‘should’ or ‘ought’ is that if we have told a person that he ‘should’ do something in this sense, then,

we must be prepared to withdraw our statement about what he should do if we find that the right relation does not hold between the action and the end – that it is either no way of getting what he wants (or doing what he wants to do) or not the most eligible among possible means.⁸³ (159)

The statement is also to be withdrawn if we were mistaken in thinking that the agent had the end in question.

By contrast, a non-hypothetical use of ‘should’ or ‘ought’ is not, as such, answerable to facts about the agent’s ends:

83 Need we always *maximise* the ‘eligibility’ of our means? I would suggest not (cf. Anscombe 1957 §44). Perhaps that point is specific to giving advice.

When we say that a man should do something and intend a moral judgement [here we are treating ‘moral judgement’ as a kind of non-hypothetical judgement] we do not have to back up what we say by considerations about his interests or his desires; if no such connexion can be found the ‘should’ need not be withdrawn. (Foot 1978a 159)

Foot’s interest is in the idea that the ‘non-hypothetical’ nature of moral judgements might capture their alleged ‘inescapability’; specifically, the thought that “the agent cannot rebut a suggestion about what, morally speaking, he should do by showing that the action is not ancillary to his interests or desires” (159). She thinks, though, that non-hypotheticality cannot exhaust what’s special about moral judgements, because there are non-hypothetical ‘should’s which do not have the “special dignity and necessity” commonly attributed to moral judgements (160). She mentions examples like the rules of etiquette, or the rules of a club to which one belongs:

The club secretary who has told a member that he should not bring ladies into the smoking-room does not say, ‘Sorry, I was mistaken’ when informed that this member is resigning tomorrow and cares nothing about his reputation in the club. Lacking a connexion with the agent’s desires or interests, this ‘should’ does not stand ‘unsupported and in need of support’; it requires only the backing of the rule. The use of ‘should’ is therefore ‘non-hypothetical’ in the sense defined. (160)

I suggest that non-hypotheticality in something like Foot’s sense captures something important about the way in which we normally think of rules. But it is important to note that this is not just confined to advice-giving scenarios. The non-hypotheticality of rule-backed ‘oughts’ manifests itself also within first-person deliberation. As children, the central occasions for learning what rules are occur when a rule is in force which conflicts with what we want to do. In adulthood it is still common to feel that one should do something one doesn’t want to, and which serves none of one’s interests, simply because some rule dictates it.⁸⁴ As Rawls highlighted, albeit in

84 ‘Want’ can bear a wide range of interpretations. My point about the possibility of conflict is not supposed to depend on giving it an especially narrow reading.

different terms, we can also *explain* an agent's acting contrary to her ends, in cases where we can interpret her as knowingly following a rule (Rawls 5).

However, the practical role of rules is not limited to cases where they conflict with an agent's ends. Where a valid rule requires of an agent that she perform some action which also happens to further some of her ends, if she performs the action 'because it furthers her ends', we would tend to think that something has gone wrong. Rawls says of someone who breaks her promise because it is 'best on the whole' to do so in the circumstances, that we would question whether she really understood what "I promise" meant (Rawls 17). But, likewise, *keeping* one's promise because "doing so in *this* case will have the best consequences [...] is said, quite rightly, to conflict with the way in which the obligation to keep promises is regarded" (13). Korsgaard makes a similar point, suggesting that it would be at least "a little odd" to say of someone who took a required logic class that she took it because of the merits of studying logic, if in fact that class was *required* (1996 27).

This last point hints at what makes the difference that Foot highlights seem like a *structural* one. We might wonder why it seems reasonable to group together things that an agent wants, and things that are in her interests, under a common heading of 'ends', and to contrast them with "the backing of a rule" as sources of warrant for an 'ought' claim. Rawls' and Korsgaard's remarks bring out that this contrast seems natural insofar as we associate the notion of a rule with some idea of *requirement*. Understanding the basis for non-hypotheticality thus leads us to consider this richer practical framework.

We can make sense of *an* idea of 'requirement' in the context of ends-based practical reason. Sometimes, at some level of description, there will only be one action which an agent can take to realise one of her ends; an action might be a 'necessary' means to some end. One is not in general 'required' to take the means to any given end one might have, as is brought out nicely by some remarks of Anscombe's:

Practical grounds may 'require' an action, when they shew that *only* by its means can the end be attained, but they are just as much grounds when they merely shew that the end *will* be obtained by a certain means. Thus, in the

only sense in which practical grounds *can* necessitate a conclusion (an action), they need not, and are none the less grounds for that. (2005 120)

However, if there is some way of weighing up all of one's ends, and deciding that some one or some combination of them is particularly pressing or important, then we might say in some sense that one is 'required' to take the *necessary* means, if any, towards fulfilling *those* ends.⁸⁵ (Though, as Anscombe's remarks suggest, it looks as though the existence of this sort of requirement will tend to be 'accidental', insofar as nothing about the way in which ends provide reasons guarantees that it will obtain.)

The idea of requirement gestured at by Rawls and Korsgaard does not work like this. Their sense of requirement shows up by making it in some sense *inappropriate* to give certain considerations a role in our deliberation, advice, and explanation, which it would normally be appropriate to give such a role. By contrast, those things Foot groups under 'ends' – an agent's wants and her interests – are not generally exclusive in this way. Let's say that my most pressing end today is to exercise, and walking is the only feasible means I have to do so. Nothing about walking for exercise prevents me from also walking for pleasure, or from calculating my route so as to allow me to carry out some less pressing errands. These ends all combine to give me even more reason to do what I am doing. Even if I can pursue only my most pressing end effectively, there is nothing in general stopping me from *considering* how to pursue others at the same time, or considering whether it is indeed my most pressing end, given the others I have.

In general it seems to be characteristic of an agent's desires and interests – and, I would suggest, the desires and interests of others, and why not also 'engaging with things of value'?⁸⁶ – that the pursuit of one such end does not preclude one from

85 Sometimes the notion of a 'conclusive' reason is invoked to describe a situation like this. (See e.g. Raz (2005 4), although his views on 'ends' bear a complex relationship to the picture I am describing.) It does not seem to me that anything about the very idea of a conclusive reason implies that for any agent at any given time, there will be some action such that she has a conclusive reason to perform that action at that time.

86 For the latter, see Raz (2005 3-4).

weighing other ends in the balance at the same time.⁸⁷ Rawls' and Korsgaard's sense of 'requirement' *does* preclude this kind of weighing, though it remains to be seen how this works in detail.

If the formulation of this point about requirement seems vague, it is because I deliberately want to give a characterisation of it which captures something about our everyday conception of rules, but which allows the exact nature of the phenomenon to be cashed out in different ways in a theory of practical reason. The next chapter will elaborate on one particular way of fleshing out this more intuitive datum. We can also interpret other accounts in the literature in this light.

For instance, Rawls tells us that when a rule is in force, it is no longer appropriate to justify one's action by appeal to what is best 'on the whole' in that particular situation (say, by appeal to the utility of the consequences of that action). Rather, the particular action is justified by appeal to the rule, and the rule itself (or the practice to which it belongs) is what gets subjected to the utilitarian test. Raz suggests that rules are generally marked by a positive 'first-order' reason to perform some action in a given kind of situation, along with a 'second-order' reason *not* to act for some other range of first-order reasons in that kind of situation (Raz 1999 Ch.2).

Hume's account of the distinction between natural and artificial virtue, and Anscombe's account of the peculiar role of certain kinds of 'stopping modal' can also be seen in part as elaborations of this theme (Hume, Book III; Anscombe, 1981f; 1981a).⁸⁸ Each of these writers is trying to do many different things, of

87 This may not be true on all accounts of the relationship between ends and practical reasoning, but it would take us too far afield to address the question fully. It is also worth highlighting that nothing here is supposed to rule out the idea that the *contents* of various ends might conflict in a way that means a rational agent ought not to adopt all of these ends simultaneously.

88 Bernard Williams's (1981a) discussion of 'Justice as a Virtue' could be seen as marking a similar distinction, although he appeals in the first place to a certain kind of *circularity* associated with justice, compared to the other virtues. The theme of circularity in connection with rules (etc.) is one which figures in a number of the other writers mentioned here, and the relationship between circularity and what I'm calling

course, but having this everyday picture of rules allows us to perceive some common features in the context of what are otherwise quite divergent conceptions of practical reason in general. Each writer is marking something like a *structural* difference within the realm of reasons, which seems to help explain the sense of requirement we have attaching to rules.

However, it is not obviously obligatory to preserve this apparent datum concerning rules and requirement.⁸⁹ So, for instance, Foot (1978a) appears to suggest that there is in fact nothing interesting on the level of *reasons* to mark out the rules of etiquette, of clubs, or even of morality. Of course, she does mark some distinction here, as we saw at the beginning of this section. But the difference in assertibility norms between the two uses of the term ‘should’ is not, for her, mirrored in a distinction between kinds of practical reasons. She claims that, “[i]n the case of etiquette or club rules it is obvious that the non-hypothetical use of ‘should’ has resulted in the loss of the usual connexion between what one should do and what one has reason to do” (Foot 1978a 168 n.8).

Foot does not offer a definite account of the reasons associated with rules, but she seems to take it as uncontroversial that the rules of etiquette and of clubs will give agents reasons to act *only if* they can be shown to engage with some of the agents’ ends. Hence in these cases, the non-hypothetical use of ‘should’ would not be associated with non-hypothetical *reasons*, and this assumption is used to launch her more controversial challenge against the categorical imperative in morals: if there *is* a

‘requirement’ is a topic deserving more time than I can give it here.

89 There are further options that intuitively fall ‘between’ Foot’s view and those described in the previous paragraph. We might think here of Pettit’s attempt to give a consequentialist account of rights. This attempt involves introducing a significant gap between “the function which determines what is the right decision”, on the one hand, and “the function which ought to be applied in decision making”, on the other (Pettit 1988 43). Such a gap opens the possibility that one ought to deliberate in the restricted way described by Rawls and Korsgaard when faced with a valid rule, even though this restriction was not straightforwardly reflected in the realm of reasons itself. This is an intriguing possibility but one which I do not have space to consider further here.

difference between the ‘ought’ of morality and the ‘ought’ of etiquette, what is it? (e.g. 164; 168 n.8)

Foot’s assimilation of all rule-based reasons to the case of hypothetical reasons was motivated at the time largely by the thought that the things she called ‘ends’ were the only unmysterious sources of practical reasons. Something like this thought is still found. But in making this assimilation, Foot’s view denies that the sense of requirement attaching to rules is to be explained by reference to a structural difference at the level of reasons. This leaves two questions. The first is how and why a rule might ever provide an agent with a hypothetical practical reason. The second is why, if these reasons are only ever hypothetical like all other practical reasons, we have this sense of requirement attaching to them.

On the first question, Foot says very little.⁹⁰ On the second, she might appeal to her distinction between uses of the word ‘should’, to psychological factors like a sense of compulsion attaching to the thought that something is ‘not done’ (1978b 177; cf. 1978a 163), and perhaps to sociological factors (1978a 162). These resources are significant, but they are not entirely satisfying.

In the next few sections I outline a popular view of rule-based reasons which is similar to Foot’s insofar as it treats rule-based reasons as deriving from the agent’s ends. This view is geared towards providing a more satisfying answer to the first question facing Foot’s view, and I will suggest it also provides an interesting way of answering the second. However, I argue that this view is unsatisfactory. This motivates my attempt in the next chapter to treat requirement at face value, as reflecting an interesting phenomenon in the structure of practical reasons.

90 It is worth remembering that Foot, at least in (1978a), was more willing than many philosophers to question widely-accepted *substantive* claims about what reasons we have. Having said that, I do not intend my claims about the *structure* of practical reasoning to depend on a comfortable acceptance of the status quo. My suggestion that our sense of requirement is rooted in the structure of *reasons* is supposed to hold only for the case of *valid* rules. I believe that this can be largely separated from the question of when rules are valid, and I favour an answer there which allows for significant criticism of existing practices.

The Priority Thesis

My claim above about our ‘everyday’ conception of rules is supposed to be neutral between the rules associated with morally-loaded institutions like promising, property, and political obligation, and those associated with what I call ‘ordinary practices’. By ‘ordinary practices’ I mean games, clubs, schools, workplaces and other such institutions.⁹¹

It may be trivially true that one has reason to follow the rules of a game only if one has reason to play it. But we regularly find the expression of a stronger claim than this in the literature. The stronger claim is nicely captured in a phrase used by Matthias Haase. Considering the relationship between games and the practices associated with justice, he suggests that it is a matter of definition that in playing a game we act on an ‘ulterior motive’, whilst by contrast, acting on an ulterior motive is completely alien to the ‘virtue of justice’ (Haase 144). The nature of these ulterior motives is left unspecified, but I take it that this is part of the point: there is nothing very special about the motives which induce us to play a game, compared to the motives which bring us to any other ordinary activity (as opposed, say, to those actions which manifest justice).

Mary Midgley makes a similar point to Haase’s, albeit in different terms. As she puts it, a book of rules for a game “will not contain rules for stopping and starting playing” (236; cf. Haase 144).⁹² Millikan makes the same claim at a number of points

91 I mention only secular institutions and practices here, partly because they are what I know best, and partly because I suspect that many religious institutions are more complex in important ways. However, it is worth noting here that anthropologists Caroline Humphrey and James Laidlaw, drawing partly on Searle’s work on ‘constitutive rules’, have offered an account of a broad swathe of religious rituals which in many respects resembles the ‘instrumental construal’ of practices that I describe in the next section. (See Humphrey and Laidlaw 1994 esp. Ch.4-5.)

92 This is certainly true of *many* games. It does seem possible that a game might have rules about when it can or should be played. (It is interesting, however, that it is not immediately obvious exactly what the status of such rules would be.)

(e.g. 2005a 14-15). In Haase's terms, we play from an 'ulterior motive', although this motive need not be a generic one: "The Chess Player's desire is not a desire for general abstract intellectual activity, curbed and frustrated by a particular set of rules. It is a desire for a particular kind of intellectual activity, whose channel is the rules of chess" (Midgley 243). This point is a helpful corrective to certain initially-tempting glosses on 'ulterior motives'. However, although the *content* of the chess player's desire is unusual, it is still just one desire amongst others, for all that's been said.

For the sake of having a label, I will call the reasons one might have in a particular situation due to the bearing of a rule on that situation, 'rule-based reasons'. Both Haase and Midgley paint the picture that rule-based reasons in the context of games *depend* in some significant sense on the agent's prior reasons for playing the game. More broadly, they suggest that rule-based reasons deriving from ordinary practices will depend on prior reasons for 'engaging in the practice', whilst allowing that things might be different in the case of *some* practices, such as promising. Call this the 'Priority Thesis'. The Thesis says roughly that rule-based reasons generally depend on prior reasons for playing the game or engaging in the practice to which the rules belong. This Thesis is also suggested at one point or another by Anscombe, Foot, Rawls, Searle, and Enoch.

The instrumental construal of the Priority Thesis

The Priority Thesis by itself is compatible with the idea that rule-based reasons are associated with a special sense of requirement. However, the Priority Thesis is sometimes elaborated in a way which seems to be in tension with that idea. For the sake of a label, I will call this way of elaborating the Priority Thesis the 'instrumental construal' of the Thesis. Roughly, the instrumental construal says that the relationship between one's reasons for engaging in a practice, and the rule-based reasons one confronts as a participant in that practice, corresponds to a means-end relationship of the sort familiar from ordinary instrumental action.

The following two quotations help to bring out some of what is involved in this construal:⁹³

If you have no reason to be playing chess, then that some aim is constitutive of playing chess gives you no reason at all, it seems to me, to pursue it, and this whether or not you are in fact playing chess. (Analogously: If you have no reason to be building a house, then that some standard is constitutive of being a house gives you no reason at all to measure up to it, and this whether or not you are in fact building a house.) (Enoch 186)

If one wants to do an action which a certain practice specifies then there is no *way to do it* except to follow the rules which define it. [...] *If* one wants to perform an action specified by a practice, the only legitimate question concerns the nature of the practice itself (“How do I go about making a will?”). (Rawls 26; my emphases)

Suggestions for an instrumental model of rule-following typically invoke something like the category of ‘constitutive rules’. This term was popularised by Searle (1969; 1995), but I will not dwell on his use of it. The basic idea is that the existence of a constitutive rule allows for certain kinds of non-trivial redescription of particular entities – paradigmatically, actions – which would not be available in the absence of the rule. Yet it does not do this by adding any new consequence or obvious intrinsic quality to the action being redescribed.

93 I will anticipate a couple of likely complaints about the use of these quotations. Enoch is explicitly talking about the ‘constitutive *aim*’ of chess, rather than its rules, but in context I see no reason why he would not say something basically similar about the rules, if there is a sharp distinction between the constitutive aim of a game and its rules. Rawls is not specifically talking about the rules of games here, but it is clear from context that his remark applies to the rules of games like baseball, which he takes as a central example in his paper. There are tensions *within* Rawls’s paper between this instrumental construal and a non-instrumental construal of rules. I discuss these tensions in the next chapter. However, I take it that this passage, amongst others, expresses the instrumental construal, even if Rawls’s support for it is ambivalent.

So understood, constitutive rules need have no immediate normative or practical import.⁹⁴ But they can interact with an agent's ends in productive ways. In general it seems plausible that the redescription in question will not be applicable unless the entity redescribed fits certain conditions specified by the rule.⁹⁵ (This thought is clearly doing some work in the quote from Rawls just above.) A toy example, suppressing a number of complexities, might be this: in cricket, hitting the ball over the boundary line without it bouncing counts as hitting a six. This means that *if* one aims to hit a six, one must hit the ball over the boundary line without it bouncing. Hitting a six seems like an intelligible end. So, if one has that end, one has a reason to 'follow the rules' and hit the ball over the boundary, which reason is derived from

94 The claim that the existence of practices or institutions can generate non-trivial redescrptions of particulars is relatively uncontroversial, and I do not wish to dispute it. Cashing this out in terms of 'constitutive rules', however, is not mandatory. Assuming that we do find use for the concept of constitutive rules, certain questions still remain. Aside from considerations of the *contents* of the rules (mentioned in the next footnote), there is room for debate as to whether constitutive rules should be understood as having normative import in and of themselves. Given the typical motivation of the instrumental construal, it is clearly an advantage if they do not (if they are 'merely constitutive', as we might put it), but rather gain their practical role only by engaging with agents' ends. Although there may be room for a conception of constitutive rules as having intrinsic normative significance, I will in the body of the text use 'constitutive rules' to refer to what I have just called 'merely constitutive rules'.

95 For Searle, and also for Thomasson (2003), the conditions an entity must fulfil in order to be a Y – where Y is a kind which depends on the existence of certain constitutive rules – are, at least in the central case, specified by the rules themselves. This may not be the best way to understand the relationship between the contents of a constitutive rule and the conditions for being a member of a rule-dependent kind (cf. Cherry 1973). Social ontology may still find use for a conception of constitutive rules on which this relationship is more complicated, but such a conception would undermine some of the intuitive plausibility of the instrumental model of rule-following, as I describe it in the body of the text. (This model gains some of its appeal from the thought that one has to do "what the rule says" in order to perform an action of a rule-dependent kind.)

that end. In this way constitutive rules, in conjunction with an agent's ends, might generate reasons for her to perform the actions specified by the rules.

Someone might object that instrumental reasoning typically involves *causal* reasoning about how to bring about some particular end. However, what is sometimes called 'constitutive reasoning' – for instance, "I want to exercise; walking counts as exercise; so, I'll go for a walk" – is typically treated as being not fundamentally different from causal reasoning (see Korsgaard 1986 20), and both are sometimes grouped together as forms of 'instrumental' reasoning. There are clear parallels between my descriptions of the cricket case and the exercise case.

Given a reason to perform an action of a type defined by the rules of a game or other ordinary practice, we have seen how one might derive a reason to follow those rules as a means to that wider end. This is the basic point of the instrumental construal. However, not all rule-based reasons obviously fit into this pattern. The rules of chess tell you that *if* you are going to move your bishop, you must do so diagonally, but they can also tell you to move your king because it is in check. Similarly, in a game of *Uno*, "[t]he moment a player has just one card they must yell "UNO!" (https://www.unorules.com). To take a somewhat different example, one might think that simply *as* a teacher one has certain duties towards one's pupils. In each case there is plausibly a rule-based reason, but we have not mentioned any broader end or reason from which this can be derived instrumentally. We might call these '*positive* rule-based reasons'.⁹⁶ However, it is quite common to reinterpret apparent examples of positive rule-based reasons so that they fit the basic 'instrumental' model which we applied in the cricket case.

The chess example has been particularly influential. When a player's king is in check, and there is a way to move it out of check, only a move which takes it out of check

96 In *The Concept of Law*, H. L. A. Hart draws a distinction among laws which he terms 'power-conferring' and those which impose 'duties' (35), and suggests that a similar distinction can be drawn between kinds of rules in non-legal contexts as well (9). This distinction has some resemblance to that I am drawing between the cricket case and the Uno case, although he builds more into his distinction than I find useful for present purposes.

will count as a ‘legal’ move by the player. In such a situation, only moving one’s king to such a square will count as ‘making a move in the game’ at all, according to the rules. So, in this situation, if one has a prior reason to *make a move in the game*, one will have a reason to *move the king out of check* (and from there, perhaps, to *move this token to square b2*). Furthermore, the connection between these two reasons looks, for all that’s been said, like the connection between reasons for *bitting a six* and for *standing here and hitting this ball in this way with this bat*.

This approach explains the apparently positive reason to *move the king out of check* in terms of a broader reason. This might seem to defer the problem, leaving the reason to *make a move in the game* as an unexplained positive reason. However, in the context of a game like chess, *making a move in the game* appears to be what *playing the game* consists in. Perhaps, then, making moves in the game is necessary to count as playing at all. In that case, if one has a reason to play the game, one will have reason to do those things that in the situation count as making a move. At this point, our account seems to link up with, and gain support from, the commonly-held Priority Thesis: the reason a player has for playing the game can unproblematically figure as an exogenous factor, because the Thesis holds that one only faces rule-based reasons if one has a prior reason for engaging in the practice to which the rule belongs. Then, all rule-based reasons (including those which initially seemed to be ‘positive’) can be derived from this over-arching reason via something like ‘constitutive means-end reasoning’. That, in brief, is the programme for an ‘instrumental construal’ of the Priority Thesis as applied to games and other ordinary practices.

This provides some answers to the first question we asked of Foot’s view. Before raising problems for the instrumental construal, I will briefly sketch how a defender of this construal *might* answer the second question, i.e. explain the sense of requirement attaching to rules.

If rule-based reasons are derived instrumentally, the sense of requirement will not be explained by the structure of reasons *per se*. However, the constitutive rules of a practice are relatively ‘stipulative’, meaning that the possible means of achieving some practice-defined action can be limited fairly arbitrarily. Sometimes a practice-defined action – say, *scoring a goal* in football – closely resembles and might on

occasion include an ordinary action – say, *getting this ball through these posts*. However, in the context of a game of football, the rules restrict the possible ways that one can achieve the former action, which do not apply to the latter just as such. If one is trying to get the ball through the posts with a view to *scoring a goal*, one is more limited in the means one can adopt to try to achieve it, compared to the means one could adopt if one was trying to *get the ball through the posts* for its own sake.

Considerations of efficiency, or of conserving one's energy, might on occasion suggest picking the ball up with one's hands and throwing it between the posts. However, given the rules of football, doing this will not count as *scoring a goal*. If one thought of one's end as being *getting the ball through the posts*, it might appear as though the rules were therefore excluding considerations of ends like *conserving one's energy*. If the latter end *were* excluded, then this would fit the account of requirement I have outlined. However, on the instrumental construal the footballer's end should not really be *getting the ball through the posts* in the first place; it should rather be *scoring a goal*. Her end of *conserving energy* does remain relevant to her pursuit of this end – it makes perfect sense for her to consider the most efficient ways to score a goal. Picking the ball up with her hands is excluded simply because it is excluded as a *means* by her true end of *scoring a goal*, not because considerations of efficiency have been excluded.

As I have suggested, though, we can understand why the resemblance between the two ends of *scoring a goal* and *getting the ball through the posts* might lead an agent to feel as though her end of conserving energy was being excluded from consideration by the rules. This is the explanation of the sense of requirement that I am suggesting a defender of the instrumental construal might adopt. It ends up being revisionary, if my description of the sense of requirement above was correct: according to the instrumental view the agent is ultimately making a *mistake* if she thinks some of her ends have been excluded. Nevertheless, it gives the defender of the instrumental construal something to say in explanation of the appearances that goes beyond Foot's talk of a 'sense of compulsion', yet without ultimately invoking a structural distinction in the realm of reasons.

Problems for the instrumental construal

Although it seems to work very neatly for certain paradigm cases, such as moving one's king out of check, the instrumental construal faces a number of problems. I will provide an outline of these problems here. There are a number of ways in which the instrumental construal could be modified to try to escape these problems, some of which I will mention. I will not try to show in detail that each of these modifications fails. However, I believe that each of them is quite unappealing, if we have an alternative to the instrumental construal. I develop what I take to be such an alternative in the next chapter. I will therefore focus in this section primarily on objections to the central and most appealing version of the instrumental construal.

Firstly, the instrumental strategy for explaining 'positive' rule-based reasons can lead to some odd-sounding glosses on the content of the rules of particular practices. Searle's suggestion that "it is a matter of rule of [sic.] competitive games that each side is committed to trying to win" (1969 34 n.1), appears to be an example.

Secondly, it is sometimes simply *incorrect* to say that something one 'must' do in the context of a game is *necessary* to continuing to play. Take the rule that "[t]he moment a player has just one card they must yell "UNO!"." It is not true that if one *fails* to say "Uno!", one will not count as playing the game anymore: play continues, and one will not be made to re-take one's turn, or have one's subsequent victory invalidated. Nevertheless, one must say "Uno!" True, there are penalties if one is *caught* not saying "Uno!", but the 'must' here is not the merely prudential or hypothetical 'must' of "must... if one is to avoid sanctions." (In addition, 'catching' someone is a move defined within the game; not just any observation that someone didn't say "Uno" will count as catching them out.) There is a difference, here, between calling "Uno!" because it is the rule, and calling "Uno!" in order to avoid sanctions.

Thirdly, it does not in general seem to be the case that if one is cheating at a game then one is no longer playing that game.⁹⁷ If it is generally always possible to *either* perform the action required by the game, *or* break the rules and cheat, whilst still counting as 'playing the game', this is a problem for the instrumental construal. The

⁹⁷ This point is highlighted by Thompson, amongst others (2008 178-179 n. 16).

problem is not that the possibility of cheating means that having a reason to play can never generate *any* reason to follow the rules. All else being equal, having a reason to pursue some end can generate a reason to carry out any means which will promote that end, and following the rules still counts as *a* means to promote the end of playing the game. Rather, the problem is to see how there could be any sense of *requirement* to do what's specified by the rules, when one has equal reason to cheat.

A defender of the instrumental construal could try to avoid the first two problems by modifying our understanding of the rules of particular games. The third might be avoided by modifying our understanding of the over-arching reason involved – perhaps instead of *playing the game*, our end should be *playing the game according to the rules*.⁹⁸ (An alternative is always to revise our estimation of the reasons we actually face in the context of ordinary practices, as Foot did.) However, such moves can seem *ad hoc*, and the second move in particular seriously undermines the support the instrumental construal can derive from the appealing Priority Thesis. I suggest that in light of the commonalities between the three problems, if we want to do justice to the sense that rules are sources of *requirement*, we would be better off acknowledging a richer structure of practical reasons than can be found in the instrumental construal, thereby allowing us to take at face value our ordinary conception of rules.

Each of these three problems stems from the point that, as we saw above, the only kind of ‘necessity’ which is internal to ‘instrumental reason’ *itself* is the kind that stems from there only being one eligible option for pursuing the end in question.⁹⁹

98 Anscombe writes of the temptation she felt at one time to construe people as acting on the end of “playing a game according to its rules”, in order to preserve the attempt to construe “all deliberate action as a matter of acting on a calculation how to obtain one’s ends” (1981b, viii). See also Millikan’s remarks that “you must follow the rules of chess if you wish to play a *conventional* game of chess; otherwise you can do what you like with the pieces” (2005d, 148; my emphasis).

99 This is, I believe, one of the least controversial claims one can make about ‘instrumental reason(s)’. Raz is highly sceptical of most traditional ways of understanding instrumental reason. However, even he would acknowledge this point with respect to the reasons which he thinks most closely reflect those traditional understandings, namely what he calls ‘facilitative reasons’ (2005 5-6).

What our examples show is that we have a sense of practical necessity associated with the rules of games even in cases where there is *not* only one eligible option for pursuing the end of playing the game. This despite the stipulative nature of many such rules, noted in the last section.

The deeper problem with the instrumental construal is that it treats reasons for engaging in a practice as just one end among others. This means that there is no general reason why one could not consider one's other ends both when deciding whether to pursue the end (engage in the practice) and also when deciding whether to pursue the means to it (act in accordance with the rules).¹⁰⁰ If that is correct then we are unable to explain what is infelicitous about keeping a promise because it is 'best on the whole', or about taking a required logic class because of the merits of studying logic.

Conclusion

Rules are associated with a sense of requirement. This shows up in the practical thought of agents who fall under rules, who tend to feel that their deliberation is more constrained than it is when they are guided by pursuit of their ends. It also shows up from a third-person perspective. For instance, we saw that we can explain an agent's acting contrary to their ends by pointing out that they are following a rule. We also saw, in passages from Rawls and Korsgaard, that the phenomenon of requirement shows up in the sense that there is often something inappropriate about deciding what to do by weighing up all of one's ends, in a situation to which a valid rule applies.

On the face of it, then, the sense of requirement seems to be explained by at least some ends being excluded from practical consideration by the claim of a valid rule. I outlined an attempt – which I called the 'instrumental construal' – to explain the sense of requirement attaching to rules without claiming that rules exclude any of

¹⁰⁰ I take this, too, to be one of the least controversial claims one can make about instrumental reason(s), though I do not deny that some might doubt it.

the agent's ends. On the instrumental construal, rule-based reasons are derived instrumentally from one of the agent's ends, specifically the end of participating in the practice to which the rule belongs. The instrumental construal typically invokes a minimal social ontology of so-called 'constitutive rules' to explain how this broader end could generate specific rule-based reasons for action.

I argued that the instrumental construal of rule-based reasons looks promising for certain cases, and has some resources for giving an account (albeit a revisionary one) of the sense of requirement attaching to rules. However, in the final section I argued that the instrumental construal ultimately faces a number of problems in accounting for rule-based reasons as we generally understand them. It implies implausible claims about the rules of particular familiar practices, such as the game of Uno. It also faces the quite general problem that it is consistent with there always being reason to cheat. If there is always reason to cheat, consistent with the end of participating in the practice, then the instrumental model cannot explain the sense of requirement attaching to the rules of games. The reasoning here depended on quite general features of the examples, though, so the same will apply to the rules of any other practice on this construal.

In the next chapter I will propose a different way of accounting for the practical role of rules. This will draw on the ontology of social practices I developed in Chapters 1-3 to provide a more satisfactory explanation of the sense of requirement.

Chapter 7. ‘Generality’ and the Rules of Ordinary Practices

In this chapter I will provide a positive account of the practical role of the rules of ordinary practices, which explains the sense of requirement attached to them. I will do so in a way that draws from the ontology of rules, norms, and conventions developed in the first three chapters. I develop this positive account through a discussion of Michael Thompson’s work on what he has called ‘practical generality’. Thompson, too, highlights the metaphysical generality of rules, and shows how this explains some of their practical features – in particular, their ability to hold in a ‘tight corner’. I suggest that we can build on this insight, showing how the generality of rules also helps to explain the sense of requirement attaching to them.

There are some apparent obstacles to this account, however. I mentioned at the beginning of Chapter 6 that the association of rules with requirement is often thought to imply that it can never be right to break valid rules. This view frequently leads to the conclusion that the rules of ordinary practices cannot really be sources of requirement, because it is surely often right to break such rules. This issue takes on a particular form in the context of the present account. It leads Thompson to restrict his view of rules to cover only the rules of rarefied practices such as that of promising. This restriction is at odds with my own aim, which was to provide an account of the rules of all sorts of ordinary practices.

Defending my account will therefore require resisting the suggestion that if rules are sources of requirement then it can never be right to break them. I will argue that, in fact, Thompson’s arguments do not give us a clear reason to accept this conditional in the first place. The positive account I develop in this chapter will only cover part of the practical role of rules. However, it will also clear the ground for a fuller

treatment in the future, by defending it from this significant objection which can otherwise make my approach seem doomed from the start.

An overview of the argument

Thompson argues that in order to do justice to our ordinary moral thought about the reasons agents have to keep their promises, we must mark a significant distinction among forms of explanation and justification of actions.¹⁰¹ Specifically, he claims that particular actions of promise-keeping acquire their goodness or rationality by manifesting the ‘practice’ of promising, which is itself good, and which moreover is ‘general’ and ‘transparent’ (in technical senses to be described below). This contrasts with those particular actions that are explained or justified by being fitted into a wider pattern of over-arching particular actions and events. (Such actions are associated with ‘instrumental’ explanation.)

I will argue that Thompson’s understanding of ‘generality’ helps to shed some light on the phenomenon of requirement. However, in his account it seems as though generality and transparency must go together. But, as I will show, transparency is not a feature that ordinary practices can possess. Even more troubling, Thompson’s argument seems to introduce a *dichotomy* between actions which are explained and

101 My discussion of Thompson’s work will be circumscribed. I focus on the central line of argument running through Chapters 9 and 10 of *Life and Action*, because this forms a relatively coherent whole, and contains the material most directly relevant to my broader purpose. I make some reference to earlier parts of the book, but only where it is needed to understand the later argument. A full treatment of Thompson’s view on promising (the main case discussed below) would need to provide some discussion of ‘bipolar normativity’ as it figures in his paper ‘What is it to Wrong Someone?’ (2004). I have not provided this partly because there is insufficient space to do so, but also because whilst bipolarity is a central part of *justice* on Thompson’s view, the category of ‘practical generality’ which he discusses in *Life and Action* seems to attach in the first instance to *virtue*, and should equally cover those virtues which do not exhibit bipolarity, such as temperance. Given that my main concern here is with a feature of the wider category, nothing crucial should be lost by omitting bipolarity.

justified through a transparent practice, and those which are explained and justified instrumentally. This suggests, given my argument against the instrumental construal in Chapter 6, that there will be no satisfactory place for actions explained and justified by the rules of ordinary practices. I suggest that an assumption I will call ‘rigourism’ might account for the appearance of a dichotomy in Thompson’s work, and that it is plausibly at work more broadly in the literature. The upshot is that rejecting rigourism might allow us to give a satisfactory account of the practical role of the rules of ordinary practices.

Thompson’s account of ‘practical generality’ develops via what he calls an ‘immanent criticism’ of the concepts of ‘disposition’ and ‘practice’, as they figure in the work of Gauthier, Foot, and Rawls.¹⁰² Before beginning the discussion of his account, three clarifications are appropriate.

First, both Rawls and Foot appear in a number of guises in this thesis. Foot’s views in the papers Thompson cites are different from those she held in ‘Morality as a System of Hypothetical Imperatives’, discussed above, and for present purposes the Foot of this chapter can be treated as a distinct figure from that of Chapter 6. More complex is Rawls’ role. I have already invoked him as a source of the idea of ‘requirement’, and as a proponent of what I called the ‘instrumental construal’. Thompson uses Rawls as a source of the idea of ‘practical generality’, which stands in contrast with the instrumental construal. Most confusing, all of these points are drawn from the same paper of Rawls! The situation, as I see it, is this. Rawls developed his account of practices partly in response to the point about requirement. However, he was somewhat ambivalent about following through its consequences, leading to the elements of instrumentalism which I highlighted in the previous part of this paper.

Second, I will largely refrain from commenting on the differences between Rawls’, Foot’s, and Gauthier’s views, and any divergences I see between their real views and those ascribed to them by Thompson. Finally, I focus on the discussion of practices,

102 Each of these writers has a large and varied body of work. The canonical texts for Thompson’s purposes are Gauthier’s *Morals by Agreement*, Rawls’s ‘Two Concepts of Rules’, and the last eight paragraphs of Foot’s ‘Moral Beliefs’ (Thompson 2008 150).

especially the role of practices in Rawls's account of promising. However, it should be borne in mind that in discussing practices Thompson is trying to characterise the broader genus to which *both* 'practice' and 'disposition' belong (along, perhaps, with notions like 'principle' and 'practical identity' (see 2008 150)).¹⁰³

Thompson's argument from tight corners

The phenomenon which sets the scene for Thompson's discussion of practical generality is what he refers to as the 'tight corners' that emerge in the context of promising.¹⁰⁴ Ordinary moral thought seems to hold that one ought to keep one's promises, simply because one has made a promise, even in cases where the thing one promised to do turns out not to be good or useful, considered in itself. A classic example is the secret deathbed promise to do something costly and trivial. Such cases puzzled Hume, and as Rawls notes they have often been thought to pose a problem for Utilitarians.

103 This is consistent with what I said above, namely that I do not claim to be describing features which are *unique* to rules. Thompson's grouping together of practices, dispositions, identities, etc. might be compared with Anscombe's grouping of rules, rights, and promises.

104 Thompson's language echoes, consciously or otherwise, Anscombe's remark that "[i]t is the tight corner that presents the theoretical difficulty" in her account of promising (1981c 20). However, Anscombe's problem and Thompson's are interestingly different. For Anscombe, the problem of how one could be acting well in keeping a promise, even when no good seems to attach to that particular action itself, has already been solved. In her account the problem arises for the agent who asks why it is necessary for her to act well. It may be true that it generally pays to act well, but there appear to be 'tight corners' where this generalisation does not hold (unless the agent has some aim which cannot be pursued by acting badly). For Thompson, the question of the tight corner is precisely the one which Anscombe is here treating as solved: How could being part of the practice of promising confer any value onto an act of promise-keeping which is, considered just in itself, worthless? As I understand it, Anscombe's later papers on promising develop an account of the puzzle they pose which is closer to Thompson's (e.g. Anscombe 1981f; 1981a).

Thompson calls cases in which we think one ought to keep one's promise, in spite of the fact that one could do more good by breaking it in this particular situation, 'tight-corner cases'. It is important to emphasise that the tight corner is a *structural* phenomenon – it does not depend on a commitment to a specific view of the *content* of moral reasoning, like Utilitarianism. So, for instance,

For Gauthier and for Foot, the problem is the more ancient one that, like any requirement of justice, promise-keeping will sometimes cross our *interests*, even if we take into account the irritating noise of complaint, the threat of never being trusted again, and the like. (2008, 152)

Here we see the same kind of structure with a different content. Putting it quite generally, "the goods and evils to be pursued and avoided in the tight-corner context are such as would make the faithful person's conduct either morally blameworthy or imprudent if it were performed by a similarly situated person, but one who had not made a promise" (153). This gives rise to a puzzle: how can we explain our sense that one ought nevertheless to keep one's promise in such a case?

For Thompson, as for Foot, Gauthier, and Rawls, the tight corner is not something to be *explained away*. We should embrace it, rather than attempting to show for each apparent tight corner, either that there is some good to be found which will tip the balance, or else that it should be included in the set of cases in which it is permissible to break one's promise. (Almost everyone acknowledges that there are *some* cases where it is permissible to break a promise, when the consequences of keeping it would be sufficiently dire.) Allowing the existence of tight corner cases seems to be required by our ordinary moral thought. That means allowing that "the faithful person is sometimes left doing what would simply be stupid or wrong for a similarly situated person who had not given her word", but that *she*, having given her word, is justified in doing it (153, n.5).

So far this characterisation of the tight corner has been given in terms of ordinary moral thought, but Thompson gives it a more *metaphysical* spin which, whilst less intuitive, is still recognisable:

On neither account is the individual act of fidelity held to acquire the relevant characteristic, *moral goodness* or *rationality*, from the agent's setting his sights on some benefit or excellence or merit that might reside in the *individual* action itself or in the series of *individual* events it is likely to entrain. [...] On neither account does the faithful agent anticipate the appearance of anything in the order of individual events that might outweigh the more readily recognizable benefits of tight-corner infidelity. (153; my emphasis)

This metaphysical gloss on the tight corner gives a specific shape to the puzzle we face in making sense of it. To see what this shape is, we need to connect this with Thompson's account of 'Naive Action Theory' from Part Two of *Life and Action*. In that account – broadly identified as an account of 'instrumental' action-explanation – it is precisely something 'in the order of individual events' that helps to explain action.

The paradigmatic form of naive action-explanation explains an action of A-ing by fitting it into some broader or over-arching action of B-ing, where the agent is knowingly engaged in B-ing, and thinks that A-ing will in some way contribute towards her doing B (93-94). I'm *breaking some eggs* because I'm *making an omelette*. The role that this over-arching action of B-ing plays in the explanation of A-ing is the role of what Thompson calls a "nexus of action and consideration", or 'nexus of thought and action' (161). Instrumental action in general is characterised by a particular "*form* of dependence of [...] action on [...] consideration", which depends on fitting one's action into a broader pattern of action in this way (95; my emphasis).

Thompson does not really spell out what a 'nexus of action and consideration' *is*, but I think we find a useful parallel in Anscombe's account of practical reasoning in *Intention*. Anscombe suggests that any *genuinely practical* reasoning (i.e. reasoning which may lead to action, as opposed to the kind of idle syllogising one might perform as a classroom exercise) will involve a deep division of labour between the material contained in the premises of the practical syllogism, on the one hand, and something

else which is not part of the syllogism, and which Anscombe calls a ‘wanting’ or ‘want’, on the other (1957, §35).¹⁰⁵

This division of labour is displayed by a dramatic example from a later paper. Take the premises,

Strong alkaloids are deadly poison to humans.

Nicotine is a strong alkaloid.

What’s in this bottle is nicotine.

Very different actions might follow from consideration of these, depending on whether the agent wishes to live or to die (Anscombe 2005 116). More to the present point, *no action whatsoever* might follow from my contemplating the fact that they have good Jersey cows at Hereford market, if I do not want a Jersey cow. The division of labour between the premises and the wanting means that a perfectly good set of premises can be expected to fall flat – to result in no action – if they do not engage with anything I want. My suggestion is that in Thompson the over-arching action of B-ing takes the place occupied by a wanting in Anscombe’s account of practical reasoning. ‘Wanting’, in Anscombe’s account, is a ‘nexus of consideration and action’, linking considerations like “They have good Jerseys in Hereford”, with actions like going to Hereford.

But, as I have suggested, the nexuses of consideration and action that figure in naive action theory – paradigmatically, an over-arching action – are precisely things that are

105 In one sense, that there is a deep division in the explanation of action between the role of a ‘want’ or ‘desire’ and that of some ‘cognitive’ element concerning how the thing desired could be obtained, is about the least controversial claim one could find in the philosophy of action. It can be found, for instance, in Davidson (2001a) and Smith (1987). I take it, however, that the role Anscombe envisages for the ‘want’ in practical reasoning is significantly different from that found in Davidson or Smith. It is no easy matter to precisely characterise this difference, but I hope that the discussion of Anscombe in the body of the text provides some indication of the distinctive features of her account, which make sense of calling the want a ‘nexus’ of thought and action.

‘in the order of individual events’.¹⁰⁶ Even when we add the resources of what Thompson calls ‘sophisticated’ action-explanation, which allow us to explain A-ing by reference not to a wider action of B-ing that is currently in progress, but to a want or intention to do B, this does not get us away from the order of individual events. The message of the last section of Chapter 8 of *Life and Action*, as I understand it, is that *wanting to B*, and *intending to B*, are no less processes than is *doing B* itself (Thompson 2008 138-146).

Now we are in a position to appreciate the problem raised by the tight corner in a new light. If the Anscombean division of labour requires something to play the role of a nexus of thought and action, but nothing in the order of individual events can play this role in the tight corner, then what *can* play it? A crucial ingredient of practical reasoning and action-explanation seems to be missing, in the tight corner.

As I see it, this puzzle is what leads Thompson to introduce practices and dispositions as entities which are ‘general’, and which can play the role of a nexus of thought and action. Treating a practice as something general allows us to think of the relation between, say, the practice of promising and a particular action of promise-keeping as one of *manifestation*. This in turn allows us to think of the practice as being manifested in indefinitely many distinct actions of promise-keeping. As Thompson puts it, the practice of promising, “is essentially one and the same, unchanged, unexhausted, and not merely similar, through a potentially unlimited series of individual acts of fidelity” (2008 158). To think otherwise would be to risk making the justification of tight-corner promise-keeping hang on its tendency to *strengthen* or *maintain* the practice itself, collapsing into an attempt to provide an “especially sophisticated analysis of the series of events into which the individual act of fidelity is fitted”, and thus denying the existence of genuine tight corners altogether (157). Some accounts of tight-corner promise-keeping might attempt that

106 My *walking to school* can explain my *crossing the road* even in the case where I never get to school and so there is never a “perfected” event considered as a ‘genuine particular’ which could be reported by “I walked to school” (Thompson 2008 137). This does not, I think, affect the point that the “imperfectly present process” of *walking to school* should be considered as something ‘in the order of individual events’, for present purposes (133).

strategy – Hume is sometimes, I think wrongly, read this way. However, we saw reasons in the previous chapter to think that that despite its initial appeal, such an approach will not be satisfactory.

The question now is how ‘manifesting the practice of promising’ could help make a particular action of tight-corner promise-keeping *morally good* or *rational*. The answer is supposed to come from the goodness of the practice itself. This is part of the thought behind Rawls’s insistence that we distinguish between the justification of a practice, and the justification of an action which falls under that practice. One justifies the practice itself by, say, referring to the utilitarian principle. But “[o]ne doesn’t so much justify one’s particular action as explain, or show, that it is in accordance with the practice” (Rawls, 27). Providing that the practice is a good one, the actions which manifest it will be good, too.

What Thompson brings out is that Rawls’s picture seems to presuppose that practices can be treated as ‘transparent’. It requires us to endorse a ‘transparency’ or ‘transfer’ principle to the effect that “a good practice makes the actions falling under it good” (Thompson 168; italics removed). This is how the problem of the tight corner is solved: the goodness of a particular action of tight-corner promise-keeping is derived, not from anything in the order of individual events, but rather from the goodness of something ‘general’, namely, the practice which the action manifests.¹⁰⁷ The practice, or the rule, can serve as the relevant nexus of thought and action. Much of the rest of Part Three of Thompson’s book is devoted to making sense of this idea of transparency as a feature of practices.

Our sense that promises can bind even in a tight corner – that is, even where there is nothing ‘in the order of individual events’ which justifies it – is thus the key premise in an argument that the action of the virtuous promise-keeper cannot be understood

107 Taking this position seems to entail that we can’t account for the ‘generality’ of practices via a construction from individual acts of promise-keeping and the like. In the account I developed in Chapters 1-3 of the present thesis, I allowed that rules, norms, and conventions might be generated in part through individual actions. However, the rules (etc.) themselves are something over and above a collection of particular actions. I take it, then, that according to my account they are ‘general’ in the present sense.

in terms of her deliberately pursuing some of her ends. To explain such actions of promise-keeping, we need to appeal to something *outside* the order of individual events (hence ‘general’ in Thompson’s sense). I suggest that we can generalise this argument to cover rules other than the rule that ‘promises must be kept’: rules are ‘general’ entities which, when they are themselves good, can stand as sources of practical reasons distinct from an agent’s ends.

Generality and requirement

According to Thompson, the virtuous agent, acting on the consideration “I promised Y I would do it”, exhibits a different *form* of practical reasoning, compared to the agent who breaks a few eggs in order to make an omelette, or who keeps her promise for an ulterior motive (2008 95; see also 95 n.14). I suggest that this difference in form can help to explain the sense of requirement attaching to rules, which I described in Chapter 6.

Thompson’s virtuous promise-keeper is not indifferent to the good – quite the opposite – but the good which she sees in her action is not derived from its participation in ‘the order of individual events’. This fact makes the balance of *particular* goods and ills that could be achieved by her keeping or breaking her promise in a given situation somewhat irrelevant in a way that it is not when her reason for performing an action is anticipation of some particular good that will arise from it.¹⁰⁸ My suggestion is that if an agent is aware that this is the situation – and in the case of explicit rules, we can expect that she will be – then this will help explain the sense of requirement she associates with rules. This may not, by itself, get us all the way to Rawls’s and Korsgaard’s remarks about requirement. However, it does get us some of the way, and it does so by marking a structural distinction within

108 Any developed account of the virtuous agent’s practical thought would have to be much more nuanced than the gloss I have just offered. That the ‘particular goods’ are in some sense ‘irrelevant’ leaves open whether the virtuous agent could still regret not being able to realise them, for instance. Some of the nuances in this topic are brought out in the exchange between John McDowell and David Wiggins in *Aristotle and Moral Realism*. (ed. Robert Heinaman)

the realm of reasons themselves. The further element we need for a full account of requirement will be introduced later, when I discuss the role of ‘complete explanation’.

Thompson’s discussion of Rawls also suggests a more general way of spelling out the key feature of rules: a genuine (and valid) rule should be such that it could, *at least in principle*, be binding in a tight corner. It must have at least that kind of ‘generality’. We can map this point onto Rawls’s ‘two concepts of rules’. Rawls draws a distinction between what he calls a ‘summary’ conception of rules, and a ‘practice’ conception. He suggests that much traditional scepticism about the practical role of rules has come from adoption of the former conception.

A summary rule of the form “Whenever C, do A”, simply registers the fact that we have recognised that for a number of instances in which C holds, it makes sense to do A. “We are pictured as recognizing particular cases prior to there being a rule which covers them” (Rawls 22). Such a rule could not apply in a tight corner, if the agent recognised it to be a tight corner, and this raises the question of whether it could actually *do* any normative work.¹⁰⁹ A summary rule may have various representational or epistemic uses, but it is hard to see it as having any normative reality. (Rawls doubts whether we should call summary rules ‘rules’ at all (23).)

In developing his alternative ‘practice conception’, Rawls makes much of the idea that one simply cannot perform a practice-defined action outside the context of the practice which defines it. We have seen that this is true in the case of some games and other practices, though it is far from clear that it holds for every act-type that a rule of an ordinary practice might require us to perform. The point, instead, is that a rule’s ability in principle to extend to cover a tight corner – which we can think of as its having a certain kind of ‘generality’ – is the way to understand its having normative ‘reality’. This, rather than the impossibility of performing practice-defined actions outside the context of a practice, should be seen as the basic contrast with

109 Raz’s notion of a ‘rule of thumb’ might seem to be a counterexample: a summary rule that can hold in a tight corner (Raz 1999 Ch.2). The relation of Raz’s view to the present question is complex, and in my view his account of rules is ultimately unsuccessful, but to explore this here would take us too far afield.

‘summary’ rules. It is also what helps to explain the sense of requirement attaching to rules.

The account of rules, norms, and conventions I developed in Chapters 1-3 of this thesis, according to which they are abstract individuals, fits well with the present account of how rules bind in tight corners, and thereby become sources of requirement. (I called rules, norms, and conventions, understood as abstract individuals, ‘AI conventions’.)

AI conventions are ‘general’ entities, capable of being multiply-instantiated. But that is not all. We saw that *some* ‘general’ entities may have no priority over their instances, but instead are better considered as something like a ‘fusion’ of a collection of particulars sharing some property. In terms of the preceding discussion, such entities would be analogous to ‘summary rules’, and indeed it is difficult to see how they could *explain* much about their ‘instances’. By contrast, with AI conventions, created through the reproduction of arbitrary patterns of behaviour, the rule itself (i.e. the AI convention) does have reality over and above that of its instantiations. It is capable of explaining the occurrence of its instantiations.

We can make a similar point concerning goodness or rationality. I will argue below that the rules of ordinary practices are not ‘transparent’ in Thompson’s sense. Nevertheless, their reality over and above their instances means that we can see how the *goodness* or *rationality* of a rule might transfer to its instances. This would be a source of goodness for an action of rule-following which would not derive from its participation in the order of individual events. If so, a rule could apply in a tight corner, explaining the goodness of instantiating the rule when performing the act in question independently of the rule might have nothing going for it. In short, on the view of rules as AI conventions there is no need to think of them as merely ‘summary’; they can apply in tight corners and therefore be sources of requirement.

It is worth pausing to note the connection back to the criticism of Bernard Williams’ formal individualism at the end of Chapter 5. We have seen in this section that an agent’s taking a rule to be such as to apply in a tight corner can explain her association of the rule with a sense of requirement. What we have also seen is that if

she takes a rule to apply in a tight corner, then this is evidence that it is *the rule itself*, considered as an AI convention, that is the object of her practical identification. This in turn was the key premise in our argument that, contrary to Williams' assumptions, it can be necessary to appeal to something non-psychological – an AI convention – as part of the proximate explanation of an individual's action.

My account of AI conventions is not the same, metaphysically, as Thompson's account of practices or dispositions. However, the way Thompson highlights the practical significance of 'generality' has helped to frame how generality might play a role in explaining what I have suggested is the sense of requirement attaching to rules. Rules, as AI conventions, are general in the right kind of way to fit with this explanation. I will not be able to provide a full *positive* account of how rules can justify their instances, here. Instead, in the rest of this chapter I will remove an apparent obstacle to my account of the practical role of rules, which arises from the way Thompson sets up his own view. Namely, the obstacle posed by the fact that ordinary practices are not transparent.

Ordinary practices are not transparent

According to Thompson, the practice of promising should be understood as both 'general' and 'transparent'. I have suggested that 'generality' helps us to get a grip on the phenomenon of requirement which is associated with a variety of rules, but transparency cannot be a feature of the rules of ordinary practices.

As Thompson notes, not just anything we might think of calling a 'practice' (or a 'disposition') could make a reasonable claim to be 'transparent' in his sense. We could collect together any old behavioural regularity and call it a 'practice', if we wished – say, the 'practice' of going to the cinema on Saturdays. But if this 'practice' passed some test of utility, it would not thereby gain a role in *explaining* the goodness of particular instances of cinema-going on a Saturday. This much seems independently plausible, but transparency is in fact so strict as to exclude all examples of what I have called 'ordinary practices' from being transparent, and hence from being

practices in Thompson's sense. Ordinary practices include things like games, and the rules of clubs, schools, etc.

The only things which have a plausible claim to be 'practices' or 'dispositions' in *Thompson's* sense, will be those things which we consider as *virtues*. Furthermore, virtues themselves will only count providing we endorse the *unity* of virtue. The reason for this is that in delineating a virtue we can allow ourselves to be guided by the thought that a bad action is, *simply by being a bad action*, not a manifestation of the virtue. The need to be guided in this way emerges directly from the transparency condition: transparency requires that every action that manifests a good practice is itself a good action; but then any good practice which has manifestations we are unwilling to count as good cannot be counted a transparent practice.¹¹⁰

Ordinary human practices, for instance games but also more serious institutions, perhaps even including a just legal system, will not count as 'practices' in Thompson's sense. Our understanding of what fixes the content of the rules of a game (or of a club, or the laws of a legal system) does not allow us to think that a bad action must, *simply by being a bad action*, not be a manifestation of the rules of a game. Thompson secures the transparency of promising in part by allowing the possibility of widespread mistakes about the conditions under which promises must be kept (188), but we cannot endorse the parallel idea that we could be collectively wrong about what the rules of chess, baseball, or a club require. (Or at least: we can't be collectively wrong about it in *this* way.) Our interpretation of ordinary practices cannot generally be carried out under the kind of regulative principle that operates for the interpretation of virtues.

The apparent dichotomy between instrumental and 'general'

Finding this difference between virtues and ordinary practices is unsurprising, in itself. Haase's and Midgley's remarks showed in passing above that it is easy to generate an intuitive sense that games and virtues differ in important ways. However,

110 This argument is implicit in Thompson's discussion of "deviant" practices of promising (2008 183-191).

Thompson's arguments in Part Two and Part Three of *Life and Action* appear to imply that generality and transparency must always go together in the explanation of action by appeal to a practice.

The picture gets worse. Recall that explanation of actions by appeal to something 'in the order of individual events' was identified with instrumental action. Then recall that the characterisation of the tight-corner case was precisely that it could *not* be explained by appeal to something in the order of individual events; that is, it could not be explained instrumentally.¹¹¹ The argument rehearsed above then seemed to be that in order to explain or justify action in the tight corner, it was necessary to appeal to something *general* and *transparent* – a practice or disposition – which could provide the 'nexus of thought and action' in this case.¹¹² But if that argument is correct, then we have instrumental explanation on the one hand, and explanation by appeal to Thompsonian practices or dispositions on the other, and *no room in between the two*.¹¹³ We saw in Chapter 6 that the rules of ordinary practices fit badly with the instrumental construal, and I have just shown that they are not transparent, so we seem to be left with no way to give a satisfying account of such rules.

I am going to argue that if we reject the existence of this dichotomy then we can find a more satisfactory home for ordinary practices. However, there is an option left open by the dichotomy which we have not explored here. This would be to try to accommodate non-transparent ordinary practices by appeal to the kinds of rules or

111 My claim is not that promise-keeping can *never* be explained or justified instrumentally.

One can of course have ulterior motives for keeping a promise, which may well be – perhaps will necessarily be – instrumental. I am not suggesting that Thompson denies this either (see e.g. 2008 95). The point is rather that our intuitive conception of the tight corner already accounts for this: we think that promises should often be kept even when there is no ulterior motive for keeping them.

112 As has been my practice throughout, I largely gloss over an important third element: the *goodness* of the practice.

113 This is not supposed to be a dichotomy within action-explanation *per se*, on his account, but rather within the explanation of distinctively human action, or perhaps within 'rational' action. Other forms of explanation will be available for the actions of non-rational animals, for instance, and perhaps of some of our less reflective actions, too.

principles which *are* transparent. This option might allow us to account for the sense of requirement attached to ordinary rules by piggy-backing off the account of requirement attached to transparent (perhaps moral) rules. Such an approach is perhaps found in Scanlon's discussion of the so-called 'Principle of Established Practices', which I mentioned in the Introduction (Scanlon Ch.8). Any such account would have to give some explanation of the cases where the rules of ordinary practices do *not* bind us. This may not be too difficult. More problematic, in my view, is the fact that this approach requires us to find transparent rules or principles which can back all the genuinely binding ordinary rules we wish to acknowledge. This seems likely to distort our understanding of both putatively transparent, and ordinary rules. In addition, as I mentioned in the Introduction, Scanlon's position seems to me to err by making the role of actual practices too indirect. However, I cannot make the full case here.

The gap between the tight corner and transparency

In this section I begin to unpick the dichotomy in Thompson's framework, which seemed to leave no room for an adequate account of ordinary practices. There is a logical gap between the claim that a practice can apply in a tight corner (and hence be 'general' in Thompson's sense), and the claim that that practice is transparent.

Foot, in 'Morality as a System of Hypothetical Imperatives', appeared to incline towards something like an instrumental construal of the reasons generated by the rules of ordinary human practices. She was moved partly by the thought that unless these reasons are dependent on the ends we have in the particular situation, there would be no way to escape even the silliest (or, we might add, the most oppressive or exploitative) rules that might attach to some institution or other (1978a 168 n.8). We can partially escape this problem by claiming that there are end-independent standards of goodness for rules, and that only a good rule can ever bind. (More generally, the discussion of practical incapacity in Chapter 5 above showed that constraints on an agent's action which are not dependent on her ends need not thereby be entirely 'external' constraints.)

However, this move leaves a problem remaining. Another way of putting the point that ordinary practices are not transparent is to say that it *must* sometimes be right to break the rules even of good ordinary practices. But how can we make sense of this possibility? In his criticism of Rawls, Thompson moves straight from the claim that we should allow room to criticise the action of an agent who continues to circle the bases – or to move her king – during an earthquake, to the claim that “the ‘justification’ of an individual act of [Rawlsian rule-following] is *nowhere* itself bound up with or colored by the necessity of the underlying practice or of the good it serves” (2008, 179; my emphasis). Rawls’ conception of practices includes games like chess and baseball, as well as promising. Thompson’s thought is that if, as in these games, rule-following actions are sometimes wrong or imprudent, then the games are unable to ever directly justify the actions which manifest their rules. But if the practice can *never* justify its manifestations, then it cannot do so in a tight corner, and so we seem to have slid from non-transparency to an instrumental model of the rules of ordinary practices. Hence the dichotomy outlined in the previous section.

But I want to push the question: why *should* we think that there is this choice to be made between an instrumental construal on the one hand, and a transparent practice or disposition on the other? To put the question bluntly, and in the terms of Thompson’s discussion: we brought in the idea of a ‘transparent’ practice in order to show how promise-keeping in the tight corner could be justified, when by definition nothing ‘in the order of individual events’ could justify it, but why should we think that the *only* way for a practice to extend to the tight corner is for it to render *every single* action which manifests it good or rational? There is a logical gap between the following two claims:

1. Practice P, if it is a good practice, renders some actions which manifest it good or rational, which would not be good or rational considered independently of P.
2. Practice P, if it is a good practice, renders *every* action which manifests it good or rational.

Claim (1) is what is minimally required for P to justify some of its manifestations in a tight corner. (2) is an expression of the ‘transparency’ of P. But (1) does not obviously entail (2). The claim that for every practice of which (1) holds, (2) holds also – that transparency *is* necessary to account for the tight corner – requires further argument. The appearance that (1) entails (2) underpins the apparent dichotomy I have been addressing in this part of the chapter, but this entailment does not follow from the nature of the tight corner itself.

A premise which might bridge the gap between (1) and (2) would be something we can call ‘rigourism’. This is the idea that *if* a genuine rule actually applies to a situation, then it completely settles what must be done in that situation.¹¹⁴ As we have seen, this is the status of the virtue of fidelity, on Thompson’s account – if someone is moved to keep her promise in a situation in which it would in fact be better to break it, he says we should not treat her action as the work of the virtue of fidelity.¹¹⁵ And in general when we have the weighty topic of justice before us rigourism can seem unobjectionable. But if my diagnosis of the gap between (1) and (2) is correct, then something like rigourism is responsible for forcing the dichotomy between the two forms of action-explanation I have been discussing. Moreover, one finds traces of rigourism in many other philosophical discussions of rules. However, rigourism is a substantive premise; there is no merely a logical inference from (1) to (2).

114 This view bears a complex relationship to the view that a genuine rule should be *exceptionless*, as that view figures in, for example Dworkin’s ‘The Model of Rules – I’.

115 It might seem that Thompson’s position can’t embody a kind of rigourism about rules, because in the case of promising, for instance, he explicitly denies that the various exceptions that we recognise to the rule that ‘promises must be kept’ are to be written into our practice of promising (2008 Ch. 10, Section 4). Yet he does allow that there are times when one ought not to keep one’s promise. However, the point of denying that we should write exceptions into the practice seems to come not from the thought that sometimes it will be right to go against the practice – that, after all, is ruled out by transparency – but rather from the thought that *any* practice of promising will have to recognise fundamentally the *same* set of excusing conditions. (A view about the ‘uncodifiability’ of the correct principles of promise-keeping may be somewhere in the background here, too.)

My interest in the first part of this thesis was primarily in the ontology of *ordinary* practices, conventions, rules, and norms. We have seen that this ontology can help to shed light on some of the distinctive features of the practical role of such practices for us. However, we have also seen that this account would be seriously limited if it only applied to transparent practices. So, what scope is there for rejecting rigourism?

Rigourism and requirement

Given the picture we have been developing here, it looks as though rigourism is a logically stronger claim than transparency. (This explains why rigourism can be used to support transparency claims for particular rules or practices.) In denying transparency for ordinary rules, then, we are committed to denying rigourism. However, the rejection of rigourism may threaten to undermine the explanation given above of the sense of requirement attaching to rules.

This is because the explanation of the sense of requirement does not rely *solely* on the generality of rules. Rather, it relies on the claim that a rule (considered as something general) is in some cases and in some sense *sufficient* to determine what an agent should do. Another way of putting this might be to say that in those cases it *completely* determines what the agent should do. I will now elaborate on why we need this further commitment, and why rejecting rigourism appears to threaten my account of requirement.

In ‘The Source of Authority of the State’, Anscombe generally treats rules and rights as two distinct species of a wider genus. At times, however, she also talks of ‘rules of right’ such as, for instance, “People have a right to get messages which have been sent to them.”¹¹⁶ Of such rules, she makes a distinction between “making an exception to some general rule of right”, on the one hand, and ‘pointing out’ that a

116 This formula is not taken directly from Anscombe’s text, but is adapted from one of her examples. In this context, to say that people have such a right may be used to stop someone from destroying a message which has not yet been seen by its intended recipient (1981g 143). It does not, I think, imply that we are obliged to do much to actively facilitate people’s receiving their letters.

given case does not “fall under the actually current rule”, or else “proposing to modify the current general rule”, on the other (1981g 143-144).

Some such distinction must be drawn by any account of ordinary rules, but it is not clear exactly what making an exception involves. Interestingly, Anscombe herself appears to envisage a fairly heterogeneous set of possible *grounds* for such exceptions, many of which do not obviously make any reference to ‘ends’. However, following Thompson’s example of playing a game rather than taking cover during an earthquake, we can say that at least *sometimes* an agent’s ends – here, that of staying safe – will constitute good grounds for ignoring an ordinary rule. If this is the case during such emergency situations, then ends are relevant in such situations. But if they are relevant *here*, are they not relevant in every case, albeit usually outweighed by the rule? If this is so, then we have lost our explanation of the sense of requirement. So it now appears that, in order to sustain my explanation of requirement, we need to be able to hold that in non-emergency situations, the rules can completely determine what the agent is to do, without reference to her ends.

It turns out, then, that rather than (1), what we need in order to explain requirement is the following, stronger, claim:

3. The rules of practice P, if P is a good practice, sometimes *completely* settle what an agent participating in P should do, in cases to which P applies (including in some tight corners).

However, a claim which seems much more plausible than our initial formulation of rigourism is the following:

4. *If* a rule *ever* completely settles what an agent should do, then it *always* completely settles what she should do, in those cases to which it applies.

(4) might be thought to gain support directly from the meaning of ‘completely’, but in conjunction with (3) it entails rigourism about the rules of practice P. So, if (3) is generally true of rules, rigourism as a general view will be true if (4) is true. We have now isolated the problem posed by the rules of ordinary practices, given the broader framework of practical generality: is there a reading of ‘completely’ which allows us

to explain the sense of requirement attaching to such rules, by endorsing (3), but which does not force us to accept (4)? In the final section I will explore a possible motivation for (4), which may be implicit in Thompson's work, before suggesting how some remarks of Anscombe's might point the way towards undermining that motivation.

Complete explanation and 'natural goodness'

A schematic argument in support of (4) might go like this. If Qs completely explain Rs, then wherever a Q is found, there must be some R that Q completely explains. If that were not the case, then Rs would depend on more than just Qs, and so a Q could *never* completely explain an R.

A move – with Aristotelian heritage – which can be made in response to such an argument distinguishes between, i) conditions which are required in addition to Q to explain R, and ii) the absence of *interference* with Q's ability to produce R. If we can understand all cases of Qs without Rs as cases in which something external is interfering with the Q's inner tendency to produce Rs, then we can say that where there is no such interference, Qs completely explains Rs.¹¹⁷

Thompson's view of living substances draws a distinction between processes which unfold according to the nature of the life-form which they manifest, and those which are interfered with. (The corresponding changes he sometimes labels 'non-accidental' and 'accidental', respectively.) It is only in the latter case, he says, that we need to appeal to factors external to the life-form itself in order to explain how things turn out with the living being in question (e.g. 2008 70-71; 204-207). To use his example, that this is a cat can be a sufficient explanation of why it has four legs, even though there are cats which never develop four legs. We need to invoke factors other than its being a cat in order to explain the latter case, but not the former, because we can construe such factors as interfering with the natural tendency of this thing to develop four legs otherwise. Thompson also suggests an analogy between the way in which a life-form can explain the characteristics of an individual bearer of it, and the

117 cf. Waterlow (28-29).

way in which the practice of promising can explain individual acts of promise-keeping (209). In a sense, the practice of promising stands to particular actions of promise-keeping as the life-form *the cat* stands to particular cats. If Thompson is in fact committed to (4), then, this does not come directly from the general line of thought about ‘complete’ explanations sketched in the first paragraph of this section.

I suspect that if Thompson is tempted by (4), this is explained by his acceptance of another broadly Aristotelian thought which is distinct from, but related to, that described above. This second claim is a particular way of fleshing out the suggestion that a living thing’s nature provides, to use Foot’s term, a standard of ‘natural goodness’ for it. In its most general form, the natural goodness framework holds that *what a thing is* determines the standards of goodness by which it can be evaluated. The claim I have in mind is a particular way of understanding this ‘determination’. This view has it that a ‘nature’ (the ‘what it is’) determines a particular way of being, and that an individual which manifests that nature is defective just insofar as it deviates from that way of being in one or more respects (Thompson 2008 81; cf. Foot 2001 27-34).¹¹⁸

According to this view, an individual thing’s nature completely explains its goodness, and in a strong sense. An individual’s nature completely explains its development, insofar as that development is in accordance with its nature. But in addition, the thing will be good just insofar as it is in accordance with its nature, because its nature provides the standard of goodness by which it is evaluated. Transposing this picture to the case of practices shows that it causes trouble for our account of ordinary practices. We want to say that such practices can (when they are themselves good)

118 Talk of ‘defect’ in connection with human beings generally makes me shudder, but my assessment of this reaction is complicated. Something like the framework of natural goodness has been used to defend what I take to be objectionable substantive ethical claims, and it is difficult at times to avoid the suspicion that this is not accidental. However, despite a murky past and present, it is not clear that the framework of natural goodness *per se* can only support conservative evaluative judgements, for instance. It is, first and foremost, a claim about the *form* of evaluative judgements, and is in principle compatible with a variety of contents. This is, I take it, how both Foot and Thompson view the framework.

completely explain the goodness of *some* of their manifestations. But we also want to allow that not all of their manifestations will be good. The framework sketched in the last few paragraphs allows for this possibility, but only by construing those manifestations as *poor manifestations* of the practice itself. But cruel or imprudent chess moves which are made in accordance with the rules of chess need not be bad *chess moves*, though they are by hypothesis bad actions.¹¹⁹ So we cannot accommodate the badness of such actions within this model.¹²⁰

However, if the first Aristotelian move worked, why could we not repeat it with the connection between manifesting a practice (or nature) and goodness? Why not say that it is possible for something to interfere with a chess move's being a good action, without making it a bad chess move? This would allow us to say that as a sort of default condition, manifesting a good practice makes an action good, though there are also cases where it is right to act contrary to the practice. That combination is what we need in order to give a satisfactory account of ordinary practices.

Anscombe seems to envisage something like this in her discussion of what is “unjust given the circumstances”, as opposed to “intrinsically unjust”, in ‘Modern Moral Philosophy’. She says that “[w]hat is unjust given e.g., normal circumstances is to deprive people of their ostensible property without legal procedure, not to pay debts, not to keep contracts and a host of other things of the kind”, but that in abnormal circumstances, doing these things may not be unjust (1981e 39).

Let's focus on the first example. Slightly obscured by her way of phrasing it here is the fact that Anscombe does not think that there is any natural right to have one's

119 In order to get a univocal evaluation of a given action, according to this framework of natural goodness, we will need to specify a single kind to which it belongs as both, say, a chess move and a piece of cruelty. The most obvious kind is, perhaps, ‘human action’. More needs to be said than either Thompson or Foot have said about how practices relate to action *per se*, such that promise-keeping could *ipso facto* be good *qua* human action. However, I suspect that any further complications involved in this picture will tend to be more favourable to the case for non-transparency than they are harmful to it.

120 If I have understood it correctly, this is the objection that Doug Lavin has raised against my account of ordinary practices, in conversation.

property protected by *some* legal procedure or other. Whilst it might – who knows? – be possible to discern in a ‘state of nature’ that it would be good to have property rights protected by some procedure or other, no-one has a *right* that such a procedure be followed until one is actually instituted, on her view (cf. 1981g). So what is ‘normally unjust’ for Anscombe is in fact to disregard the *actual* legal procedures in force in one’s context (providing that those procedures meet certain requirements of validity, etc.).

Another way of putting this point is to say that if an action is one of disregarding the relevant procedures, then that is normally enough to make it unjust. In such cases, that an action disregards the procedures completely explains its injustice. An objection to this reading might be that ‘normal’ is just doing *statistical* work here, and so cannot ground the required distinction between facilitating conditions and the absence of interference. However, this does not seem to be Anscombe’s intention. She says that sometimes it is the ‘consequences’ of such an action which mean it will not be unjust. But it is part of her argument against ‘consequentialism’ in that paper that the fact that consequences are sometimes relevant does *not* mean that they are always relevant. So, more particularly, it does not mean that they are always relevant but usually outweighed by other factors. It would be a mistake to identify ‘consequences’ with ‘ends’, but this position has the same structure as the one I have been describing with regard to ordinary practices.¹²¹

In many ways, what I have done in these last few sections is just to describe more precisely the *problem* which appears to be posed by the rules of ordinary practices. Speaking quite generally, the task for future work is to provide a positive account of how the rules of ordinary practices might justify the actions of agents who follow those rules. We have now seen that this positive account must allow the following three things. First, following the rules of a good practice must sometimes completely

121 It is worth noting in this connection one of Anscombe’s remarks from ‘On Brute Facts’:

Every description presupposes a context of normal procedure, but that context is not even implicitly described by the description. Exceptional circumstances could always make a difference, but they do not come into consideration without reason. (1981d 23)

explain the goodness or rationality of the rule-following action. Second, this must be compatible with *some* rule-following actions not being good or rational (it will sometimes be right to break the rules). Third, the cases mentioned in the second point cannot all be accommodated through the idea that bad actions manifesting a good practice must be *bad manifestations of the practice* (cruel chess moves need not be *bad chess moves*).

At the same time, though, setting out the problem in these terms has allowed us to get a clearer view of the limitations of Thompson's picture. We saw that there is a dichotomy in Thompson's account between instrumental justification, and justification by appeal to transparent, general practices or dispositions. This dichotomy can be read, now, as a *dilemma* for Thompson himself. The presence of this dichotomy in his framework means he cannot do justice to the practical role of the rules of ordinary practices: he must construe them either instrumentally, or as transparent.

In Chapter 6 we saw good reasons to think that the rules of ordinary practices cannot be understood through the instrumental model. But it is also a common point between me and Thompson that ordinary practices are not transparent. The premise which seemed to be responsible for forcing the dichotomy was what I called 'rigourism'. In the last two sections of this chapter I have given reasons for thinking that rigourism – whilst it gains some support from the framework of natural goodness defended by Thompson and Foot – has not been conclusively established by anything in Thompson's argument.

Given the arguments of the last two chapters, then, I suggest that the nature of ordinary practices gives us good reasons to reject rigourism, and along with it the view that practices must be transparent if they are to have explanatory and justificatory power. We can then use the concept of practical generality to explain the sense of requirement attaching to the rules of ordinary practices, without facing the dilemma created by Thompson's dichotomy. This frees us up to exploit my account of the ontology of ordinary practices to explain their practical role.

If I am right about this, then my ontology of practices is a key element in the explanation of the sense of requirement attaching to rules. A *full* positive account of the practical role of rules may require an alternative to the picture of natural goodness sketched by Thompson and Foot. In particular, it will require a more complex or fragmented picture of the relationship between goodness and explanation in the realm of human activity. But it may be that this complexity is entirely appropriate when it comes to ordinary human practices. The sense that this is so is, really, the basic *datum* of this chapter. In Chapter 5 I argued that AI conventions can be objects of practical identification. But they nevertheless remain significantly ‘external’ to us, and as we saw in Chapter 1 they are things which possess their own *histories*. So, the picture of rules, norms, and conventions which I have developed in this thesis as a whole certainly gives us no special reason to doubt the existence of this kind of complexity or fragmentation when it comes to ordinary practices.

Conclusion

In this chapter I have offered a positive account of the sense of requirement attaching to the rules of ordinary practices, drawing on the ontology of rules developed in the first part of the thesis. The metaphysical generality of rules, understood as AI conventions, allows them to apply in a tight corner, and so to do real normative work. Because of this, they can explain why an agent’s ends can become irrelevant when the agent falls under the scope of a valid rule.

My account drew on Thompson’s discussion of ‘practical generality’ in connection with promising. However, we saw that Thompson is unable to accommodate the rules of ordinary practices in his account, which requires that practices be ‘transparent’ if they are to have a role in explaining or justifying rule-following actions. If we drop this commitment, and the rigourism which appeared to prompt its inclusion, then we will be able to find a more satisfactory place for the rules of ordinary practices in practical reasoning.

This picture will ultimately need to be supplemented with a positive account of exactly how the goodness of a practice can provide a complete explanation of the goodness or rationality of its manifestations. However, we now have a much more distinct image of what this positive account will look like, and have cleared away some key obstacles to developing it in the future.

Conclusion

My opening move in this thesis was to suggest that the study of rules, norms, and conventions would benefit from greater attention to their ontology. The subsequent chapters have put this suggestion to work. We have seen that treating rules, norms, and conventions as abstract individuals explains how we individuate them and the behaviours which manifest them. It also helps to account for their explanatory role in ordinary and social-scientific discourse. In the final chapter we saw that the metaphysical ‘generality’ of rules, considered as abstract individuals, can explain a distinctive part of their normative profile: the sense of requirement which typically attaches to them.

Conventions are generated through the reproduction of patterns of behaviour. There have been two linked tendencies in much analytic philosophical work on human behaviour. One is to over-intellectualise the concept of human agency or human action. The paradigm has been intentional actions undertaken for reasons that the agent could endorse, and other cases have tended to be understood as defective approximations of this one. In a similar way, intentional action has been taken as the central example of *socially-articulate* action, and as the basis for social explanation and social ontology.

We have seen, though, that the scope of human agency is broader than the intellectualist picture suggests. Our agency is manifested in non-intentional actions which are in no way defective. Moreover, this is not something confined to the background against which we act on the social stage. We have seen that such actions can be just as much a part of the social world as our more reflective actions. There are such things as implicit norms, and norms governing non-intentional behaviour, in addition to the explicit norms which have been the focus of much philosophical discussion.

The second tendency has been to think that there is a sense in which all human action must be explained ‘in the first place’ (as Bernard Williams put it) by individual psychology. Williams, in his discussion of ‘formal individualism’, took this to be almost a truism. However, we have seen that when explaining the actions of those who follow conventions, it may be necessary to make reference to the agent’s relationship with something non-psychological: the AI convention itself. Although this is a significant departure from Williams’ apparent truism, properly understood this fact does not threaten the existence of human agency in the way that memetics appeared to do.

I have said that AI conventions can equally be implicit or explicit. This is not, however, to say that there are no interesting differences between the two cases. To the contrary, in Chapter 5 I outlined a number of different ways in which reflection on a norm, or on the behaviour which conforms to a norm, can transform our relationship to it. Another important possibility was the topic of Chapter 7. Rules, norms, and conventions, whether implicit or explicit, are ‘general’ and so are capable in principle of applying in a ‘tight corner’. But when agents are reflectively aware of this situation, as they typically are when following rules, this can give rise to what I have described as a sense of ‘requirement’ attached to following a rule. This all goes to show that in the practical realm, reflection upon some phenomenon often transforms the nature of that phenomenon itself. It need not be the kind of detached enquiry which leaves its object untouched.

Over the course of this thesis we have, then, explored important interrelations between the three categories named in my title: convention, reflection, and agency. But as always in philosophy, many questions remain, and one of the values of the approach I have adopted here is that it helps to clarify what these questions are and how to answer them in the future. I will close by suggesting how it does this.

I mentioned some of these questions throughout, particularly at the end of Chapter 3. There, we saw that understanding conventions as abstract individuals allowed us to suggest a new way of framing the questions raised by certain puzzling examples. Intuitively, it appears that something like the power of ‘branding’ can create a new kind of repeatable artefact, as in the case of the counterfeit jeans. I also mentioned

the appealing but elusive idea of a 'living tradition', which crops up fairly frequently in writing on culture. The theory of abstract individuals suggests that we will make progress in accounting for these cases by asking whether they can be understood through modifications of the notion of the 'reproduction' of an abstract individual, or not.

We can also open out from here to consider broader questions of interest in the philosophical literature which have not yet come up explicitly. Two themes in practical philosophy which come to mind are the aspiration to be fully reflective in action, and the extent to which our reasons for following rules derive from reasons to preserve the rule or practice in existence.

The thought that a responsible agent should aspire to be as reflective as possible concerning her actions, consistently with being able to get on with things, is implicit in quite a lot of philosophical writing. It has sometimes been thought that this aspiration is in tension with treating social practices as binding sources of practical reasons. Others have questioned the aspiration, whilst yet others have questioned the supposed incompatibility of reflectivity with the following of social practices. The theory of AI conventions suggests that these debates would do well to acknowledge an important distinction between the *objects* of reflection. Treating rules, norms, and conventions as entities with significant social reality means that they themselves, in addition to particular actions which may conform to or flout them, are appropriate objects of reflection. Moreover, the two corresponding forms of reflection can come apart: it is possible to reflect on the merits of a norm which one continues to follow without thought in the particular case;¹²² or to self-consciously follow a rule whose validity one has never stopped to interrogate. Given this, any claim for the importance of reflection should specify at which level(s) it is important (bearing in mind, too, that reflection can transform the practical situation itself, in the ways discussed above).

122 Despite his emphasis on the distinction between justifying a rule and justifying action which falls under that rule, Rawls did not appear to give any attention to this particular possibility. It may be that this is because his account of practices did not make room for what I have called 'implicit' norms.

The literature also contains expressions of the idea that good rules can bind us in part because the rules themselves have value, and because breaking the rules has a tendency to undermine them. The arguments of Chapter 6 speak against a certain way of construing this thought, according to which the reason to follow the rule would derive instrumentally from the end of preserving the rule in existence. However, this does not mean that there are no interesting alternative ways of understanding it.

Spelling out this thought will, however, require an account of what rules are and how they might be undermined by those who break them. The ontology of practices developed here bears on this question, by providing an account of the persistence-conditions for AI conventions. Roughly, an AI convention continues to exist as long as it is possible to reproduce it in the right sort of way.

Given this, we can expect to find interesting generalisations concerning how different sorts of conventions persist, depending on how they are reproduced. One plausible thought here is that there will be differences between how implicit and explicit conventions get reproduced. Explicitly recognising the existence of a norm or convention allows it to spread through explicit instruction, suggesting that the persistence of the convention will in general depend less directly on the ongoing possibility of acquaintance with instances of the convention. The differences here may be similar to those between oral and written literary works, discussed in Chapter 2. These differences in persistence-conditions may in turn be relevant to the question of whether and how the need to preserve conventions in existence relates to our reasons for following them.

Over the last few paragraphs, I have outlined some questions about social practices which have not been answered here. Each requires further investigation, and the last one in particular will need to be informed by relevant empirical work. The theory of AI conventions developed here does not answer all of these questions, and we shouldn't expect it to. I have tried to show, however, that what it does do is provide a valuable framework through which to approach these further topics in future work.

Bibliography

- Acerbi, A., & Mesoudi, A. (2015). 'If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution'. *Biology & Philosophy*, 30(4), 481–503.
- Aiello, John R. & Aiello, Tyra De Carlo. (1974) 'The Development of Personal Space: Proxemic Behavior of Children 6 Through 16'. *Human Ecology*, 2(3), pp.177-189.
- Anscombe, G. E. M. (1957) *Intention*. Basil Blackwell.
- , --. (1981a) 'The Question of Linguistic Idealism' in *The Collected Philosophical Papers of G. E. M. Anscombe. Volume 1: From Parmenides to Wittgenstein*. Basil Blackwell.
- , --. (1981b) 'Introduction' in *Ethics, Religion, and Politics: Collected Philosophical Papers Volume 3*. Basil Blackwell.
- , --. (1981c) 'On Promising and its Justice, and Whether it Need be Respected In Foro Interno' in *The Collected Philosophical Papers of G. E. M. Anscombe. Volume 3: Ethics, Religion, and Politics*. Basil Blackwell.
- , --. (1981d) 'On Brute Facts' in *The Collected Philosophical Papers of G. E. M. Anscombe. Volume 3: Ethics, Religion, and Politics*. Basil Blackwell.
- , --. (1981e) 'Modern Moral Philosophy' in *The Collected Philosophical Papers of G. E. M. Anscombe. Volume 3: Ethics, Religion, and Politics*. Basil Blackwell.
- , --. (1981f) 'Rules, Rights, and Promises' in *The Collected Philosophical Papers of G. E. M. Anscombe. Volume 3: Ethics, Religion, and Politics*. Basil Blackwell.
- , --. (1981g) 'On the Source of Authority of the State' in *The Collected Philosophical Papers of G. E. M. Anscombe. Volume 3: Ethics, Religion, and Politics*. Basil Blackwell.

- -, - -. (2005) 'Practical Inference' in *Human Life, Action, and Ethics: Essays by G. E. M. Anscombe*. Exeter: Imprint Academic.
- Austin, J. L. (1956-1957) 'A plea for excuses', *Proceedings of the Aristotelian Society*, New Series, Vol. 57, pp.1-30.
- Bakker, M. J., Dijk, J. G. Van, Pramono, A., Sutarni, S., & Tijssen, M. A. J. (2013). Latah : An Indonesian Startle Syndrome. *Movement Disorders*, 28(3), pp.370–379.
- Barnes, Barry. (2005) 'Practice as collective action', in *The Practice Turn in Contemporary Theory*. Edited by Theodore R. Schatzki, Karin Knorr Cetina, and Eike von Savigny. Routledge.
- Becker, Howard S. (1982) *Art Worlds*. University of California Press.
- Bennett, Charles H., Li, Ming, and Ma, Bin. (2003) 'Chain Letters & Evolutionary Histories', *Scientific American*, June 2003.
- Bicchieri, Cristina. (2005) *The Grammar of Society*. Cambridge: Cambridge University Press.
- -, - -. (2017) *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford: Oxford University Press.
- Blackmore, Susan. (1999) *The Meme Machine*. Oxford University Press.
- Bourdieu, Pierre. (1977) *Outline of a Theory of Practice*. Translated by Richard Nice. Cambridge University Press.
- -, - -. (1990) 'Belief and the Body' in *The Logic of Practice*. Translated by Richard Nice. Polity.
- Borges, Jorge Luis. (1998) 'Pierre Menard, Author of the *Quixote*', in *Fictions*. Translated by Andrew Hurley. London: Penguin.
- Boyd, Richard N. (1991) 'Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds'. *Philosophical Studies*, 61(1/2), pp.127-148.

- , --. (1999) 'Kinds, Complexity and Multiple Realization: Comments on Millikan's "Historical Kinds and the Special Sciences"'. *Philosophical Studies*, 95, pp.67-98.
- Brown, M. J., & Feldman, M. W. (2009). 'Sociocultural epistasis and cultural exaptation in footbinding, marriage form, and religious practices in early 20th-century Taiwan', *PNAS*, 106(52), pp.22139-22144.
- Bunzl, M., & Kreuter, R. (2003) 'Conventions Made Too Simple?' *Philosophy of the Social Sciences*, 33(4), pp.417–426.
- Burge, Tyler. (1975) 'On Knowledge and Convention'. *The Philosophical Review*, 84(2), pp.249-255.
- Buskell, A. (2016). 'Cultural longevity: Morin on cultural lineages'. *Biology & Philosophy*, 31(3), pp.435–446.
- Carpenter, M., & Nielsen, M. (2008). 'Tools, TV, and trust: Introduction to the special issue on imitation in typically-developing children'. *Journal of Experimental Child Psychology*, 101, pp.225–227.
- Cherry, C. (1973). Regulative Rules and Constitutive Rules. *The Philosophical Quarterly*, 23(93), 301–315.
- Claidiere, N., Scott-phillips, T. C., & Sperber, Dan. (2014). 'How Darwinian is cultural evolution?' *Philosophical Transactions of the Royal Society of Biology*, 369.
- Davidson, Donald. (2001a) 'Actions, Reasons, and Causes' in *Essays on Actions and Events: Collected Philosophical Essays Volume 1*. Oxford University Press.
- , --. (2001b) 'Agency' in *Essays on Actions and Events: Collected Philosophical Essays Volume 1*. Oxford University Press.
- , --. (2001c) 'Freedom to Act' in *Essays on Actions and Events: Collected Philosophical Essays Volume 1*. Oxford University Press.
- , --. (2001d) 'Knowing One's Own Mind' in *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3*. Oxford University Press.

- Davies, D. (2013) 'What Type of "Type" is a Film?' in *Art and Abstract Objects*. Oxford University Press.
- Dawkins, Richard. (1982). *The Extended Phenotype*. Oxford University Press.
- Dennett, Daniel C. (1996) *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Penguin.
- -, - -. (2001) 'The Evolution of Evaluators', in *The Evolution of Economic Diversity*. Ed. A. Nicita and U. Pagano. Routledge.
- Dworkin, Ronald. (2013) 'The Model of Rules – I', in *Taking Rights Seriously*. London: Bloomsbury Academic.
- Elster, Jon. (1983) *Explaining Technical Change: A Case Study in the Philosophy of Science*. Cambridge University Press.
- Enoch, David. (2006) 'Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action', *Philosophical Review*, 115(2), pp.169-198.
- Epstein, Brian. (2015) *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford University Press.
- Finnegan, Ruth. (2012) *Oral Literature in Africa*. Open Book Publishers.
- Foot, Philippa. (1978a) 'Morality as a System of Hypothetical Imperatives' in *Virtues & Vices*. Berkeley: University of California Press.
- -, - -. (1978b) 'A Reply to Professor Frankena' in *Virtues & Vices*. Berkeley: University of California Press.
- -, - -. (2001) *Natural Goodness*. Clarendon Press.
- Fry, Anna M. and Willis, Frank N. (1971) 'Invasion of personal space as a function of the age of the invader'. *The Psychological Record*, 21, pp.385-389.
- Geertz, Clifford. (1973) 'Thick Description: Towards an Interpretive Theory of Culture' in *The Interpretation of Cultures: Selected Essays by Clifford Geertz*. Basic Books.

- Geertz, Hildred. (1968) 'Latah in Java: A Theoretical Paradox'. *Indonesia*, 5, pp.93-104.
- Ghiselin, Michael T. (1974) 'A Radical Solution to the Species Problem'. *Systematic Zoology*, 23(4), pp.536-544.
- Gilbert, Margaret. (1989) *On Social Facts*. Routledge.
- , --. (1990) 'Walking Together: A Paradigmatic Social Phenomenon'. *Midwest Studies in Philosophy*, 15(1), pp.1-14.
- Godfrey-Smith, Peter. (2009) *Darwinian Populations and Natural Selection*. Oxford University Press.
- , --. (2012) 'Darwinism and Cultural Change'. *Philosophical Transactions of the Royal Society B*, 367, pp.2160-2170.
- Goffman, Erving. (1966) *Behavior in Public Places: Notes on the Social Organization of Gatherings*. The Free Press.
- Goody, Jack. (1986) *The Logic of Writing and the Organization of Society*. Cambridge University Press.
- Haase, Matthias. (2014) 'For Oneself and Toward Another: The Puzzle about Recognition', *Philosophical Topics*, 42(1), pp.113-152.
- Hacking, Ian. (1996) 'The looping effects of human kinds', in *Causal Cognition: A Multidisciplinary Debate*. Ed. Dan Sperber, David Premack, and Ann James Premack. Oxford University Press.
- Hall, Edward. T. (1955) 'The Anthropology of Manners'. *Scientific American*, 192(4), pp.84-91.
- , --. (1966) *The Hidden Dimension*. New York: Doubleday.
- Hare, R. M. (1964). 'The Promising Game', *Revue Internationale de Philosophie*, 62(243), 51-78.
- Hart, H. L. A. (1997) *The Concept of Law*, 2nd edition. Oxford: Oxford University Press.

- Hediger H. (1950) *Wild Animals in Captivity*. Translated by G. Sircom. Butterworths Scientific Publications.
- Heinaman, R. (Ed.). (1995). *Aristotle and Moral Realism*. London: UCL Press.
- Hempel, Carl G. (1965) 'The Logic of Functional Analysis' in *Aspects of Scientific Explanation And Other Essays in the Philosophy of Science*. The Free Press.
- Heyes, Cecilia. (2018) 'Enquire within: cultural evolution and cognitive science'. *Philosophical Transactions of the Royal Society B*, 373: 20170051.
- Hull, David L. (1978) 'A Matter of Individuality'. *Philosophy of Science*, 45(3), pp.335-360.
- Humberstone, I. L. (1996) 'Intrinsic/Extrinsic'. *Synthese*, 108, pp.205-267.
- Hume, D. (2000) *A Treatise of Human Nature*. Ed. David Fate Norton and Mary J. Norton eds.) Oxford University Press.
- Humphrey, Caroline, and Laidlaw, James. (1994) *The Archetypal Actions of Ritual: A Theory of Ritual Illustrated by the Jain Rite of Worship*. Clarendon Press.
- Hurley, S. and Chater, N. (2005) 'Introduction' in *Perspectives on Imitation: From Neuroscience to Social Science. Volume 2: Imitation, Human Development, and Culture*. MIT Press.
- Hursthouse, Rosalind. (1991). 'Arational Actions'. *The Journal of Philosophy*, 88(2), pp.57–68.
- Hyman, John. (2015) *Action, Knowledge, & Will*. Oxford University Press.
- Ingold, Tim. (2000) 'Hunting, gathering and perceiving' in *The Perception of the Environment: Essays on livelihood, dwelling and skill*. London: Routledge.
- James, Susan. (1984) *The Content of Social Explanation*. Cambridge University Press.
- Kaplan, David. (1990) 'Words'. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 64, pp.93-199.

- Kaya, Naz. and Erkip, Feyzan. (1999) 'Invasion of personal space under the condition of short-term crowding: a case study on an automatic teller machine'. *Journal of Environmental Psychology*, 19, pp.183-189.
- Korsgaard, Christine. (1986). 'Skepticism about Practical Reason'. *The Journal of Philosophy*, 83(1), 5-25.
- , --. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Lanska, Douglas J. (2018) 'Jumping Frenchmen, Miryachit, and Latah: Culture-Specific Hyperstartle-Plus Syndromes'. *Frontiers of Neurology and Neuroscience*, 42, pp.122-131.
- Lear, Jonathan. (2006) *Radical Hope: Ethics in the Face of Cultural Devastation*. Harvard University Press.
- Lewis, David K. (1969) *Convention: A Philosophical Study*. Harvard University Press.
- , --. (1983) 'Languages and Language' in *Philosophical Papers Volume 1*. Oxford University Press.
- Lord, Albert B. (1971) *The Singer of Tales*. New York: Atheneum.
- Marmor, Andrei. (1996) 'On Convention'. *Synthese*, 107, pp.349-371.
- Matthen, Mohan. (2013) 'Millikan's Historical Kinds', in *Millikan and Her Critics*. Edited by Dan Ryder, Justine Kingsbury, and Kenneth Williford. Wiley-Blackwell.
- Mauss, Marcel. (1973) 'Techniques of the Body', *Economy and Society*, 2(1), pp.70-88
- Merton, Robert K. (1949) 'Manifest and Latent Functions' in *Social Theory and Social Structure*. The Free Press.
- Midgley, Mary. (1974) 'The Game Game', *Philosophy*, Vol.49, No.189, pp.231-253.
- Millikan, R. (1984) *Language, Thought, and Other Biological Categories*. MIT Press.
- , --. (1990) 'Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox'. *The Philosophical Review*, 99(3), pp.323-353.

- , --. (1999a) 'Historical Kinds and the "Special Sciences"'. *Philosophical Studies*, 95, pp.45-65.
- , --. (1999b) 'Reply to Boyd's Commentary'. *Philosophical Studies*, 95, pp.99-102,
- , --. (2000) *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press.
- , --. (2005a) 'Language Conventions Made Simple' in *Language: A Biological Model*. Oxford University Press.
- , --. (2005b) 'In Defense of Public Language' in *Language: A Biological Model*. Oxford University Press.
- , --. (2005c) 'Why (Most) Kinds are not Classes' in *Language: A Biological Model*. Oxford University Press.
- , --. (2005d) 'Proper Function and Convention in Speech Acts' in *Language: A Biological Model*. Oxford University Press.
- , --. (2013) 'Reply to Matthen' in *Millikan and Her Critics*. Edited by Dan Ryder, Justine Kingsbury, and Kenneth Williford. Wiley-Blackwell.
- , --. (2014) 'The Tangle of Natural Purposes That Is Us', in *Contemporary Philosophical Naturalism and Its Implications*. Edited by Bana Bashour and Hans D. Miller. London: Routledge.
- Miton, H., Claidière, N., & Mercier, H. (2015). 'Universal cognitive mechanisms explain the cultural success of bloodletting'. *Evolution and Human Behavior*, 36, pp.303-312.
- Morin, O. (2013). 'How portraits turned their eyes upon us: Visual preferences and demographic change in cultural evolution'. *Evolution and Human Behavior*, 34(3), pp.222–229.
- , --. (2016a). 'Reasons to be fussy about cultural evolution'. *Biology & Philosophy*, 31(3), pp.447–458.
- , --. (2016b) *How Traditions Live and Die*. Oxford University Press.

- Negley, Elaine Nelson, and Manley, Joanne T. (1990) 'Environmental interventions in assaultive behaviour'. *Journal of Gerontological Nursing*, 16(3), pp.29-33.
- Ortner, Sherry. B. (1984) 'Theory in anthropology since the sixties'. *Comparative Studies in Society and History*, 26(1), pp.126–166.
- O'Shaughnessy, Brian. (1980) *The Will: A Dual Aspect Theory. Volume. 2*. Cambridge University Press.
- Otsuka, Jun. (2016) 'A critical review of the statisticalist debate'. *Biology & Philosophy*, 31, pp.459-482.
- Perry, A., Lwi, S. J., Verstaen, A., Dewar, C., Levenson, R. W., & Knight, R. T. (2016). 'The role of the orbitofrontal cortex in regulation of interpersonal space: evidence from frontal lesion and frontotemporal dementia patients'. *Social Cognitive and Affective Neuroscience*, 1894–1901.
- Pettit, P. (1988) 'The Consequentialist Can Recognise Rights'. *The Philosophical Quarterly*, Vol. 38, No. 150, pp.42-55.
- , --. (1990) 'Virtus Normativa: Rational Choice Perspectives'. *Ethics*, 100(4), pp.725-755.
- Peyton-Young, H. (1996) 'The Economics of Convention'. *The Journal of Economic Perspectives*, 10(2), pp.105-122.
- Popper, Karl. (1945) *The Open Society and Its Enemies. Volume One: The Spell of Plato*. Routledge.
- , --. (1963) 'Towards a Rational Theory of Tradition' in *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge.
- Pournelle, Jennifer R. and Algaze, Guillermo. (2010) 'Travels in Edin: Deltaic Resilience and Early Urbanism in Greater Mesopotamia' in *Preludes to Urbanism: Studies in the Late Chalcolithic of Mesopotamia in Honour of Joan Oates*. Edited by H. Crawford, A. McMahon, and N. Postgate. Oxford: Archeopress.

- Provine, R. R. (1992). 'Contagious laughter: Laughter is a sufficient stimulus for laughs and smiles'. *Bulletin of the Psychonomic Society*, 30(1), pp.1-4.
- Raz, Joseph. (1999) *Practical Reason and Norms*. Oxford University Press.
- -, - -. (2005) 'The Myth of Instrumental Rationality'. *Journal of Ethics & Social Philosophy*, 1(1), pp.2-28
- Richerson, Peter J. and Boyd, Robert. (2005) *Not By Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.
- Rousi, Rebekah. (2013) 'An uplifting experience – adopting ethnography to study elevator user experience'. *Ethnography Matters*.
ethnographymatters.net/blog/2013/04/02/an-uplifting-experience-adopting-ethnography-to-study-elevator-user-experience/
- Scanlon, T. M. (1998) *What We Owe to Each Other*. London: The Belknap Press of Harvard University Press.
- Schatzki, Theodore R. (2005) 'Introduction: practice theory', in *The Practice Turn in Contemporary Theory*. Edited by Theodore R. Schatzki, Karin Knorr Cetina, and Eike von Savigny. Routledge.
- Scott-phillips, T. C. (2017) 'A (Simple) Experimental Demonstration that Cultural Evolution is not Replicative, but Reconstructive — and an Explanation of Why this Difference Matters'. *Journal of Cognition and Culture*, 17, 1–11.
- Searle, John R. (1969) *Speech Acts*. Cambridge University Press.
- -, - -. (1995) *The Construction of Social Reality*. Penguin Books.
- -, - -. (2010) *Making the Social World: The Structure of Human Civilization*. Oxford University Press.
- Shove, E., & Pantzar, M. (2005) 'Consumers, Producers and Practices: Understanding the invention and reinvention of Nordic Walking'. *Journal of Consumer Culture*, 5(1), pp.43–64.

- Simons, Ronald C. (1996) *Boo!: Culture, Experience, and the Startle Reflex*. Oxford University Press.
- Sommer, Robert. (2002) 'Personal Space in a Digital Age' in Bechtel and Churchman (eds.) *Handbook of Environmental Psychology*. New York: John Wiley & Sons.
- Sorokowska, A., Sorokowski, P., Hilpert, P., Cantarero, K., Frackowiak, T., Ahmadi, K., ... Pierce, J. D. (2017) 'Preferred Interpersonal Distances'. *Journal of Cross-Cultural Psychology*, 48(4), 577–591.
- Sperber, D. (1996) *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.
- , --. (2001) 'An objection to the memetic approach to culture' in Aunger (ed.) *Darwinizing Culture: The Status of Memetics as a Science*. Oxford University Press.
- Strawson, P. F. (1996) *Individuals*. London: Routledge.
- Sterelny, Kim. (2006) 'Memes Revisited', *British Journal of the Philosophy of Science*, 57, pp.145-165.
- , --. (2011) 'Darwinian spaces: Peter Godfrey-Smith on selection and evolution'. *Biology and Philosophy*, 26, pp.489-500.
- , --. (2017) 'Cultural Evolution in California and Paris'. *Studies in History and Philosophy of Biological and Biomedical Sciences*,
- Steward, Helen. (2000) *The Ontology of Mind: Events, Processes, and States*. Oxford University Press.
- Stotts, Megan Henricks. (2017) 'Walking the tightrope: unrecognized conventions and arbitrariness'. *Inquiry: An Interdisciplinary Journal of Philosophy*, 60(8), pp.867-887.
- Swidler, Ann. (2005) 'What anchors cultural practices' in *The Practice Turn in Contemporary Theory*. Edited by Theodore R. Schatzki, Karin Knorr Cetina, and Eike von Savigny. Routledge.
- Taylor, Charles. (1971) 'Interpretation and the Sciences of Man'. *The Review of Metaphysics*, 25(1), pp.3-51.

- Thomasson, Amie. (2003) 'Foundations for a Social Ontology'. *ProtoSociology*, 18, pp.269-290.
- Thompson, Michael. (2004) 'What is it to Wrong Someone? A Puzzle about Justice', in *Reasons and Value: Essays on the Moral Philosophy of Joseph Raz*. Edited by J. Wallace, P. Pettit, S. Scheffler, and M. Smith. Oxford University Press.
- , --. (2008). *Life and Action: Elementary Structures of Practice and Practical Thought*. London: Harvard University Press.
- Tylor, E. B. (1871) *Primitive Culture: Researches into the development of mythology, philosophy, religion, art, and custom, Vol. 1*. London: John Murray.
- Velleman, D. (1997) 'How To Share An Intention'. *Philosophy and Phenomenological Research*, 57(1), pp.29-50.
- Voelkl, B., & Huber, L. (2000). 'True imitation in marmosets'. *Animal Behaviour*, 60, pp.195–202.
- Walters, L. (2013) 'Repeatable Artworks as Created Types', *British Journal of Aesthetics*, 53(4), pp.461-477.
- Watt-Smith, Tiffany. (2014) *On Flinching: Theatricality and Scientific Looking from Darwin to Shell-Shock*. Oxford University Press.
- Weber, Max. (1947) *The Theory of Social and Economic Organization*. Translated by A. M. Henderson and Talcott Parsons. Oxford University Press.
- Wetzel, L. (2009) *Types and Tokens: On abstract objects*. MIT Press.
- Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W., Boesch, C. (1999). 'Cultures in Chimpanzees'. *Nature*, 399, pp.682–685.
- Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009) 'Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee'. *Philosophical Transactions of the Royal Society of Biology*, pp.2417–2428.

- Williams, Bernard. --, --. (1981a) 'Justice as a Virtue' in *Moral Luck. Philosophical Papers 1973-1980*. Cambridge University Press.
- , --. (1981b) 'Internal & External Reasons' in *Moral Luck. Philosophical Papers 1973-1980*. Cambridge University Press.
- , --. (1981c) 'Practical Necessity' in *Moral Luck. Philosophical Papers 1973-1980*. Cambridge University Press.
- , --. (1993) 'Moral Incapacity' in *Proceedings of Aristotelian Society*, New Series, 93, pp.59-70.
- , --. (1995a) 'Making Sense of Humanity' in *Making Sense of Humanity and other philosophical papers 1982-1993*. Cambridge University Press.
- , --. (1995b) 'Formal & Substantial Individualism' in *Making Sense of Humanity and other philosophical papers 1982-1993*. Cambridge University Press.
- Wollheim, Richard. (2015) *Art and Its Objects*. Cambridge University Press.
- Yap, P. M. (1952) 'The Latah Reaction: Its Pathodynamics and Nosological Position'. *The Journal of Mental Science*, XCVIII(413), 515–564.